



OPEN

Novel synaptic memory device for neuromorphic computing

Saptarshi Mandal¹, Ammaarah El-Amin¹, Kaitlyn Alexander¹, Bipin Rajendran² & Rashmi Jha¹

¹Department of Electrical Engineering and Computer Science, University of Toledo, University of Toledo, OH 43606, USA,
²Department of Electrical Engineering, Indian Institute of Technology, Bombay, India.

This report discusses the electrical characteristics of two-terminal synaptic memory devices capable of demonstrating an analog change in conductance in response to the varying amplitude and pulse-width of the applied signal. The devices are based on Mn doped HfO₂ material. The mechanism behind reconfiguration was studied and a unified model is presented to explain the underlying device physics. The model was then utilized to show the application of these devices in speech recognition. A comparison between a 20 nm × 20 nm sized synaptic memory device with that of a state-of-the-art VLSI SRAM synapse showed ~10× reduction in area and >10⁶ times reduction in the power consumption per learning cycle.

Emulating “artificial intelligence” in computational devices, inspired by energy-efficient, robust, cognitive, and emergent computational ability of a biological-brain, has inspired the scientists from multiple disciplines for the last several decades. On one-hand tremendous efforts have been made to understand how information processing, learning and decision-making processes are actually performed in a biological brain¹, while, on the other-hand, significant efforts have been devoted to implement some of these understandings in computational systems using software based neural network algorithms². In spite of these significant progresses, software-based neuromorphic approaches impose severe challenges in terms of energy-efficiency and scalability in emulating the complexity and diversity of a biological-brain^{3–4}. For example, a human brain, because of its massively parallel and reconfigurable architecture spanning a complex network of ~10¹² neurons and 10¹⁵ synapses, is able to perform a simple cognitive task by consuming only around 20 W of power as compared to multi-core based supercomputers that require 10,000 times more power⁵. To overcome the limitations of scalability and energy requirement, there have been tremendous efforts to implement hardware-based neuron and synapse network in task-specific neural circuits^{6–7}. However, this approach also suffers from fundamental scalability limitation of Complementary Metal Oxide Semiconductor (CMOS) devices³, given each neuron will require at least 6 transistors for the axon hillock⁸ and a plastic synapse will require more than 10 components⁹. Therefore, over the past few years several 2-terminal (2-T) devices have been discovered which can emulate synaptic behaviour. A fundamental property of a synapse is the analog change in its efficacy when subject to different input conditions. The 2-T devices that have gained attention include WO_x¹⁰ and Ag:Si¹¹ based devices whose conductance or strength can be modulated using different input bias, much like a synapse subject to various inputs. The mechanism of operation of these devices has been shown to be either the movement of oxygen vacancies (WO_x) or the migration of dopants (Ag) in the semiconductor material. Resistive random access memory devices based on formation and rupture of conducting filaments¹², and phase-change memories based on resistance modulation by bias dependent change of phase¹³, have also shown synaptic characteristics. In spite of these recent advancements, a device model supported by experimental results has been lacking. In this paper a synaptic memory device is presented which shows a reconfiguration in conductance as a function of input pulse parameters— amplitude and width, caused due to the generation and annihilation of defects. The mechanism of generation of defect was explained by stress induced leakage current (SILC)¹⁴. The physics behind this model is investigated and a comprehensive device model is presented. The model is then used to simulate learning in a 16 × 16 crossbar array consisting of these synaptic devices by spike timing dependent plasticity (STDP) learning algorithm. A simulation of speech recording and discrimination was shown using these devices as the synaptic component. Finally a comparison of the performance specifications of the proposed synaptic memory device when scaled to nanometer dimensions with that of biological systems and existing VLSI circuits for neuromorphic computation is presented.

