We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,900 Open access books available
116,000 International authors and editors
120M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Review of Recently Progress on Neural Electronics and Memcomputing Applications in Intrinsic SiO$_x$-Based Resistive Switching Memory

Cheng-Chih Hsieh, Yao-Feng Chang, Ying-Chen Chen, Xiaohan Wu, Meiqi Guo, Fei Zhou, Sungjun Kim, Burt Fowler, Chih-Yang Lin, Chih-Hung Pan, Ting-Chang Chang and Jack C. Lee

Abstract

In this chapter, we focus on the recent progress on memcomputing (memristor + computing) in intrinsic SiO$_x$-based resistive switching memory (ReRAM or called memristor). In the first section of the chapter, we investigate neuromorphic computing by mimicking the synaptic behaviors in integrating one-diode and one-resistive switching element (1D-1R) architecture. The power consumption can be minimized further in synaptic functions because sneak-path current has been suppressed and the capability for spike-induced synaptic behaviors has been demonstrated, representing critical milestones and achievements for the application of conventional SiO$_x$-based materials in future advanced neuromorphic computing. In the next section of chapter, we will discuss an implementation technique of implication operations for logic-in-memory computation by using a SiO$_x$-based memristor. The implication function and its truth table have been implemented with the unipolar or nonpolar operation scheme. Furthermore, a circuit with 1D-1R architecture with a 4 × 4 crossbar array has been demonstrated, which realizes the functionality of a one-bit full adder as same as CMOS logic circuits with lower design area requirement. This chapter suggests that a simple, robust approach to realize memcomputing chips is quite compatible with large-scale CMOS manufacturing technology by using an intrinsic SiO$_x$-based memristor.

Keywords: resistive switching, synaptic device, silicon oxide, neuromorphic computing
1. Background

In recent 20 years, emerging memory has drawn a lot of interest and attention as a promising candidate for next generation nonvolatile memory (NVM) [1–3]. Traditional “charge”-based NVM (Flash) will face the potential scaling challenge below the 10 nm node with reliability and power consumption issues [4, 5]. Resistive switching (RS) memory, or we call resistive random access memory (ReRAM), operates by controlling device “resistance” with an external electrical bias [6–9], leading to better electrical performance, smaller design area (4F²), and excellent cycling endurance [10] based on the 2015 International Technology Roadmap for Semiconductors (ITRS) (ReRAM is one of two recommended candidate technologies (the other one is the STT-MRAM) for emerging memory devices) [11]. Moreover, RS-based memories represent a new class of devices compatible with applications that go beyond traditional electronics configurations, for example, three-dimensional (3D) stacking, nanobatteries, neuromicroelectronics, and Boolean logic operations [12–17].

In 1971, Chua presented the theoretical basis for a passive two-terminal circuit device called a “memristor” (a contraction of memory and resistor) [18]. If realized, the memristor would then join the resistor, inductor, and capacitor to provide four basic circuit elements. In 2010, researchers in HP lab realized the memristor in nanoscale titanium dioxide (TiO₂) cross-point structure [17], and the field has advanced quickly growth over the past decade as a result. Having demonstrated the existence of memristors in the lab, additional research efforts focused on the potential applications that this emerging new circuit element enables [19]. In recent years, memristors have been extensively studied as a nonvolatile memory called resistive random-access-memory (named ReRAM or RRAM) to potentially replace dynamic random-access-memory (DRAM) and flash memory [20]. Memristors have also gained tremendous interest in the field of neuromicroelectronics and synaptic electronics, which aims to build artificial synaptic devices that emulate the computations performed by biological synapses [21–25]. Jo et al. described possible applications in artificial intelligence using memristors as synapses in neuromorphic circuits [15]. Another interesting application is to use memristors for arithmetic/logic operations, such as an adder circuit or a multiplier circuit.

In the literature, arithmetic operations are proposed using the memristor as a: (1) switch, (2) programmable interconnect, and (3) computational element. In the first approach, crossbar arrays of memristor switches are connected to a row of weighting resistors and sensing logic to build an analog arithmetic processor [26]. The switches control the current flow (ON/OFF) through the weighting resistor, which then controls the analog voltage at the sensing amplifier end. The resistance of the weighting resistor assigns the appropriate bit significance to the each row’s current contribution. The memristor-CMOS technology may be used to realize the same types of arithmetic circuits that are developed in CMOS/FPGA (field programmable gate array) technology [27, 28]. Last but not the least, a more universal approach for constructing the logic operations from memristors is via “material implication” (or an “IMP” operation). In 2010, researchers showed that all fundamental Boolean logic functions can be realized by using memristors with the IMP operation [17]. Later work built on these findings to construct larger logic blocks such as adders and multipliers [29–31], linear feedback shift
registers [32], and counters [32]. The advantages of memcomputing (memristor + computing) are not only to store and process information on the same physical platform, but also to allow massively parallel computations in a simple crossbar array architecture.

