Memristive devices for computing

J. Joshua Yang1, Dmitri B. Strukov2 and Duncan R. Stewart3

Memristive devices are electrical resistance switches that can retain a state of internal resistance based on the history of applied voltage and current. These devices can store and process information, and offer several key performance characteristics that exceed conventional integrated circuit technology. An important class of memristive devices are two-terminal resistance switches that exceed conventional integrated circuit technology. These devices were originally conceived in the late 1960s and recent progress has led to fast, low-energy, high-endurance devices that can be scaled down to less than 10 nm and stacked in three dimensions. However, the underlying device mechanisms remain unclear, which is a significant barrier to their widespread application. Here, we review recent progress in the development and understanding of memristive devices. We also examine the performance requirements for computing with memristive devices and detail how the outstanding challenges could be met.

The search for new computing technologies is driven by the continuing demand for improved computing performance, but to be of use a new technology must be scalable and capable. Memristor or memristive nanodevices seem to fulfil these requirements — they can be scaled down to less than 10 nm and offer fast, non-volatile, low-energy electrical switching. Memristors are two-terminal ‘memory resistors’ that retain internal resistance state according to the history of applied voltage and current. They are simple passive circuit elements, but their function cannot be replicated by any combination of fundamental resistors, capacitors and inductors2,3. Moreover, their microscopically modified internal state is easily measured as an external two-terminal resistance. Memristors were originally defined as components that linked charge and magnetic flux, but they can be more usefully described as devices with a pinched-hysteresis loop whose size is frequency dependent4. The natural computing application for such devices is resistive random access memory (ReRAM), but their dynamical nonlinear switching also suggests that they could be used to develop alternative computer logic architectures.

Memristive switching behaviour can be traced back two centuries4, but its theoretical inception came only 40 years ago1,2 and the link between theory and experiment was only established in 20085. In the 1960s, advances in thin-film technology allowed very high electric fields in ultrathin metal/oxide/metal films to be obtained and memristive behaviour was pronounced enough to be observed6. However, after a decade of intensive study7, research efforts into these devices faded, which is perhaps not surprising given the rise of silicon integrated circuit technology. Interest remained low until the late 1990s, when a gradual reduction in the progress of silicon technologies led to renewed interest in alternative switching devices.

The successful commercialization of any application demands a robust and predictive understanding of its underlying mechanisms. For memristors, understanding their electrical switching has been limited by the chemical, stochastic and localized nature of the effects. However, in the past decade advances in the growth and characterization of nanoscale materials has led to significant progress in the understanding and optimization of microscopic memristive switching. These developments have recently been covered by a variety of excellent review articles8–19. Here, we focus on the chemical and physical mechanisms of memristive devices, and try to identify the key issues that impede the commercialization of memristors as computer memory and logic.

Memristive materials and nanoscale devices

Memristive devices can be classified based on switching mechanism, switching phenomena or switching materials. Here we loosely group all ionic switching devices into two categories — anion devices and cation devices — to simplify our discussion of their mechanisms.

Anion devices. The switching materials of anion-based devices include oxide insulators, such as transition metal oxides, complex oxides, large bandgap dielectrics, and some non-oxide insulators, such as nitrides and chalcogenides. In most metal oxides, for example, the oxygen anion (or equivalently the positive-charged oxygen vacancy) is thought to be the mobile species. Anion motion then leads to valence changes of the metal (cations), which causes the resistance change of the metal oxide material. Therefore, these devices are also called valence change memories8. Resistance switching can originate from a variety of defects that alter electron transport rather than a specific electronic structure of insulating materials, and consequently almost all insulating oxides exhibit resistance switching behaviour. In principle, it should also be possible to observe this switching in most other insulating compound materials, such as halides, borides, carbides and phosphides. The key to obtaining controllable switching seems to be engineering structural and/or chemical defects into the material either through fabrication processes or electrical operations.

Since the early demonstration of resistance switching phenomenon in oxides around 50 years ago6, oxides have been extensively studied as anion-based switching materials20–26, including MgO, TiO2, ZrO2, HfO2, VO2, NbOx, TaOx, CrO2, MoO3, WOx, MnO2, FeOx, CoOx, NiOx, CuOx, ZnxOy, AIOx, GaOx, SiOx, SiOxNy, GeOx, SnOx, BiOx, SbOx, oxides of rare-earth metals including Y, Ce, Sm, Gd, Eu, Pr, Er, Dy and Nd, and perovskites (SrTiO3, Ba0.6Sr0.4TiO3, SrZrO3, BiFeO3, Pr0.7Ca0.3MnO3, La0.7Sr0.3FeO3, PrLa0.65Ca0.35MnO3). Among the even larger class of non-oxide insulators, a few examples of switching materials have been demonstrated in nitrides (for example, AlN),

1Hewlett-Packard Laboratories, Palo Alto, California 94304, USA, 2Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, California 93106, USA, 3Steacie Laboratories, National Research Council of Canada, Ottawa K1A 0R6, Canada.

E-mail: jianhuay@hp.com; strukov@ece.ucsb.edu; duncan.stewart@nrc.ca
tellurides (for example, ZnTe) and selenides (for example, ZnSe). References and some detailed information of these switching materials are given in Supplementary Table S1.

The switching mode can be bipolar or unipolar (also called nonpolar when devices can be operated with both voltage polarities). Bipolar switching requires opposite voltage polarities to be used for switching ON (set) and OFF (reset) respectively, whereas nonpolar and unipolar switching has no such requirement. Most of the above materials can switch in both nonpolar and bipolar modes27,28 depending principally on the structure of the device, and in particular the asymmetry obtained through device fabrication or electrical operations.