SUBJECT AREAS:
ELECTRONIC DEVICES
ELECTRICAL AND ELECTRONIC
ENGINEERING

Received
9 January 2014

Accepted
1 May 2014

Published
18 June 2014

Correspondence and
requests for materials
should be addressed to
R.J. (rashmi.jha@
utoledo.edu)

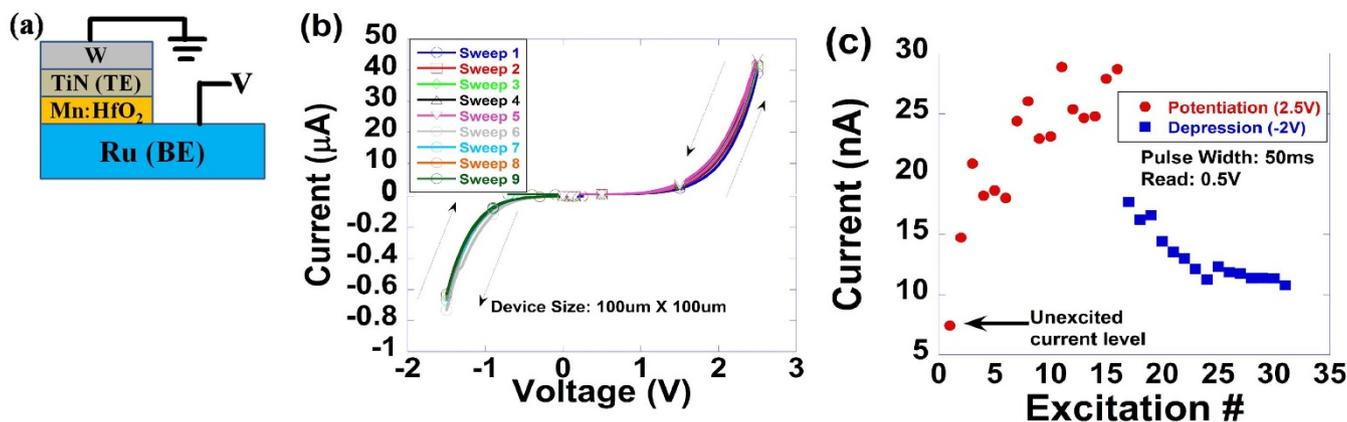


Figure 1 | (a) Device structure and testing configuration (b) IV hysteresis. Incremental loops were observed when the device was swept to 2.5 V and back. The conductance of the device decreased during negative sweep to -1.5 V. (c) Potentiating (2.5 V) pulses of 50 ms width were applied for 15 cycles. Then depressing (-2 V) pulses were applied for next 15 cycles. The plot shows the read (0.5 V DC) after each excitation. The first data point is initial conductance level. The first increase/decrease is always greatest. Saturation of conductance tends to occur after a number of cycles.

Results

The device structure and testing configuration is shown in Figure 1(a). Mn doped HfO_2 forms the switching layer while Ru and TiN form the bottom electrode (BE) and top electrode (TE) respectively. Figure 1(b) shows hysteresis in I-V with repetitive DC sweeps. Positive voltage sweeps increase the conductance, while negative voltage sweeps decrease it. Figure 1(c) shows the current reads at 0.5 V after each excitation. A +2.5 V, 50 ms wide pulse was applied to the device repeatedly for 15 times and the current was measured after each excitation. It is evident that the first increase of conductance (from 1 to 2) is always the highest, while the conductance tends to saturate with increased number of pulses. Once the device was driven to near saturation of increased conductance, -2 V, and 50 ms wide pulses were applied. Here, too, the first decrease is highest; while subsequent reduction in conductance tends to saturate.

Next, capacitance voltage (CV) characteristic of the device was obtained at several frequencies as shown in Fig. 2(a). From here, assuming a dielectric constant of 24, the film thickness was estimated to be 9.93 nm. To understand the mechanism of charge transport, I-V sweeps were performed at temperatures ranging from 260 K to 350 K. The conduction mechanism was found to be Frenkel-Poole (F-P) emission¹⁵ based on excellent r-square (R^2) values obtained for the F-P fitting, shown in Figs. 2(a)–2(d). The equation for F-P can be given as:

$$I = q\mu EA n_0 \exp\left(-\frac{\phi_B - \sqrt{\frac{qE}{\pi\epsilon}}}{kT}\right) \quad (1)$$