Otherwise, neuroelectronics and synaptic electronics are interesting applications for ReRAM that aim to build artificial synaptic devices that emulate the computations performed by biological synapses [15, 33]. These emerging fields of research potentially have better efficiency in solving complex problems and outperform real-time processing of unstructured data than conventional von Neumann computational systems [34]. There have been many studies of binary metal oxide-based and perovskite oxide-based resistance switching characteristics for synapse-like electronic device development [35, 36], which can have operating instability issues due to difficulty in controlling stoichiometric compositions [37, 38]. Therefore, a simple process that is compatible with conventional complementary metal-oxide semiconductor (CMOS) fabrication allows multilayer compositional engineering and provides good electrical stability and high yield, which are critical requirements for neuroelectronics realization [39]. Silicon oxide (SiO\(_2\)) has long been used as gate dielectrics for metal-oxide-semiconductor field-effect transistors. In addition to excellent insulating properties, resistive switching properties have been observed in SiO\(_x\) materials as early as 1962 by Hickmott and 1967 by Simmons and Verderber [40–42]. Yao et al. also have reported SiO\(_x\)-based RS behaviors in vacuum, indicating that this traditional material can be converted to an active component by controlling the external electrical manipulation [43–45]. Several recent reports describe using SiO\(_x\) as the active switching medium in resistive switching memory devices [46–49]. We have further demonstrated a Si diode (1D) with low reverse-bias current integrated with a SiO\(_x\)-based memory element (1R) using nanosphere lithography and deep Si etching to pattern a P\(^{++}\)/N\(^{+}\)/N\(^{++}\) epitaxial Si wafer [50].

2. Introduction

In this chapter, first SiO\(_x\)-based resistive switching memory elements (1R) are integrated with Si diodes (1D) using conventional CMOS processing to demonstrate a 1D-1R device with synaptic behaviors. Compared with our previous work (in most cases investigating only the 1R device system), the Si diode provides low reverse-bias current and high power efficiency for future neuromorphic computing array architectures. Unlike other binary or complex metal oxide materials [51], SiO\(_2\) has been used in CMOS manufacturing for over 50 years due to its excellent electrical isolation properties, low-cost, high chemical stability, compatibility with mainstream integrated circuit materials, high-throughput processing, and large-area production using chemical vapor deposition (CVD). A 1D-1R architecture fabricated at the wafer-scale using conventional CMOS processing can, therefore, be well controlled in thickness, size, and electrical characteristics by precisely controlling the doping levels of the diode layers and the temperature and flow-rate of the oxide CVD process [52]. Synaptic device performance is characterized in a prototype 1D-1R array configuration. Robust biological synaptic behaviors such as long-term potentiation (LTP), long-term depression (LTD), and spike-timing-dependent plasticity (STDP) are demonstrated with excellent uniformity, low
operational variability, and good suppression of static power consumption [51]. A bio-inspired proton exchange resistive switching model is used to help characterize this novel application for SiO$_x$ materials. The SET transition in the resistive switching memory is modeled as hydrogen (proton) release from the (Si-H)$_2$ defect to generate a conductive hydrogen bridge, and the RESET transition is modeled as an electrochemical reaction (proton capture) that reforms nonconductive (SiH)$_2$. The synaptic behaviors exhibited by the 1D-1R device demonstrates good potential for using a simple and robust approach for large-scale integration of programmable neuromorphic chips using CMOS technology.

Second, the application of SiO$_x$-based memristors for material implication operations is examined. A bidirectional implication scheme is demonstrated and tested in an actual circuit using SiO$_x$-based memristors. The symmetric unipolar memristive behavior of the SiO$_x$-based memristor enables the use of two sets of implication voltage setups, one positive and the other negative, hence the name “bidirectional”. Progressing one step further from the initial concept demonstrated by Borghetti et al. and our previous work, a one-bit full adder is realized by using the material implication technique on a crossbar structure with a one-diode one-memristor (1D-1R) array. Several potential application problems such as sneak current paths within an array and using a select transistor as the load resistor are discussed in detail. The results suggest that a memristor-enabled logic circuit is most suitable for applications requiring low-speed, low-power, and high-density.