**Switching mechanisms.** Caution needs to be taken when identifying the exact switching material and mechanism. Many chemical interactions29,30 are possible inside a device under a high electric field and Joule heating, and the entire material stack should be taken into consideration. For example, the electrode31 or even the substrate material32,33 could be responsible for the observed switching phenomena. In most cases, a new material phase is formed in the insulating material during the initial electroforming processes31,34,36 or device fabrication, and this new phase accounts for the switching. Therefore, identifying the actual switching material is a challenging but crucial first step towards understanding the switching mechanism. The specific questions that need to be addressed are: what are the mobile species, and where (location), why (driving force) and how (microscopic picture) are they moving under electrical excitations?

Insulating oxides can also be viewed as semiconductors with native dopants, resulting from oxygen deficiency (n-type, for example, TiO2-x; refs 37,38) or oxygen excess (p-type, for example, Co2+,O; refs 39,40). The thermally and/or electrically activated motion of these native dopants results in chemical changes in the oxide, such as valence state change, leading to resistance switching. The mobility and concentration of oxygen vacancies and cation interstitials are sufficiently high in the temperature range in question31,34, especially in transition metal oxides. Therefore, it is generally believed that they are the mobile species responsible for switching, and this seems to be supported by experimental evidence32,38,42–44. However, more direct evidence would still be needed to clarify the identity of the mobile species in the representative switching oxides. Experiments using isotope tracers should be particularly useful45. In addition to the native dopants, other impurities, such as hydrogen, have been suggested to play a role in switching46,47, and this seems likely in cation devices48.

The switching region in a micrometre-size device or larger is typically a localized conduction channel tens or hundreds of nanometres in diameter48,44. Switches with a laterally (that is, parallel to the electrode surface) uniform switching region typically have a shorter retention and lower switching speed. The switching channel is usually created in the electroforming process31,34,36 in which the virgin device is preconditioned to a switchable state by applying a higher-than-usual voltage or current. For a symmetric bipolar device, the switching polarity is usually determined by the voltage polarity of this electroforming process39,46,48,50. In contrast, for an asymmetric bipolar device, the switching polarity is often governed by the asymmetry of the device stack rather than the polarity of electroforming46. Even interface switching (as opposed to bulk switching) is localized to a small area of the interface57,50,51.

The important role of the interface in the switching is reflected by a commonly observed rectifying current–voltage (I–V) relation in the high-resistance state (OFF state). In most cases, one interface of a bipolar device is essentially ohmic with a large density of dopants and the other interface is more resistive with fewer dopants. As a voltage divider, most of the applied voltage drops on the resistive interface to force switching. Only a small fraction of the applied voltage is dropped on the highly conductive ohmic-like interface, which always remains at low resistance. The rectification orientation of the I–V curves could be used to identify the interface that blocks the current and this interface is usually the switching interface52. However, it is not uncommon that two opposite switching polarities coexist in the same device53,54, which could be rationalized by considering that both interfaces might be switchable under a certain circumstance and they switch to the opposite resistance states because they always see the opposite electric fields53,55. A family of nanodevices have been demonstrated in simple metal/oxide/metal structures with rich switching behaviours by tailoring the switching properties of the two interfaces. Nonpolar devices can switch either at the interface region56–58 or in the interior of the switching film14.

An electrical bias generates two main effects in a switching material: an electric field and Joule heating. Joule heating is expected given the high current densities (of the order of >108 A cm−2) typically required for switching in anion-based memristors and has been confirmed in both bipolar and nonpolar devices49,52,59,60. Electric field and Joule heating generally coexist in all memristive switching, although their relative importance varies depending on the device stack, materials, electrical operation history and more. This leads to four main classes of switching commonly observed in oxide-based switches: bipolar nonlinear, bipolar linear, nonpolar bistable and nonpolar threshold switching. The four classes are shown schematically in Fig. 1a–d. In all cases there are four main driving forces that work independently or together to influence atomic motion or rearrangement: electric potential gradient (field), electron kinetic energy, species concentration gradient and temperature gradient (Fig. 1e–h).

The microscopic picture of how exactly these factors drive the mobile species to actuate a particular type of switching is still under debate. Therefore, experiments that could visualize the switching in real time and at nanoscale resolution, and with chemical and/or structural information would be extremely valuable, such as those using in situ transmission electron microscopy64,65 and in situ scanning transmission X-ray microscopy66.

A high electric field gradient can move charged dopants, as illustrated in Fig. 1e. High-speed electrons under a high electric field can also move atoms by momentum transfer or ‘electron wind’ (Fig. 1f), which has been demonstrated to reversibly switch the resistance of a Au nanowire6. Both electric field and electron kinetic energy are polarity dependent and can cause bipolar switching. The role of Joule heating is more complicated and not yet fully understood. High temperatures from Joule heating significantly enhance drift (Fig. 1e) and diffusion (Fig. 1g). Moreover, the high temperature is localized around the conduction channel, which generates a high temperature gradient laterally. Similar to an electric potential gradient or an element concentration gradient, a high temperature gradient itself might also induce a net mass transfer of atoms67 (Fig. 1h). Drift and electromigration move dopants vertically along the conduction channel, while the temperature gradient might contribute to lateral motion of dopants. Diffusion can happen in both ways, because element concentration gradients exist both vertically (within the channel) and laterally (into and out of the channel). Finally, for each microscopic device state with a particular arrangement of impurities, defects or dopants, the I–V characteristic is determined by the available electron conduction mechanisms, as summarized in Fig. 1i for n-type materials. This I–V behaviour is the key device property for applications and also provides crucial information to decipher the switching mechanism.