Here, μ is the mobility of dielectric, E is the electric field, A is the area of the device, n_0 is defect concentration and Φ_B is the depth of the trap from the conduction band of HfO_2 which is corrected for the electric field in the exponential. Figures 2(b) and 2(c) show $\ln(I/V)$ vs. $\sqrt{\text{V}}$ for different temperatures for positive bias and negative bias respectively. Beyond 0.2 V, a straight line fitting with R^2 values between 0.998 and 0.999 is obtained for the plots at all temperatures indicating the conduction mechanism to be dominated by F-P. At low bias (<0.2 V) some other mechanism can be dominant, such as trap-assisted tunnelling¹⁷ at low temperatures and thermionic emission at higher temperatures (>330 K). The parameters for emission were determined by extracting the slope (E_a) of $\ln(I/V)$ vs. $1/kT$ plot for different bias points as shown in figure 2(d) for positive bias and 2(e) for negative bias. For comparison, Schottky emission fittings for $\ln(I/T^2)$ vs $1/kT$ were also tried as shown in the inset of figures 2(d) and 2(e). However, R^2 values for F-P (0.983–0.999 for negative biases

and 0.998–0.999 for positive biases) were found to be better than Schottky fittings indicating the dominant mechanism of conduction to be F-P in these samples. The E_a was then plotted as a function of the square root of V for positive and negative bias as in Figure 2(f). The extracted Φ_B for positive and negative bias were found to be 0.207 and 0.232 eV respectively. Assuming $\mu \sim 0.15 \text{ cm}^2/\text{V}\cdot\text{s}$ ¹⁶, n_0 was estimated to be around $3 \times 10^{11} \text{ cm}^{-3}$. Using Mn:HfO₂ thickness of 9.93 nm, extracted from CV, the dielectric constant of ~ 24 was extracted for both positive and negative bias using F-P fitting. F-P emission is usually associated with symmetric I-Vs²⁹ due to bulk defects. However, asymmetric I-Vs in our devices could be a result of different Φ_B observed for the positive and negative biases. It is possible that TiN, being an oxygen gettering layer, can getter oxygen from HfO_2 near TiN/ HfO_2 interface which can lead to different oxidation states of Mn. As a result, defects of different depths in the band-gap of HfO_2 can exist which can preferentially participate under positive and negative bias. It has also been reported that the dielectric constant extracted for F-P emission in HfO_2 corresponds to optical frequencies³⁰. However, the optical dielectric constant is valid for very high fields, while the E-field for our samples was much lower.

To gain an insight into the physical processes that govern the hysteretic behaviour, constant voltage stressing (CVS) of the device was performed. It was observed that the current increases during the stress. Figure 3(a) shows the increase of current during CVS of the device when a pulse of 2.5 V and 100 ms duration was applied to it. Such an increase of current under stressing is usually observed in high-k dielectrics where the mechanism is described by SILC model¹⁴. The equation for SILC is given as:

$$I = I_0 + N \left[1 - \exp\left(-\frac{t}{\tau_0}\right) \right] + \alpha \cdot t^\gamma \quad (2)$$

Here, I_0 denotes the current at the start of the 2.5 V bias, N is the saturation value of electron de-trapping, α is the leakage current from SILC traps and γ is the trap generation rate. The experimental data was fitted with this equation and the parameters were extracted as shown in Figure 3(a). The physical process of the model in our device can be explained as follows. The Mn:HfO₂ initially has V_o based defects with electrons trapped in them. When a positive bias is applied, the electrons de-trap and participate in conduction. The de-trapping of electrons from pre-existing defects is a field-dependent phenomenon. Therefore, the device has different increase of current under different bias. The de-trapping effect is usually modelled by the second term of (2). However, as supported by the extracted parameters, de-trapping is a fast process and the second term quickly reaches its maximum. The subsequent increase of current in the remainder of the pulse can thus be attributed to gen-