3. Method and experiment

Secondary electron microscopy (SEM) images show a top-down view of a 1D-1R test structure (Figure 1a), a tilted (45°) view of the 1R device (Figure 1b) and a cross-section image of the 1R device showing layer information (Figure 1c). The devices were fabricated at XFAB in Lubbock TX using the XC06 CMOS process technology. The 1R device was fabricated by first implanting the Si substrate to form an n-type lower electrode. The active SiO$_x$ memory layer was then deposited to a thickness of 40 nm using plasma-enhanced chemical vapor deposition (PECVD). This thickness is known to provide high electroforming yield and good memory endurance [53]. An n-type polysilicon layer was deposited onto the SiO$_x$ layer to form the top electrode. An opening in the polysilicon layer was made after all thermal oxidation and implant anneal steps are complete (Figure 1b). A first dielectric layer was then deposited over the polysilicon top electrode. Tungsten plugs were used to make electrical contact to the n-type Si lower electrode and the polysilicon top electrode. After all the back-end dielectrics and a passivation layer were deposited, the back-end dielectric layers were removed using reactive ion etch (RIE) to the Si substrate. This RIE step cleared-out the SiO$_x$ layer inside the hole, and created a SiO$_x$ sidewall where the memory device is formed (Figure 1c). Polymer residue that remained after the post-RIE cleaning steps was removed by a 30-s buffered oxide etch (BOE). The pn diode used in the 1D-1R test structures was formed by an implanted p-well inside a deep n-well with 40 V reverse-bias breakdown voltage, 1 nA reverse-bias leakage current and 0.5 V forward voltage. The active memory area of the 1R device is $2 \times 2$ μm$^2$ and the overall size including metal interconnects is $21.9 \times 21.9$ μm$^2$. The overall size of the 1D device is $41 \times 19$ μm$^2$. A lake shore cryotronics vacuum probe chamber (<1 mTorr) and Agilent B1500A device analyzer were used.
to electroform devices and measure the DC/AC $I-V$ response. The SET process programs the device to a conductive, low-resistance state (LRS). The RESET process programs each device to a low-conductance, high-resistance state (HRS). A Kratos Axis Ultra HSA X-ray photoelectron spectrometer (XPS) equipped with a monochromatized aluminum X-ray source was used to analyze several SiO$_x$ materials deposited in our laboratory using different methods. Calibration of the binding energy scale was set by fixing the C-(C,H) peak at 284.4 eV. Figure 1d shows XPS analysis results for the O-1s and Si-2p binding energies in thermal oxide grown by low-pressure chemical vapor deposition (LPCVD) and PECVD oxide. The existence of stoichiometric SiO$_2$ can be observed in thermal oxide (binding energy Si: 103.2 eV; O: 532.5 eV) with essentially no suboxide bonding being detected. In contrast, the PECVD oxide has nonstoichiometric SiO$_x$ ($x$ is about 1.6 based on the peak position and orbital valence) composition in the switching layer, as indicated by the peak-binding energies in the XPS spectra (O: 530.5 eV; Si: 101.9 eV, and 100.9 eV) [54, 55], which may promote low-energy defect generation during the electroforming process.

4. Results and discussions

Figure 2a–d show $I-V$ characteristics for DC voltage sweeps applied to the SiO$_x$-based 1D-1R devices fabricated by the conventional CMOS process. Voltage was applied to the 1D top electrode (p-type Si) with bottom 1R electrode (n-type Si) at ground. All testing was done in
vacuum. To establish reversible resistive switching in each SiO$_x$-based 1R ReRAM device, a forward/backward voltage sweep (Figure 2a) was used to electroform each device, where current is observed to increase dramatically at 22.5 ± 2.9 V during the forward voltage sweep. Electroforming is completed during the backward voltage sweep from the maximum sweeping voltage to 0 V, resulting in the formation of a conductive filament (CF) and setting the device to a LRS. After electroformation, RS performance of 1D-1R can be stabilized by 10 times cycles (Figure 2b). For SET process, a 10 V forward/backward sweep is applied without any compliance current limitation (CCL) to change the device from HRS to LRS; for RESET process, a 17 V, single sweep is done to change the device from LRS to HRS. The HRS/LRS resistance ratio can be read out at 1 V bias with satisfying sensing requirements (~10$^3$) [3, 26]. For diode characteristics, the forward current can reach 100 mA at 2 V (current density 1.15 × 10$^{-5}$ A/μm$^2$ at 1 V), which indicates a forward current level high enough to support the RESET process. The reverse current is below 1 × 10$^{-12}$ A at −5 V. Compared with Schottky diodes (potentially useful for 3D arrays), the advantages of Si-based PN diodes include low reverse current, high reverse-bias breakdown voltage, and fewer stability issues [45]. The quality of the Si-based PN diode can dramatically affect diode reverse or forward current characteristics, as well as power consumption (describe below). Also, the chosen Si-based PN diode configuration has high reverse breakdown voltage (>40 V), which is important for SiO$_x$-based ReRAM operating in an array. Figure 2c demonstrates the gradual change of