Generally speaking, the switching tends to be bipolar if the electric field plays a significant role and nonpolar if thermal effects are dominant. Figure 1a presents schematically a device driven by an electric field. The growth and retraction of the conduction channel vertically under the electric field (drift or electromigration) in the interface region results in the typical switching loop (inset in Fig. 1a), where the rectifying I–V curve in the OFF state and the symmetric nonlinear I–V curve in the ON state reflect a Schottky-like barrier and a residual tunnel barrier in the OFF and ON states, respectively.
Figure 1 | The driving forces, electrical characteristics, transport mechanisms of ions and electrons for the switching of anion-based devices. a–d, Simplified schematics of conduction channels (red) in switching matrix materials (blue) in four typical switching devices, where both electric field and Joule heating drive the switching. Switches with a laterally uniform switching region across the entire device area are not included here. The grey arrows indicate the idealized ionic motion. Different from e, the channel in d usually completely disappears in the high-resistance state and may or may not (for example, Mott transition) involve ionic motion. Inset to each schematic shows switching current–voltage loops typical of bipolar nonlinear switching (a), bipolar linear switching (b), nonpolar nonvolatile switching (c) and nonpolar threshold switching (d). e–h, Schematic illustration of the factors that influence oxygen anion motion for drift (electric potential gradient, e), electromigration (electron kinetic energy, f), Fick diffusion (concentration gradient, g), and thermophoresis (temperature gradient, h). Any of the four factors shown in e–h may contribute with others simultaneously to produce switching of type a–d, which are idealized limiting cases; real-world switching is usually combinations of these. I, Schematic of the possible electron transport paths in the devices. (1) Schottky emission: thermally activated electrons injected over the barrier into the conduction band. (2) Fowler–Nordheim tunnelling: electrons tunnel from the cathode into the conduction band; usually occurs at high electric field. (3) Direct tunnelling: electrons tunnel from cathode to anode directly; only when the oxide is thin enough. When the insulator has localized states (traps) caused by disorder, off-stoichiometry or impurities, trap-assisted transport contributes to additional conduction, including the following steps: (4) tunnelling from cathode to traps; (5) emission from traps to the conduction band (Poole–Frenkel emission); (6) tunnelling from trap to conduction band; (7) trap-to-trap hopping or tunnelling, ranging from Mott hopping between localized states to metallic conduction through extended states (see Fig. 2b); and (8) tunnelling from traps to anode: $E_F$, Fermi energy level; $E_C$, valence band; $E_v$, conduction band; $E_s$, Schottky barrier height; $E_b$, trap barrier height. Panel i reproduced with permission from ref. 16,© 2012 IEEE.

Figure 1b shows another type of bipolar switching, which has a linear I–V curve in both the ON and OFF states. In this class of switching, there is a conduction channel bridging the top and bottom electrodes all the time in the entire switching cycle. The resistance change originates from the change in composition, volume or geometry inside the channel, which is a result of the combined effect of the electric-field-induced vertical drift and the thermally enhanced lateral diffusion.

An increasing role of the thermal effect leads to nonpolar switching (Fig. 1c) where the same voltage polarity can be used to set and reset the device. The physics of nonpolar switching is still controversial. On the one hand, the temperature gradient shown in Fig. 1h might, for example, attract oxygen vacancies to the conduction channel region, switching the device ON. On the other hand, the set switching initially resembles ‘soft breakdown’ of dielectrics under an electric field. It is likely, it is a combined two-step process, where a purely electronic effect (breakdown) leads to a high current and then heat-assisted ionic motion follows. The reset switching of the nonpolar device is normally described as a thermal rupture (fuse) of the conduction channel, possibly caused by thermal diffusion driven by concentration gradient and/or reduction of free surface energy of the filament or even a phase-change process induced by heat and/or electric field.

In contrast, the switching shown in Fig. 1d is the threshold switching, which is essentially a volatile switching with a hysteresis loop. With increasing current, the insulating device suddenly becomes metallic at a certain current level accompanying a steep current increase in the I–V plot. However, after reducing the current level, the device becomes insulating again. Threshold switching has been observed in NiO$_2$ when its electrode is thin, and attributed to a spontaneous rupture of filaments at high temperature due to an insufficient heat dissipation. Threshold switching has been more commonly observed in metal–insulator–transition materials, such as VO$_2$, NbO$_3$ and Ti$_2$O$_3$, where negative differential resistance (NDR) phenomenon is frequently seen. A localized high temperature from Joule heating converts the heated region of a metal/insulator/transition material from insulating to metallic via mechanisms still under debate, giving rise to the abrupt current increase. Reducing the current decreases the local temperature to below the metal/insulator/transition temperature, recovering the insulating state. Therefore, this type of switching (Fig. 1d) is viewed as a purely thermal effect.

Most switching phenomena are thus a result of both thermal and electric field effects — indeed, a purely electric-field-induced switching is yet to be unambiguously demonstrated in anion-based devices. As the device resistance changes during switching (and with it the voltage and current), even the relative contributions of the coupled thermal and electric field effects are dynamically changing.

Material selection criteria. Joule heating is essentially unavoidable in these devices, and this has a major impact on material selection. To
Figure 2 | Material selection criteria for high endurance and repeatability. a, Simplified schematic phase diagram of a metal–oxygen (Me–O) system with only two solid-state phases at low temperature. The MeO$_x$ phase is an insulating stoichiometric phase, serving as the matrix material in the switching device illustrated schematically in Fig. 1. The (Me) phase is a metal–oxygen solid solution, serving as the conduction channel. These two phases are thermodynamically stable with each other and do not mix by reaction to form an intermediate phase even at high temperature, for example, locally induced by Joule heating. The metal (Me) has a large solubility of oxygen, readily accommodating mobile oxygen anions or vacancies during switching. b, With increasing oxygen content in the channel region, the electron transport mechanism changes, producing corresponding resistance changes. The schematic depicts a sequence of evolving conduction centre density and conduction mechanisms (top), and corresponding photoemission core-level (CL, middle) and valence-band (VB, bottom) measurements, from a disordered transition metal oxide in the course of oxidation. The red circles are the localization radii. Unoxidized metallic state (i), weakly localized variable range hopping (VRH) regime (ii), more strongly localized nearest-neighbour hopping (NNH, iii), strong localization regime on the verge of percolation breakdown (iv), and final highly insulating sub-percolation insulating state (v). $X$ is the fraction of conduction centre sites and $X_{cr}$ is the critical fraction at the percolation threshold. $E_F$, Fermi energy. Panel b reproduced with permission from ref. 73, © 2012 Springer.