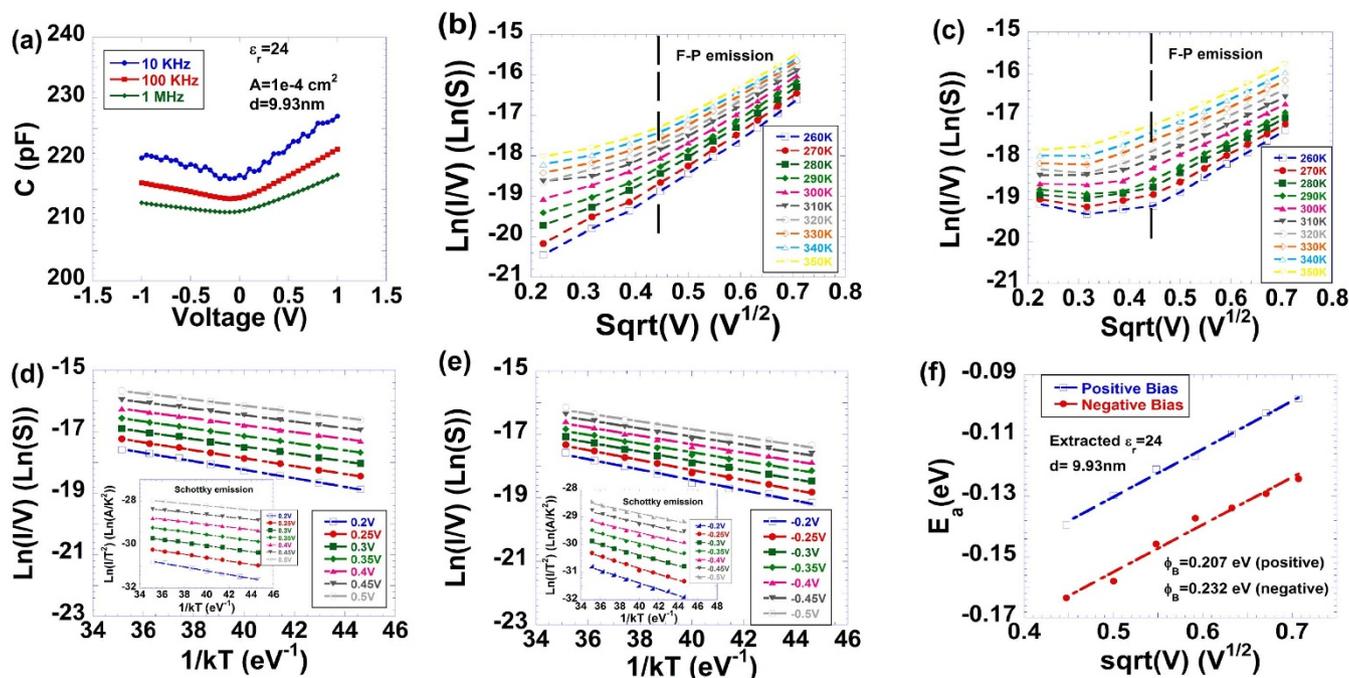


Figure 2 | (a) CV measurement is shown for the sample. Assuming a dielectric constant ~ 24 , the thickness was estimated to be 9.93 nm. The dielectric constant was verified by the transport study. $\ln(I/V)$ vs $\text{sqrt}(V)$ plots are shown here for (b) positive and (c) negative bias. Straight line fittings are obtained beyond 0.15 V which indicate that the dominant current conduction mechanism is F-P. At low bias and/or high temperatures (>330 K) different conduction mechanisms set in. $\ln(I/V)$ vs $1/kT$ plot is shown for different bias points for (d) positive bias and (e) negative bias. The activation energy for carriers in F-P emission is extracted from the slopes of these plots. The intercepts provide the average carrier concentration in the device, assuming a mobility of $0.15 \text{ cm}^2/\text{V}\cdot\text{s}$. $\ln(I/T^2)$ vs $1/kT$ plots for Schottky emission are also shown in the insets of (d) for positive bias and (e) for negative bias. The fits obtained for F-P were better than that of Schottky which confirmed the dominant conduction mechanism in the dielectric to be F-P. (f) The activation energy is plotted as a function of $\text{sqrt}(V)$. The slope provides the dielectric constant of the material. For a thickness of 9.93 nm the dielectric constant was estimated to be 24. The average trap depth was 0.207 eV for positive bias and 0.232 eV for negative bias.