Figure 2. DC sweep resistive switching behaviors of 1D-1R architecture: (a) Forward/backward voltage sweeps during electroforming process averaged for 256 devices in a 16 × 16 array (gray curves). The electroforming voltage ($V_{\text{DMC, FORM}}$) is defined as the voltage where maximum current change occurs during the forward sweep. (b) 10-I-V resistive switching SET/RESET cycles. The inset shows the average of 100 measurement cycles of diode I-V behavior. (c) Effects of voltage modulation on I-V curves in SET process plotted on linear-scale, where the applied SET voltage sweep increases from 3.5 to 9.5 V in 0.5 V steps. The inset shows effects of voltage modulation on I-V curves in RESET process plotted on log-scale, where the applied RESET voltage sweep increases from 11.0 to 18.0 V in 0.5 V steps. (d) The resistance states of initial fresh device, SET DC voltage modulation, and RESET DC voltage modulation. For SET voltage sweep, increases from 3.5 to 10 V in 0.5 V steps; for RESET voltage sweep, increases from 11 to 20 V in 0.5 V steps. The resistance reads at 1 V for each state. (e) Retention measurement results of multi-state programming obtained by controlling the SET voltage. (f) Proton exchange induced resistive switching model and defect transitions. Figure reprinted by [19].
resistive states by modulating the voltage sweep range continuously during the SET and RESET (inset) process, respectively. Specifically, SET and RESET voltages were changed from 3.5 to 9.5 V in 0.5 V increments and from 11 to 18 V in 0.5 V decrements, respectively, thus potentially enabling multilevel programming in a single memory cell and demonstrate the status stability before/after sweeps. It may be noted that the electroforming voltages measured here (~ 28 V) are somewhat higher than those measured in previous work on metal-oxide-semiconductor device architectures or nanopillar type 1D-1R architectures [50, 56, 57], which may be due to fewer electrically active defects being near the SiO₂ sidewall as a result of the fabrication process. For example, several high temperature steps (>650°C) were done after PECVD SiO₂ deposition, namely: polysilicon deposition, thermal oxidation, and implant anneals, which might densify the SiO₂ layer, reduce the as-deposited defect levels, increase the soft breakdown threshold, and thus increase the filament formation energy during the subsequent electroforming process (resulting in forming voltage increase). Interestingly, the RESET voltage (the voltage at which LRS current begins to decrease) has been found to be greater than or equal to the SET voltage (where HRS current increases sharply), which is a unique characteristic of the SiOₓ-based ReRAM as compared to other materials systems [36, 58]. The difference between RESET and SET voltages can potentially be controlled by optimizing the series resistance in the circuit, choice of electrode materials, and by doping effects that modulate the interfacial contact resistance [59]. The switching voltage is largely independent of device size and SiO₂ thickness. Figure 2e shows multilevel retention performance of SiOₓ-based 1D-1R devices obtained by controlling the maximum SET voltage from 3 to 9 V. The readout current of LRS and HRS is measured at 1 V every 60 s after each programming operation. Although the state’s stability still needs to be improved (no equal split of resistance states), the retention reliability test demonstrates operation by using different SET voltages, and no degradation is observed for more than 10⁴ s, thus confirming the stable, nonvolatile nature of the SiOₓ-based 1D-1R devices. In recent studies, a possible proton exchange model consistent with the observed resistive switching I-V response has been proposed, as shown in Figure 2f [59, 60]. Several studies have used transmission electron microscopy (TEM) to document the presence of Si nanocrystals within the CF [43, 61, 62], but it is not yet clear whether resistive switching (RS) is the result of an overall increase in nanocrystal size or whether switching occurs in “GAP” regions in between nanocrystals. Most models of ReRAM switching involve the drift or diffusion of O²⁻ ions (or oxygen vacancy defects) [39], but these models cannot explain the unconventional I-V response. For example, the backward scan effect (see Figure 2a, backward scan) is very difficult to explain using a simple oxygen vacancy-switching model. The backward scan effect is a phenomenon where the duration of the reverse sweep during electroforming or RESET determines whether a state change occurs, and has been characterized using DC and AC pulse response in a previous study investigating our resistive switching model [57]. In addition, ambient effects on resistive switching suggest that the defects responsible for switching are hydrogen-passivated or are in some way protected from direct reaction with ambient oxygen and water until a switching events occurs [56, 63]. The detailed interactions between ambient gases and proton (or cation) mobility is an important topic that may provide a deeper understanding of resistive switching mechanisms [64–68], specifically those in oxide-based valence change memory (VCM)-type ReRAMs [69–71]. The models used here to describe the possible
SiO$_x$-based RS mechanisms differ from most conventional models by considering that the defects responsible for RS may remain localized within the switching region so that resistive switching occurs when a collection of defects are driven between conductive and nonconductive forms [56]. A thorough review of the reported electrical and structural properties of known SiO$_x$ defects has identified a plausible model for the conductive filament that is similar to models used to describe stress-induced leakage current and breakdown in SiO$_x$ materials, where defect concentration increases as a result of electrical stress to the point where percolation pathways capable of conducting appreciable current (>1 uA) are formed [59]. Incorporating known proton exchange reactions that can dramatically alter the conductivity of specific defects further leads to a model where the LRS has a large concentration of conductive defects within the switching region, and, conversely, when the device is programmed to the HRS, most of the defects are converted to their nonconductive form. The electrically conductive hydrogen bridge (Si-H-Si) is viewed as the most likely defect responsible for the LRS due to the location of its energy levels relative to the oxide conduction band and its small effective bandgap energy [59, 60]. Adding a proton to Si-H-Si forms the nonconductive (SiH)$_2$ defect and proton desorption from (SiH)$_2$ reforms Si-H-Si, which are well-understood electrochemical reactions that could enable localized switching without incorporating ion diffusion or drift mechanisms into the model. The SET transition voltage from HRS to LRS occurs at ~2.5 V in the I-V response, and is very near the activation energy for proton desorption from SiH (~2.5 eV), thus making the defect transformation from (SiH)$_2$ to Si-H-Si a logical assignment for the SET transition [59, 60]. In this model, the proton that is lost from (SiH)$_2$ reacts electrochemically with (SiOH)$_2$, which is simply chemisorbed H$_2$O, to form the fixed positive charged H$_3$O$^+$ defect. The transition from LRS to HRS is modeled as being initiated by electron injection into H$_3$O$^+$ that induces proton release and electrochemical reaction with Si-H-Si to reform (SiH)$_2$ [59, 60]. The localized proton exchange switching model can thus be written as (SiH)$_2$ + (SiOH)$_2$ ↔ Si-H-Si + Si$_2$O$=O$H$3$O$^+$, where a voltage drop of ~2.5 V across the switching is required to drive the reversible reaction. The RS model not only provides insights into multilevel operational characteristics but also implies a possible biomimetic chemical reaction similar to reactive oxygen species (ROS-like) production for future device characterizations [72].