form a resistance switching system, both conductive and insulating phases are required. Reliable switching requires that these two phases do not react with each other chemically to form a new phase, even at the high temperatures induced by Joule heating. This would require a simple material system, such as those with the phase diagram shown in Fig. 2a. In this system there are only two thermodynamically stable solid-state phases, meaning that they will not react with each other to form a new phase even at high temperature. The MeO$_x$ phase is the insulating phase and the (Me) phase is relatively conductive, serving as the conduction channel. What distinguishes any (Me) phase as an ideal conduction channel material is a large solubility of oxygen, which allows the channel to readily accommodate mobile species in and out without losing them after many switching cycles. A very large amount of oxygen can be accommodated in the nanoscale amorphous material (Me), especially under a fast quenching process during switching. Changes in the volume, geometry or composition of the channel phase can cause resistance switching.

Figure 2b illustrates a resistance change from metallic to insulating in a disordered oxide that can be attributed to changes in conduction-centre concentration and thus hopping transport mechanisms. Multilevel cell or even analogue switching behaviour can be readily obtained with such conduction channels. Reliable switching up to a trillion cycles has been demonstrated in TaO$_x$-based devices and endurance close to ten billion switching cycles has been demonstrated using the similar Hf–O system. Based on the above criteria of two stable phases and a large solubility, other oxide systems, such as Er–O and Y–O, may be expected to have a similar electrical performance.

Cation devices. Cation-based devices are often called electrochemical metallization memory, conductive bridging RAM, programmable metallization cells or atomic switches. They were first reported in the 1970s and have been further developed for memory applications since the late 1990s. In most cases, the mobile species is believed to be the metallic cation. The physical device stack is similar to that of anion-based devices, being an electrode/insulator/electrode trilayer.

A signature of the cation-based devices is to have an electrode made from (or the insulator doped with) an electrochemically active material, such as Cu (ref. 79), Ag (ref. 80) or an alloy of these metals (CuTe, for example). Devices without this signature are discussed as anion-based devices in this Review, although their mobile species might be cation interstitials, such as with Ni$_3$O devices. The counter electrode is usually an electrochemically inert metallic material, such as W, Pt, Au, Mo, Co, Cr, Ru, Ir, doped poly-Si, TiW or TaN (ref. 12). The insulating materials have traditionally been electrolytes, including sulphides, As$_x$S$_{3–x}$, CuS, ZnCd$_x$S, iodides (AgI, RbAg$_5$I$_4$), tellurides (Ge–Sb–Te), ternary chalcogenides (Ge–Sb–Te) and water. Other materials have also been studied for this purpose, such as methylsilsequioxane, doped organic semiconductors, amorphous Si, C and even vacuum gaps. More and more insulating oxides or nitrides have recently been used, including Ta$_2$O$_5$, SiO$_2$, HfO$_2$, WO$_3$, MoO$_3$, ZrO$_2$, SrTiO$_3$, TiO$_x$, CuO$_2$, ZnO, AlO$_2$, GdO$_2$ and AlN. References and some detailed information of these switching materials are given in Supplementary Table S2. Changing from traditional electrolytes to oxide materials increases the switching voltages from below 0.3 V to above the operating voltage of CMOS (complementary metal–oxide–semiconductor) devices, making them suitable for some special applications, such as non-volatile switches in large-scale integrated circuits. Retention may also be improved using oxides, which are usually inexpensive and CMOS compatible.
Cation switching mechanisms strongly resemble those of anion-based devices, and can also be roughly described by the schematics in Fig. 1. Electroforming in cation-based devices causes structural changes to the electrolyte and forms nanoscale channels that persist and serve as the template to host the electrochemically active metal atoms for subsequent switching17. Most of the cation-based switches are bipolar, suggesting that the dominating effect is the electric field (Fig. 1a,b). Unipolar switching shown in Fig. 1c has also been demonstrated in cation-based devices18,19, substantiating a possible role of Joule heating. The switching mechanism is relatively well understood with some electrolytes20,21,22,23, such as a Ag/H2O/Pt device20 (Fig. 3): during electroforming or set switching, a positive high voltage on the Ag electrode oxidizes the electrode atoms into Ag+ cations, which are dissolved into the electrolyte. These cations drift across the electrolyte to the inert counter electrode (Pt, cathode) under a high electric field; then they are electrochemically reduced to Ag atoms at the cathode and deposited on the surface of the cathode electrode, which allows the Ag atoms to grow towards the anode (Ag electrode) as a filament(s), reaching the Ag electrode and switching the device ON. Any of these steps could be the speed-limiting process for the entire switching event, depending on the material systems, in particular the insulating materials. An opposite voltage (positive) is then applied to the electrochemically inert Pt electrode for reset switching, which anodically dissolves the Ag filament (the virtual anode now) starting from the interface of the Ag electrode/Ag filament, and switches the device OFF. In this picture, the electric field is the only driving force and Joule heating is negligible given the small current usually involved in these devices.