eration of additional defects during the stressing, which take part in conduction. This is denoted by the third term. Therefore, when repetitive positive pulses are applied, the conductance increase is highest in the first pulse due to fast de-trapping of electrons. Thereafter, the increase in subsequent pulses is small due to the leakage current term. This explains the current saturation trend in figure 1(c).

An analogous decrease of current is observed when the device is stressed using a negative CVS. Figure 3(b) shows the decrease of current when a -2 V pulse of 100 ms duration was applied to the device which had previously been excited using positive CVS (Figure 3(a)). The current was fitted to the following equation:

$$I = I_n + N_n \left[\exp\left(-\frac{t}{\tau_n}\right) \right] + \alpha_n t^{-\gamma_n} \quad (3)$$

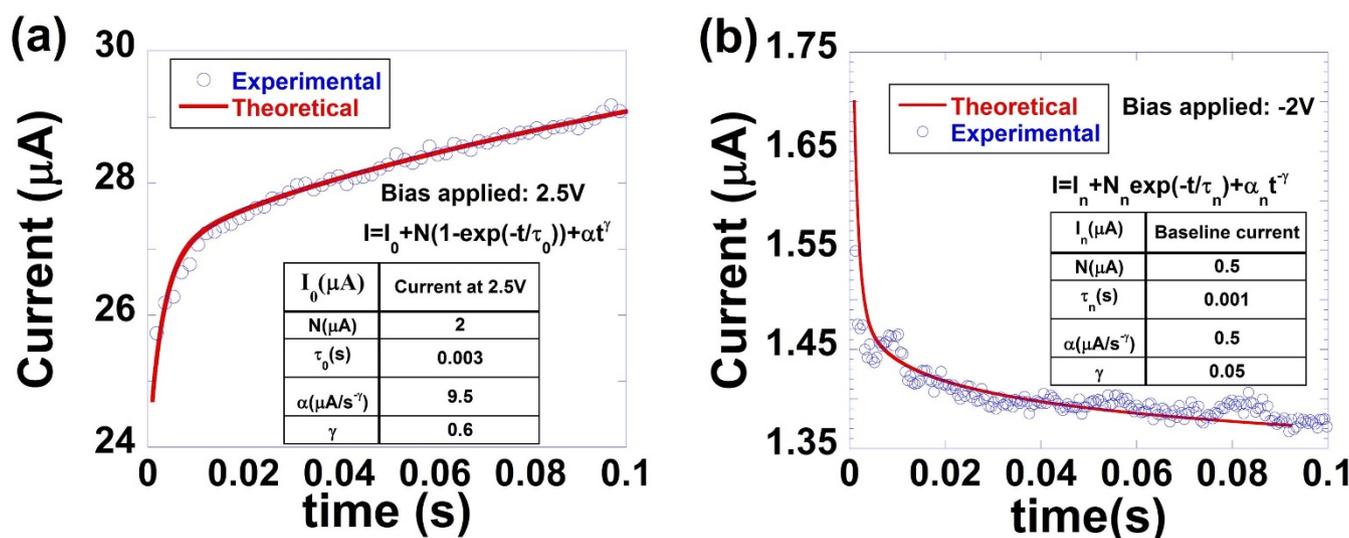


Figure 3 | SILC fitting. (a) Increase in current during the constant voltage pulse (2.5 V) for 100 ms. SILC model (equation (2)) was used to fit the data. (b.) -2 V pulse of width 100 ms was applied to a device driven to high conductance by the pulse in (a). The decrease was modelled according to equation (3).