Figure 3a–h show contour plots of the current-change ratio achieved by modulating the AC pulse height and pulse width applied to 1D-1R devices for both SET and RESET switching events, leading to optimized waveform designs for a biological synaptic device. The current-change ratio is defined as $\log_{10}(I_{\text{FINAL}}/I_{\text{INITIAL}})$, where $I_{\text{INITIAL}}$ and $I_{\text{FINAL}}$ are the currents measured at 1 V before and after applying the programming waveform, respectively. The SET/RESET sweeps from same initial resistance state (precondition programming) is to eliminate the accumulating SET/RESET effect after each cycle. One can observe by inspecting the contour lines in Figure 3 that when larger pulse heights (higher voltages) are applied to the device, shorter pulse widths are needed to achieve a similar current-change ratio. In general, we find that a single 1R device operates at higher speed and requires lower programming voltages as compared to a 1D-1R device. The higher operating voltages and lower operating speed of the integrated 1D-1R device may result from higher parasitic resistance in the Si electrodes, their contacts and the diode, as well as higher parasitic
capacitance in the diode, all of which can act to degrade the pulse mapping results shown in Figure 3a and b. It should be noted that current sneak-path issues in arrays and writing disturbance of 1R devices would cause misread problems and state disturbance, and substantially increase standby power consumption and information instability. The 1D-1R devices are used to suppress sneak-path currents, and perform much better than 1R devices in an array architecture (potential 1 Gbit array support in 10% readout-margin at 1V read).

From Figure 3a and b, it can be calculated that the switching energies to achieve at least a one-order-of-magnitude change in resistance in the 1D-1R architecture are about 0.01 pJ for SET and 1.54 nJ for RESET operations. However, due to the suppression of sneak-path current, the standby power during a 1 V read operation can be dramatically reduced in 1D-1R devices (1 pW) as compared to 1R devices (1 μW, due to 1R nonpolar switching behaviors) [73]. Minimizing the total power consumption due to sneak-path current is as crucial as reducing the synaptic dissipation.

Most importantly, the pulse mapping results not only demonstrate the potential for multilevel programming by properly designing the pulse waveforms for SET and RESET operations, but also demonstrate the potential to realize biological synaptic behaviors. Figure 3c–h demonstrate...
the optimization waveform design for biological synaptic behaviors in 1D-1R SiO\textsubscript{x}-based resistive switching memories. The long-term potentiation (LTP) and long-term depression (LTD) are a long-lasting enhancement/reduction in signal transmission between two neurons (similar with long-lasting conductance increase/decrease between HRS and LRS for resistive-type memory devices), which can be realized by designing the SET and RESET pulse waveform to use either identical (fixed pulse width and pulse height, as shown in Figure 3c–f) or nonidentical (variable pulse width or pulse height, as shown in Figure 3g and h) pulsing techniques. The trade-offs between high dynamic range and gradual multilevel programming performance (Figure 3e–h) needed to be considered, and it was found that the nonidentical pulse waveform method may have the advantages (larger than identical pulse waveform method). Although nonidentical pulsing might require a more complex neuromorphic circuit, our results show that this approach enables more efficient programming to target states while maintaining a larger dynamic range (Figure 3g–h). The use of nonidentical pulse heights ranging from 4 to 10 V in 0.3 V increments (for LTP) and ranging from 11 to 17 V in 0.3 V decrements (for LTD) allow the dynamic range to be mapped for pulse widths ranging from 100 ns to 1 ms, thereby realizing biological synapse behaviors in the SiO\textsubscript{x}-based 1D-1R architecture (Figure 3g–h). The switching energy is defined as $I \times V \times \delta t$, where $\delta t$ is the pulse width. For $\delta t = 100$ ns, the smallest switching energies are ~6 and ~130 pJ for LTP and LTD, respectively. The larger energy for LTD is mainly due to the lower resistance of the LRS (~93 k\textohm) compared to the HRS (~260 M\textohm), which results in higher switching current (118.28 $\mu$A) for the RESET process than for the SET process (15.38 nA). In order to minimize synaptic energy consumption all three components—programming current (~nA level switching), pulse amplitude (<1 V) and programming time (<10 ns)—need to be minimized. In SiO\textsubscript{x}-based ReRAM and in other material systems, an exponential voltage–time relationship is commonly observed. A small increase in programming voltage will decrease programming time exponentially, as shown in Figure 3a. For RESET process (both 1R and 1D-1R structures, Figure 3b), the process integration may result in certain level of distortion (parasitic resistance/capacitance and possible parasitic depletion region capacitance from 1D) to affect the pulse mapping results. Hence, low programming energy is obtained by minimizing the programming time (traded off by increasing the pulse amplitude slightly) for ReRAM. Further decreases in synaptic energy consumption during the switching process to fJ levels will be challenging but important to build very large-scale systems (the designed pulse waveform optimization and generation is in process).