However, more research is still needed to clarify the microscopic picture of switching in cation devices, especially for those with non-traditional electrolyte materials, such as some oxides or amorphous Si (refs 12,64,65,94). These materials are usually not good ionic conductors and require a higher voltage (thus a high electric field and substantial heating) for switching. A temperature-dependent switching study in a Cu/Ta2O5/Pt system suggests that the reset switching is likely to be thermal-diffusion assisted (similar to in Fig. 1b). Furthermore, some experimental results have shown that the conduction channel24,25 has an opposite geometry to that in Fig. 3, suggesting a possibility that the switching occurs at the interface between the inert electrode and the conduction channel in some cases. The reason for the observed discrepancy among different electrolytes remains unclear. It is worth noting that the conduction channel(s) might be composed of nano-islands rather than a continuous filament, especially in the OFF state. This possibility is supported by the observation of a Coulomb blockade effect at low temperature24 and also by some direct imaging of the filament25.

Physical switches. Both the cation- and anion-based devices are chemical switches because chemical reactions (redox) are involved in the switching mechanisms. There are also various switching phenomena where only physical changes are involved and some of these may also belong to generalized memristive devices26. There are, for example, electronic, magnetic, ferroelectric and microstructural processes, which lead to electronic resistance switches, magnetic tunnel junctions27, ferroresistive switches28,29 and phase-change switches30,31, respectively. Among them, magnetic tunnel junctions and phase-change switches have been the most intensively studied.

In electronic switches, resistance change can stem from charge trapping (de-trapping) at an electrode/insulator interface to increase (decrease) the contact barrier32, or inside a disordered thin film (for example, Pt-doped SiO2)33 to increase or decrease the degree of Anderson localization. Electronic switches have also been demonstrated in metal-doped polymers and are attributed to charge transfer between the metal nanoparticles and the organic materials34,35,36. Another type of electronic resistance switch relies purely on an electronic phase change such as the Mott transition, which applies to systems (for example, Ga2O3)37 that should be metals according to band-structure theory but turn out to be insulators due to strong electronic Coulomb interactions below a certain critical density of electrons, as determined by electric or magnetic fields, pressure, or carrier doping38,39. A ferroelectric tunnel junction, an example of ferroresistive switches40, utilizes an ultrathin ferroelectric material (for example, BaTiO3)41 as a tunnel barrier in a tunnel junction, where electric-field-induced polarization reversal in the ferroelectric material changes the interface transmission, thus the tunnelling current and the device resistance. These device concepts seem promising, but experimental data need to be carefully examined as different switching mechanisms, such as the chemical switching discussed before, are also possible in these devices.

Other device geometries. In addition to the most commonly used two-terminal, vertical-stack and crosspoint devices discussed above, other device geometries have also been explored, such as three-terminal and lateral (planar) two-terminal devices. Three-terminal resistance switches, which resemble a transistor, utilize a third electrode as a gate to control the formation and annihilation of the conduction channel by controlling the drift of dopants. This allows the signal and control lines to be separated. Functioning three-terminal devices have been demonstrated both in cation42 and anion-based switches43,44, which are related to the 1960 ‘memistor’45. A lateral device has two electrodes on the same plane separated by a small gap, which is filled with a switching material46. This device configuration is mainly used for research because it allows easy access to the switching region for material characterization.

Figure 3 | Switching of a traditional cation-based device (electrochemical metallization memory) with the cell stack of Pt/H2O/Ag. a, Scanning electron microscopy (SEM) image showing the high-resistance state of the device with shorter and smaller Ag dendrites. b, SEM image of the same cell at low-resistance state with longer and larger Ag dendrites, obtained by applying a positive voltage on the Ag electrode side. Figure reproduced with permission from ref. 93, © 2007 AIP.
narrow but very thick metal electrodes, such as those synthesized by a Cu damascene technique, are needed to reduce the electrode resistance in a crossbar array.

Nonlinear switching dynamics. Switching dynamics reveal much about the nature of the resistance switching and hence deserve detailed attention. For example, Fig. 4a shows typical experimental data for switching dynamics in TiO$_2$-based devices. In particular, no change to the memristor state occurs for small biases applied, even over long time scales. On the other hand, the state changes abruptly (potentially super-exponential) on further increase of the voltage bias. Note that there is no well defined or ‘sharp’ switching threshold voltage for memristive devices.

For TiO$_2$-based memristive devices that rely on modulation of the oxygen vacancy concentration profile, such switching dynamics are explained by nonlinear ionic transport (Fig. 4b). The vacancy motion in solids is via thermally activated hopping of oxygen ions and can be crudely approximated with hopping in the net potential of the constituent ions (Fig. 4c). (Such a model is best suited for ionic solids but can also be applied to mixed ionic-covalent and purely covalent solids with some corrections, for example, to electron-electron interactions.) The hopping rate, and hence the ionic drift velocity and current, depends exponentially on both the applied electric field and local temperature, $T$. Under normal conditions, deformation of the lattice potential energy by an external electric field $E < 10^8$ V cm$^{-1}$ is typically much less than the activation energy for ion hopping $U_A$ (with $U_A > 1$ eV for practical cases), so that a familiar linear approximation for ionic velocity $v = \mu E$ (where $\mu$ is ionic mobility) is used to describe ionic drift. Memristive devices operate in a different regime: much larger electric fields $E = V/d >> 10^8$ V cm$^{-1}$ consistent with nanoscale film thicknesses $d$ and few-volts applied biases $V$ produce exponential ionic transport $v = \mu E \exp[ET/E_0]$, where $E_0$ is a characteristic field for a particular mobile atom. Joule heating also exponentially enhances ionic drift and diffusion, and super-exponential dependence of the switching dynamics on the applied voltage might take place due to super-linear (exponential) dependence of the current (and hence dissipated power and temperature) on the applied voltage itself, for example, $v(V) \propto \exp[\exp(V)]$.

Similar nonlinear switching dynamics are observed in cation-based memristive devices, although the underlying detailed mechanisms may be different. In addition to hopping of the ions, the motion of the active electrode ion involves two redox reactions, that is, electron transfers, at the two interfaces (Fig. 4b). The reduction mechanism may be different in some cases.