Such flexible artificial control built with synaptic devices could provide a suitable platform for a broad range of computing applications, as shown Figure 4. Some of the advantages that SiO\textsubscript{x}-based synaptic devices provide over other resistive switching materials include a higher dynamic range ($\sim 10^4$) [57] and the potential to achieve as many as 10–60 multilevel states (depend on the stability) in both LTP and LTD by changing the increment/decrement of the voltage step, as shown in Figure 4a. These advantages may arise as the result of there being a large number of defects within the switching region of the memory device. Switching is modeled as a change in conductivity of a group of defects within the switching region. In this framework, defects are not created or destroyed, but are simply driven between conductive and nonconductive forms by proton exchange reactions that are known to occur in SiO\textsubscript{x} materials (Figure 2f) [60]. The SET and RESET switching transitions can be described in more
Detail with the aid of the electron energy band diagrams shown in Figure 4b, which were constructed using the thermodynamic and switching charge-state energy levels reported by Blochl in 2000 [74]. The ideal energy band diagrams in Figure 4b represent only a single electron pathway through the memory device, whereas in reality there are likely many such percolation pathways in parallel. The SET transition is modeled as being the result of trap-assisted electron tunneling through \((\text{SiH})_2\) defects (a voltage-triggered mechanism, due to less current flow in the initial stage of SET process) that stimulates \(H^+\) desorption and reaction of \(H^+\) with absorbed water \((\text{SiOH})_2\) to form conductive \(\text{Si-H-Si}\) and \(\text{H}_3\text{O}^+\) (Figure 2f). Trap-assisted tunneling can only occur when the bias across the switching region is \(\geq 2.6\) V, which is the effective bandgap of the \((\text{SiH})_2\) defect and compares well with the observed minimum SET voltage of \(-2.5\) V in the \(I-V\) response [59, 60]. The RESET transition is modeled as being the result of Fowler-Nordheim electron tunneling into the \(\text{H}_3\text{O}^+\) defect (possibly current-induced Joule heating due to large current flow through the filament) that stimulates proton release and electrochemical reactions to reform \((\text{SiH})_2\) and \((\text{SiOH})_2\) (Figure 2f) [60]. The band diagrams shown in Figure 4b are found to be consistent with measured electron energy barriers [60] and electroluminescence results reported for similar devices [62].

Figure 4. Demonstration of a SiO\(_x\)-based synaptic device. (a) Sequential LTP/LTD behaviors as a function of increment/decrement voltage steps (0.1, 0.2, and 0.3 V) by non-identical pulse form. For the non-identical pulse method, pulse height modulation changes continuously from 4 V to 10 V for LTP, and changes continuously from 11 to 17 V for LTD. Pulse width is fixed at 10 μs in both cases. (b) Energy band diagrams: For HRS and SET process, showing theoretical bandgap of \((\text{SiH})_2\) defect within gap region of length \(l_\text{GAP}\), theoretical bandgap of \(\text{Si-H-Si}\) defects outside the gap region, and trap-assisted-tunneling SET transition (green arrow). Barrier height to electron transport is \(\phi \sim 0.8\) eV. For the LRS and RESET process, showing theoretical bandgap of \(\text{Si-H-Si}\), \(\text{H}_3\text{O}^+\) energy level, switching region of length \(l_\text{SW}\), and Fowler-Nordheim tunneling RESET transition (red arrow). (c–d) A pulse waveform design using the non-identical pulse method for demonstration of spike-timing-dependent plasticity (STDP) as a function of spike pulse width intervals. For the potentiation of conductance strength change, the overall pulse waveform (pulse width fixed at 10 μs in this case) based on the delay of spike timing between neurons is shown in (c). Similarly, for the depression of conductance strength change, the overall pulse waveform (pulse width fixed at 10 μs in this case) based on the delay of spike timing between neurons is shown in (d). (e–f) A demonstration of spike-timing-dependent plasticity (STDP) using the non-identical pulse method with different spike widths. Each colored bar shows the average of 3–5 measurements. (e) Emphasizes potentiation direction of STDP with positive delta time (45° tilted). (f) Emphasizes depression direction of STDP with negative delta time (225° tilted). The definition of conductance change is as \(\log_{10}(I/I_\text{initial})\), where \(I/I_\text{initial}\) is current ratio measured at 1 V after/before the pulse is applied. Figure reprinted by [19].
Figure 4c–f demonstrate that the SiO$_x$-based 1D-1R architecture can mimic spike-timing-dependent plasticity (STDP), a biological process that adjusts the strength of connections between two neurons in a synapse gap junction region that is an electrically conductive link between the pre- and postsynaptic neurons. Two pulse generator sources are used to simulate the pre- and postsynaptic neurons. This provides the pulse waveforms using the nonidentical pulse method (also used in various types of emerging memory devices or materials systems) for demonstration of STDP. By design of pre-neuron and postneuron spikes in neuromorphic circuits, the strength of the conductance change can be modulated based on the spike-timing delta ($\Delta t$) between the two neurons (Figure 4c–d). Figure 4e–f demonstrates a total of 10 different states of STDP biological behavior for depression and potentiation with $n = 2, 4, 6, 8, 10$ and as a function of spike width modulation, ranging from 100 ns to 1 ms. For example, the depression of conductance change strength can be achieved by using multistep spike heights from −4 to 0 V in the preneuron state and a single spike height fixed at 13 V in the postneuron state, with both neurons having a fixed pulse width of 10 μs and a firing period of 20 μs, as shown in Figure 4e–f. When the time delay difference is $-10 \times (n-1)$ μs, where $n$ is an even number, the total spike waveform (postneuron spike minus preneuron spike) applied to the synapse gap junction region can adjust the conductance ratio between two neurons over the range from $10^{-3}$ to 0.1 in the depression direction (RESET process) as compared with the initial LRS conductance (Figure 4f). Similarly, the potentiation of conductance change strength can be achieved by using multistep spike heights from 4 to 8 V in the preneuron state and a single spike height also fixed at 13 V in the postneuron state, with both neurons having a fixed pulse width of 10 μs and a firing period of 20 μs. When the time delay difference is $10 \times (n-1)$ μs, where $n$ is an even number, the total spike waveform (postneuron spike minus preneuron spike) applied to the synapse gap junction region can in this case adjust the conductance ratio between neurons over the range from $10^2$ to 0.01 in the potentiation direction (SET process) as compared with the initial HRS conductance (Figure 4e). It may be noted that the 1D-1R architecture not only avoids sneak-path issues and lowers standby power consumption, but also helps to realize STDP behaviors. Without the 1D rectification characteristics in reverse-bias polarity, the above spiking forms cannot be implemented due to the unipolar nature of the 1R device, specifically in the potentiation behaviors under negative bias. In the 1R case, an applied voltage above the RESET threshold voltage (for example, −9 V) can trigger the RESET process and induce depression behaviors instead of potentiation behaviors. Also, for depression behaviors, when the time delay difference is smaller than the spiking width, the remaining 4 V spike height in this case would not fire the synapse toward a LRS in the depression direction (see Figure 3h). Therefore, by carefully designing the firing pulses between neurons in the neuromorphic circuit, a biological synapse behavior can be demonstrated with 1D-1R SiO$_x$-based resistive switching memories.

The 1D-1R architecture with SiO$_x$-based resistance switching devices and the structure of artificial neural networks map naturally onto hybrid CMOS/synapse circuits that can be designed on a single chip (Figure 5) to provide predictable results with an ultimate scaling potential of CMOS technology to the sub-10-nm level, which could possibly challenge the complexity and connectivity of the human brain.

The other topic is material implication operations by using the same device architecture in SiO$_x$-based memristor (Figures 1 and 5). Based on our recent reports, implication operation
(IMP) has been performed by two SiO$_2$ memristors and a 5.7 kΩ standalone resistor are configured as shown in Figure 6a. Furthermore, three memristors connected in the circuit shown in Figure 6b and two steps of IMP are required to perform a NAND operation. It may be noted
that the final logic value pNANDq is stored as the last value of memristor s, or s″ in Figure 6b. This row of three memristors, namely P, Q, S, can be expanded to a row consisting of more memristors all sharing the same load resistor. Implication operations can be performed on any two memristors in the row, as long as the rest of the memristors are kept unbiased. Since we are able to perform implication on one row, similarly, implication can be done on one column.

However, when we put multiple rows and columns together to form a crossbar array, several problems arise. The first issue is providing multiple voltage signals as well as a common load resistor to an arbitrary pair of memristor on the same row or column. Based on the crossbar RRAM structure, the bit-line/word-line selection transistor can serve as the common resistor. By varying the gate voltage bias of the select transistor, it can serve as an ON state switch, OFF state switch or a resistor with channel resistance of $R_{\text{Load}}$. From Figure 6, it is noted that four voltage signals ($V_P$, $V_Q$, $V_S$, and $V_{\text{Load}}$) are required during an implication operation, two different voltages along the same bit-line/word-line. Therefore, a total of four voltage lines, each connected to all NMOS select transistor, will provide voltage signals for implication operation.

The concepts are demonstrated by a $4 \times 4$ memristor crossbar array (Figure 7a and b) and a circuit with an $8 \times 8$ memristor crossbar array. In addition to 1D-1R device arrays (Figure 5a), the hybrid CMOS/1D-1R device architecture shown in Figure 5e has been successfully demonstrated as shown in Figure 5f by the I-V resistive switching plots. Using the NMOS/PMOS transistor also fabricated on the same chip (Figure 5e), an implication circuit is realized using two 1D1R memory elements and a transistor. In Figure 7a and b, the design is quite different from the RRAM crossbar array architecture, the circuit consists of two rows of bit select transistors for the same column of memristors, one on the top, and one on the bottom. Similarly, there are two column word select transistors for the same row of memristors, one on the left and one on the right. This redundancy ensures two distinctive voltage signals can be applied on any pair of memristors on the same bit line/word line. The implication voltages ($V_P$, $V_Q$, $V_S$)

Figure 7. (a) and (b) demonstrate $4 \times 4$ crossbar structure memristor arrays with select transistors to achieve a one-bit full adder function. The implication circuit performs (a) $M_{13}$ IMP $M_{33}$ and (b) $M_{11}$ IMP $M_{12}$ on a $4 \times 4$ crossbar structure memristor array. Blue arrows show the current flow directions, and red solid squares cover all OFF-state transistors. Voltage signals and memristor numbers are labeled.
are biased on three of the four lines depending on the configuration, and the voltage applied to each line is labeled as follows, $V_{\text{bit\_up}}$, $V_{\text{bit\_down}}$, $V_{\text{word\_left}}$, and $V_{\text{word\_right}}$. The gate of each select transistor is also independently biased, to either isolate the bit lines/word lines from the implication voltages or provide that implication voltage to one particular bit line/word line.