Nonlinearity in ionic transport is crucial for simultaneously realizing faster switching speeds and longer retention times in memristive devices. For example, for a system with $U_A = 1$ eV, ionic diffusion is enhanced 12 orders of magnitude from $D_{\text{diff}} \approx 4 \times 10^{-24}$ cm$^2$ s$^{-1}$ (at room temperature) to $D_{\text{diff}} \approx 5 \times 10^{-22}$ cm$^2$ s$^{-1}$ on Joule heating by 500 K. Assuming that ions need to move only a few nanometres, say $d = 5$ nm, to cause significant change in resistance state, $E = 5 \times 10^8$ V cm$^{-1}$ would correspond to a switching speed of about $t \approx 70$ ns (ref. 111). The retention time is determined by the rate of ion

Figure 4 | Nonlinear switching dynamics. a-c. Typical switching dynamics for an anion (TiO$_2$-) device characterized by a voltage pulse stress (a, top) with variable pulse duration and amplitude. In particular, panel a (bottom) shows 16 curves, that is, 8 each for set (green) and reset (blue), with each curve showing evolution of the normalized resistance ($R$, measured at specific bias) for the device that is initially set to the OFF (ON) state and then continuously switched to the ON (OFF) by voltage pulses with fixed amplitude and exponentially increasing duration. Such dynamics can be attributed to an exponential dependence of the ion transport (b) on the electric field and/or local temperature. The other mechanisms for nonlinear dynamics are due to an exponentially enhanced redox reaction (c) happening at the electrode/solid-state electrolyte interface (b) for cation-based devices and/or phase transitions (for example, Magnéli phases) affecting activation energy. $U_A$ is the activation energy for ion hopping, $E$ is electric field, $q$ is elementary charge, and $k_B$ is the Boltzmann constant. Panel a reproduced with permission from ref. 110, © 2012 IOP.
diffusion back towards a thermodynamic equilibrium ionic profile and, in this example, can be approximated as $t \approx d^2/D_{\text{diff}} = 75$ days, where the diffusion (retention) is assumed at room temperature. Retention times might be shorter when taking into account built-in internal electric fields and elevated temperatures due to Joule heating or thermal crosstalk in memristive circuits\(^{111}\). Additionally, the activation energy might need a correction when the ion movement is followed by chemical reaction or phase transition, for example, to a Magnéli phase in TiO\(_2\) (Fig. 4c)\(^{38,44}\), which might even improve the retention. One important observation from this analysis is that speed and retention are related properties that are usually traded off against each other. Engineering the devices with a smaller activation energy for ion hopping or redox reaction will improve the switching speed but it will also reduce the retention time.

High stress conditions (electric fields and/or elevated temperatures) seem crucial for nonlinear ionic transport, but they can be detrimental to other performance characteristics of memristive devices. For example, electrode melting, penetration into the insulating matrix, and permanent shorting of the device have been identified as some of the principal failure mechanisms limiting switching endurance of the memristors. Clearly, these failure mechanisms and others, for example, electromigration in electrodes and electron trapping-based oxide degradation, are enhanced (exponentially) by high stress conditions. Furthermore, strong nonlinearity in ion transport may also lead to larger dispersion in switching dynamics. Identifying and engineering nonlinear ion transport mechanisms that do not impact endurance and variations in the memristive devices is therefore an important goal\(^{118}\).

**Prospective applications**

Research activity in resistance switching has been primarily driven by the search for an ideal memory device. Indeed, hybrid CMOS/memristor circuits (Fig. 5 and 6a), and in particular those with the passive crossbar architecture (Fig. 6), could potentially combine all the desired properties of ‘universal memory’ — high speed, low energy and high endurance of static random access memories, and high density, low cost and non-volatility of flash memories (Table 1)\(^{119}\). Although many research efforts remain focused on single memristive devices, demonstrations of larger crossbar circuits are increasingly common\(^{20,121}\).

Logic applications of memristors are also being explored. Hybrid CMOS/memristor circuits (Fig. 5a,b) might eliminate the main inefficiency of field programmable gate arrays (FPGAs) — namely the large overhead associated with storing the circuit configuration information in local memory. In traditional FPGAs configuration memory is implemented as static random access memory or flash memory and typically consumes a large area of the chip — on the order of 50–90% (ref. 122). In hybrid FPGA circuits, configuration information could be stored in nanoscale memristors to improve density by more than 10 or 100 times for conservative or aggressive architectures as compared with conventional circuits\(^{123–126}\) (Fig. 6d–f). The conservative version of a hybrid FPGA has recently been demonstrated with nano-imprinted 100-nm-scale TiO\(_2\) memristor devices integrated with 0.5-µm CMOS technology\(^{127}\). Memristors have also been proposed to implement material implication logic\(^{128}\), in which Boolean logic states are stored as resistance states of the devices rather than voltage levels as in conventional logic. More research on circuits is needed to understand and optimize these hybrid circuits for all of these Boolean logic applications.

The idea of using resistance switching devices in artificial neural networks and for mixed signal computing in general (Fig. 6g–i) has a long history and can be traced back to at least the 1960s. This early work, however, was largely overlooked, principally because of the emergence of the first successful digital microprocessor just a few years later, which quickly became the dominant platform for computing. Some significant developments in this context were thin-film tungsten oxide and a-Si:H memristive devices\(^ {129}\), but it seems that the results were not encouraging enough to motivate continued investigation. More recently, the topic has been revived by publications on titanium dioxide memristors\(^ {37}\), and demonstrations of spike-time-dependent plasticity\(^ {130–132}\), short- and long-term potentiation in resistive switching devices\(^ {133}\), and hybrid circuits performing analogue dot-product computation\(^ {134}\). Motivation for the development of artificial neural networks comes from the fact that mammalian brains remain much more efficient than conventional Boolean machines for many computational tasks such as pattern recognition and classification, despite the exponential progress in CMOS performance. The structure of the artificial neural networks map naturally onto hybrid CMOS/memristor circuits and crude estimates have suggested that circuits with ultimately scaled CMOS technology and sub-10-nm scale memristive devices could challenge the complexity and connectivity of the human brain\(^ {134}\).