To perform $M_{13}$ IMP $M_{33}$ (negative voltage implication operation performed along bit line 3):

Assuming all memristors are initialized to HRS, bias $V_{\text{bit\_up}} = 2$ V, $V_{\text{word\_left}} = -1.5$ V, $V_{\text{word\_right}} = 0$ V, $V_{\text{word\_left}}$ and $V_{\text{word\_right}}$ with all other transistor gate voltages at $V_{\text{full\_off}}$. The equivalent circuit is shown in Figure 5a with the path of conduction current flow marked. In Figure 5a–e, P is $M_{1y}$, Q is $M_{3y}$, and the transistor $\text{bit\_3u}$ is used as the load resistor. The implication is a negative voltage scheme. All transistors that are biased at fully OFF states are covered by red squares, effectively keeping the voltages of irrelevant columns/rows floating; to perform $M_{11}$ IMP $M_{12}$ (positive voltage implication operation performed along word line 1):

Assuming all memristors are initialized to HRS, bias $V_{\text{bit\_up}} = 0$ V, $V_{\text{bit\_down}} = 1.5$ V, $V_{\text{word\_left}} = -2$ V, $V_{\text{bit\_1u}}$ and $V_{\text{bit\_2d}}$ with all other transistor gate voltages at $V_{\text{full\_off}}$. This equivalent circuit is shown in Figure 5b with the current flow path marked. In Figure 5b, P is $M_{1y}$, Q is $M_{3y}$, and the transistor word_1l is used as the load resistor. In this case, the implication uses a positive voltage scheme. As before, all transistors biased to fully OFF states (covered by red squares) effectively keep the voltages of irrelevant columns/rows floating.

In Figures 1a and 5a, each memristor is placed in series with a pn diode in order to avoid current sneak-path problems. Originating from the crossbar device structure itself, the sneak-path problem has been identified and analyzed by many previous researchers [20, 76–122]. Because the bit line or word line select transistor raises the voltage across the whole bit or word, a group of memristors in the LRS may form a highly conductive path and cause misreading of certain memristors. The solution to the sneak-path problem is using a selection element together with a memory element, as shown in Figure 5f. Such a selection element is used to allow current conduction in one direction while suppressing current flow in the other direction. The most common selection element is a low-leakage pn diode, limiting the current flowing through the sneak paths down to reverse the bias leakage current level and reducing the power consumption during implication operations.

5. Summary

In summary, we have demonstrated potentiation, depression and spike-timing-dependent plasticity in a synaptic device built using a SiO$_x$-based 1D-1R architecture. Proton-induced resistive switching behaviors in the SiO$_x$ memory element were discussed, where the SET threshold is modeled as proton desorption from the (SiH)$_2$ defect to generate the conductive hydrogen bridge, Si-H-Si, and the RESET transition is modeled as proton release and capture to reform nonconductive (SiH)$_2$ [82–89]. The electrical results demonstrate that the technology has good potential for providing a simple and robust approach for large-scale integration of programmable neuromorphic chips using CMOS technology, and represent a critical milestone regarding the potential use of SiO$_x$-based resistive memory as a synaptic device in future synthetic biological computing applications. Moreover, a logic circuit consisting of a
4 × 4 array of crossbar structure memristor 1D1R memory elements and select transistors are proposed together with bidirectional implication schemes. Then a one-bit full adder is theoretically realized with a total of 48 operation steps performed on the circuit. A comparison between CMOS-enabled logic circuits and memristor-enabled circuits shows advantages in real estate and power consumption, as well as disadvantages in speed. This result suggests the memristor-enabled logic circuit is most suitable for high-speed, low-power, high-density applications. Further study is still required to make a few steps in various implication schemes as well as lower power consumption in synaptic computations.

Author details

Cheng-Chih Hsieh1, Yao-Feng Chang1*, Ying-Chen Chen1, Xiaohan Wu1, Meiqi Guo1, Fei Zhou1, Sungjun Kim2, Burt Fowler1, Chih-Yang Lin3, Chih-Hung Pan3, Ting-Chang Chang3 and Jack C. Lee1

*Address all correspondence to: yfchang@utexas.edu

1 Microelectronics Research Center, the University of Texas at Austin, Austin, TX, USA
2 Inter-university Semiconductor Research Center (ISRC) and the Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea
3 Department of Physics, National Sun Yat-Sen University, Kaohsiung, Taiwan

References

[1] Lundstrom M. Science. 2003;299:210-211. DOI: 10.1126/science.1079567