In general, memristive devices combining all the desired properties would be beneficial for all such applications (Fig. 7). (The exception is ON-state resistance, which should be large enough to avoid voltage drop across crossbar wires, whose typical length depends on a particular application.) Because contemporary memristive devices are still far from that goal (partially due to immature fabrication technology), it is useful to analyse which performance metrics are more important for the various applications\(^ {6,7,10,11,128,129,130–134}\) (Fig. 7). In this context, targeted memory applications are subdivided into two categories. At the initial stage of development memristive memories are likely to compete and replace conventional storage technologies (which are slow, but very dense, cost effective and non-volatile) and, in the long run, reach the speed of fast memory technologies.

**Key challenges and outlook**

Cation- and anion-based devices have shown encouraging
**Figure 6 | Prospective applications of memristive circuits.** a–c. Digital memories and storage: in memories a resistance state represents one or more bits of information that are read by sensing current flowing through the device at a smaller non-perturbative bias. When combined in circuits, memory cells containing a memristor and a select device are organized in roughly square arrays with multiple cells sharing row (word) and column (bit) lines. The select device, implemented for example, with a transistor in a conventional 1-transistor/1-resistor (1T1R) architecture (a) enables unique access (read/write) to a particular row of cells at a time via bit lines by asserting the corresponding select word line. A much denser alternative to the 1T1R architecture without a dedicated transistor is shown in b, c (passive crossbar memory). In passive memory, select functionality is implemented via the diode-like I–V of the crosspoint device (shown in Fig. 5c). If the current is negligible for \( V < |V_{\text{th}}| \) then choosing a read voltage such that \( V_{\text{th}} < V_{\text{READ}} < 2V_{\text{th}} \) allows suppression of leakage current when sensing a particular cell (b). Strong nonlinearity in switching dynamics helps avoid another problem specific to 0-transistor/1-resistor (OT1R) architectures: disturbance of half-selected devices (c), which are typically biased at half of the write voltage \( V_{\text{WRITE}}/2 \) applied to switch a particular cell in the array. A number of biasing schemes as well as different approaches for current sensing have been proposed, see, for example, ref. 155 for more details. d. Programmable logic: in the conservative approach memristors serve as the configurable interconnect in FPGA-like circuits. A specific memristor would control whether two wires (for example, the input of some Boolean gates and output of another) are electrically connected. The operation of the FPGA consists of two stages: first, a particular computation task is mapped to the FPGA structure by setting connectivity of the gates via programming appropriate crosspoints to the ON state. During the second much longer stage, the connectivity pattern is fixed (that is, memristors do not change their state), and the FPGA runs a specific task with the programmed circuit. e, f. In a more aggressive approach, in addition to serving as a programmable interconnect memristors also implement part of the logic gate functionality. The I–V nonlinearity of the memristive devices is used to implement diode-resistor logic (f), whereas a CMOS subsystem is used for signal restoration and inversion. g–i. Bio-inspired and mixed-signal information processing: hybrid CMOS/memristor circuits may also enable efficient analogue dot-product computation, which is a key operation in artificial neural networks and many other information processing tasks. In the simplest model, artificial neural networks can be represented by a graph (g) with nodes corresponding to neurons and graph edges that correspond to synapses (d). In parallel, each node processes input information from the preceding nodes and then passes it to the next layer of nodes. h. A typical operation of the node would involve summation of input signals, with each scaled by the corresponding edge weights, followed by a specific threshold function of the node (not shown). In hybrid CMOS/memristor circuits memristive devices implement density-critical configurable analogue weights, crossbar wires serve as axons and dendrites, and CMOS is used for the summing amplifier to provide gain and signal restoration. As a result, individual voltages applied to memristors can be multiplied by the unique weight (conductance) of the memristor and summed up by a CMOS amplifier, thus implementing dot-product computation in analogue fashion. Here \( x_1, x_2, \) and \( y \) are input and output voltages for the considered synapses and neuron, correspondingly, whereas \( w_1 \) and \( w_2 \) are conductances of the corresponding memristive devices.
properties (Fig. 7). However, challenges remain, among which the device variability (or reproducibility) and $I$–$V$ nonlinearity are the main two. Furthermore, retention for cation-based devices and yield for anion-based devices remain to be demonstrated with nanodevices at the wafer scale. A SPICE (simulation program with integrated circuit emphasis) model and statistics of device parameters from a large number of devices and switching events are also needed for circuit designers.

Device variability. The variance from switching cycle to cycle and from device to device can be very large. For example, it has been seen experimentally that active conduction channels (filaments) can be inactivated and other inactive or even new ones can be activated during switching cycling of a single device. The materials selection criteria described above can lead to a stable channel–matrix system and thus greatly reduce the variance from cycle to cycle. The variance from device to device stems mainly from the random electroforming process that creates a different filament structure in each device. A high-voltage electroforming can be obviated by thinning the stoichiometric oxide film and adding a thick sub-oxide layer. However, this does not guarantee a similar channel in each device. Engineering control to reduce the randomness of channel creation is the key to reduce this variance, which can for instance be realized by fabricating uniformly distributed small precursors in the device for channel formation. Some experimental evidence has suggested that the problem is also partially resolved by scaling down the device to a size that is comparable to that of the filament and therefore has a limited active area.

**Current–voltage nonlinearity.** Using a transistor as the select device at each crosspoint in a crossbar array limits the footprint of each cell and its three-dimensional stacking capability, and would therefore not realize the potential scalability of memristive devices. A passive crossbar array utilizing the $I$–$V$ nonlinearity (different from the nonlinearity of switching dynamics mentioned above) of each cell is an alternative for ultrahigh density. Providing diode-like functionality (for example, Schottky or p–n junction) is useful for unipolar devices, but must be approached cautiously for bipolar type devices, as this limits heating and the electric field across the active device when attempting to write devices in the reverse bias configuration. A plausible solution is to use a special diode, such as a soft breakdown diode or a Zener diode, which should permit enough current density to switch the memristive device.

Two memristive devices can also be stacked anti-serially (head-to-head) to form a complementary resistance switch. Under a switching voltage pulse, these two devices see opposite voltage polarities and switch in opposite directions, thus limiting the current through the device pair as there is always one device in the OFF state in the pair. This device concept can also be realized in a single memristive device with two switchable interfaces. It is, however, difficult to operate the switches as multilevel cells or analogue devices with the scheme. A destructive read in complementary resistance switch devices is also much more demanding for device endurance.

The other option is to integrate a series metal/insulator/metal tunnel barrier, other S-type bistability or mixed ionic–electronic conductors, to suppress current at low voltage biases. In this case the select functionality is achieved with a half-biasing scheme as shown in Fig. 6b,c. A large $I$–$V$ nonlinearity can be engineered by integrating an oxide layer, such as one made of TiO$_2$, with a memristive TaO$_x$ layer. A single-layer TaO$_x$-based nanodevice normally exhibits a linear $I$–$V$ relation in the low-resistance state, meaning a high current at half of the switching voltage. However, for the bilayer oxide device the sneak path current is about two orders of magnitude smaller than that of the linear device. A detailed study suggests that a new NDR phase formed inside the bilayer oxide may be responsible for the observed $I$–$V$ nonlinearity. To realize a large OFF/ON resistance window in this case, a reading voltage large enough to trigger the threshold switching (volatilize) of the NDR element itself might be necessary when the memristive

---

**Table 1 | Comparison of memory and storage technologies.** Note that circuit-level overheads for the listed performance metrics are in general different among different device technologies and could often dominate individual device performance.

<table>
<thead>
<tr>
<th>Memristor</th>
<th>PCM</th>
<th>STTRAM</th>
<th>SRAM</th>
<th>DRAM</th>
<th>Flash (NAND)</th>
<th>HDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reciprocal density ($F^2$)</td>
<td>&lt;4</td>
<td>4–16</td>
<td>20–60</td>
<td>140</td>
<td>6–12</td>
<td>1–4*</td>
</tr>
<tr>
<td>Energy per bit (pJ)</td>
<td>0.1–3</td>
<td>2–25</td>
<td>0.1–2.5</td>
<td>0.0005</td>
<td>0.005</td>
<td>0.00002</td>
</tr>
<tr>
<td>Read time (ns)</td>
<td>&lt;10</td>
<td>10–50</td>
<td>10–35</td>
<td>0.1–0.3</td>
<td>10</td>
<td>100,000</td>
</tr>
<tr>
<td>Write time (ns)</td>
<td>~10</td>
<td>50–500</td>
<td>10–90</td>
<td>0.1–0.3</td>
<td>10</td>
<td>100,000</td>
</tr>
<tr>
<td>Retention</td>
<td>years</td>
<td>years</td>
<td>years</td>
<td>As long as voltage applied</td>
<td>&lt;&lt;second</td>
<td>years</td>
</tr>
<tr>
<td>Endurance (cycles)</td>
<td>$10^{22}$</td>
<td>$10^9$</td>
<td>$10^{11}$</td>
<td>$&gt;10^{16}$</td>
<td>$&gt;10^{16}$</td>
<td>$10^4$</td>
</tr>
</tbody>
</table>

---

*The energy to operate NAND Flash is typically hundreds of picowatts per bit primarily because accessing the memory cells requires charging word and bit lines to high voltages. Smaller number represents an effective area for multi-level cells. PCM, phase-change memory; STTRAM, spin torque transfer random access memory; SRAM, static RAM; DRAM, dynamic RAM; HDD, hard disk drive.
switch is ON, meaning that every reading operation switches the NDR element once. Therefore, reading endurance may be a potential concern of this scheme. A threshold switching (Fig. 1d) with unlimited switching cycles would be very valuable for obtaining high nonlinearity in memristive devices.

In conclusion, the development of memristive devices has recently witnessed remarkable progress. Figure 7, for example, shows device performance requirements for a range of applications and illustrates that most of these requirements have been successfully achieved with memristors, at least in devices fabricated in specialized laboratories. Nevertheless, it remains to be seen if memristive devices can combine these characteristics in a single commercially competitive device design. Further research into device mechanisms — particularly the microscopic processes of the initial and subsequent switchings — is crucial to achieve reliable and predictable nanodevices at the wafer scale. Furthermore, it is time for device designers to be more actively involved in making use of existing device properties and providing guidance on the development of future device properties for applications with different performance requirements.
This article demonstrated switching by the motion of a few atoms.

This article demonstrated switching by the motion of a few atoms.

This article demonstrated switching by the motion of a few atoms.
memristors. Measuring the switching dynamics and energy efficiency of tantalum oxide for nanoelectronic circuits. nanoelectronic crossbar memories.


ITRS International Technology Roadmap for Semiconductors, 2011 edn; (Dover, 1940).


Acknowledgments

Additional information
Supplementary information is available in the online version of the paper. Reprints and permission information is available online at http://ngp.nature.com/reprints. Correspondence and requests for materials should be addressed to J.J.Y., D.B.S. or D.R.S.

Competing financial interests
The authors declare no competing financial interests.