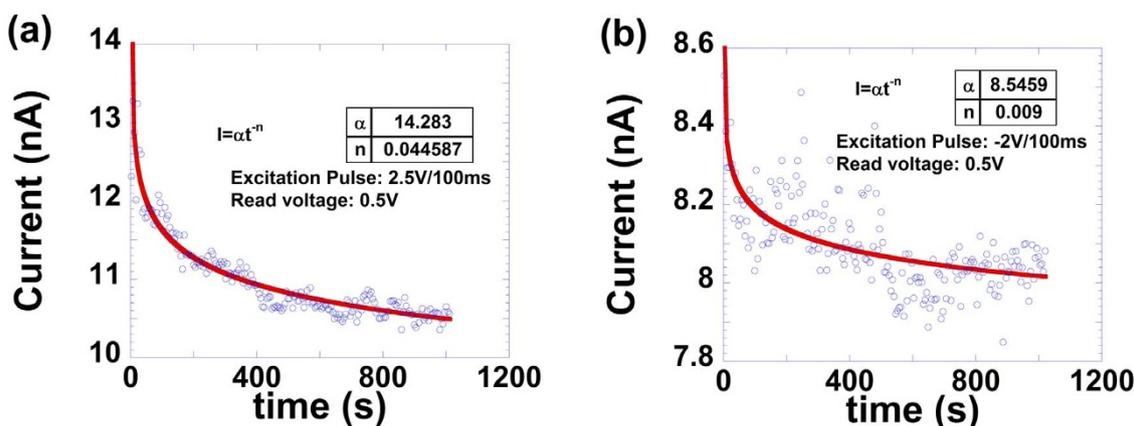


Figure 4 | Dielectric relaxation. (a) relaxation after positive voltage stressing. 2.5 V/100 ms pulse was used for the stressing and the current was measured once the stress was removed. The conductance saturates after 1000 s. (b) relaxation after -2 V/100 ms pulse was applied to a positively stressed device. Saturation of conductance is evident here too.

Here I_n is the unstressed conductance level of the device. Hence the device can be reconfigured to its unexcited condition only when a long pulse is applied to the device. It is hypothesized that the first term indicates re-trapping of electrons in the defects that were emptied during positive stressing, while the second term denotes annihilation of the oxygen vacancies that were generated during positive CVS. Here, too, the saturation of conductance decrease in subsequent negative pulses can be modelled by neglecting the second term of (3) in subsequent pulses.

Once the stress is removed from the device, its conductance tends to decay. Such a transient decay of conductance under low bias is usually attributed to dielectric relaxation in high k -dielectrics. This process can be modelled using the Curie-von Schweidler (CS) equation for relaxation¹⁸.

$$I = c.t^{-n} \quad (4)$$

Figure 4(a) shows the relaxation of conductance after removal of the positive CVS. Based on the fit using CS equation, the time to reach the initial un-excited conductance was estimated to be 3.3 months. However, the device conductance already seems to be saturating towards the end of 1000 s, which would suggest that the conductance

is retained. Similar fitting was done for relaxation after a negative pulse was applied to the device. The fitting is given in figure 4(b).

Discussion

From the results in the previous section, it was clear that the hysteresis in I - V s is caused by the increase of n_0 during positive sweep and decrease in n_0 during negative sweep. Therefore, in order to model the hysteresis, it was necessary to obtain the transient current increase and decrease as a function of applied bias. During a voltage sweep, each bias point is applied to the device for some time before a measurement is done. This stress during each bias point increases/decreases the conductance of the device and hence the current increases/decreases depending on the polarity of the bias. CVS was applied to the device in increasing amplitudes of bias and the SILC increase and the current decrease parameters were extracted as a function of bias. Figure 5(a) shows positive CVS on a device with increasing voltages ranging from 1.25 V to 2.5 V. No significant change of current was observed below 1.75 V constant stress which indicates that the activation of SILC and hence hysteresis requires a minimum electric field. The parameters for SILC were extracted by fitting the I - t curves with equation (2) and their values are provided

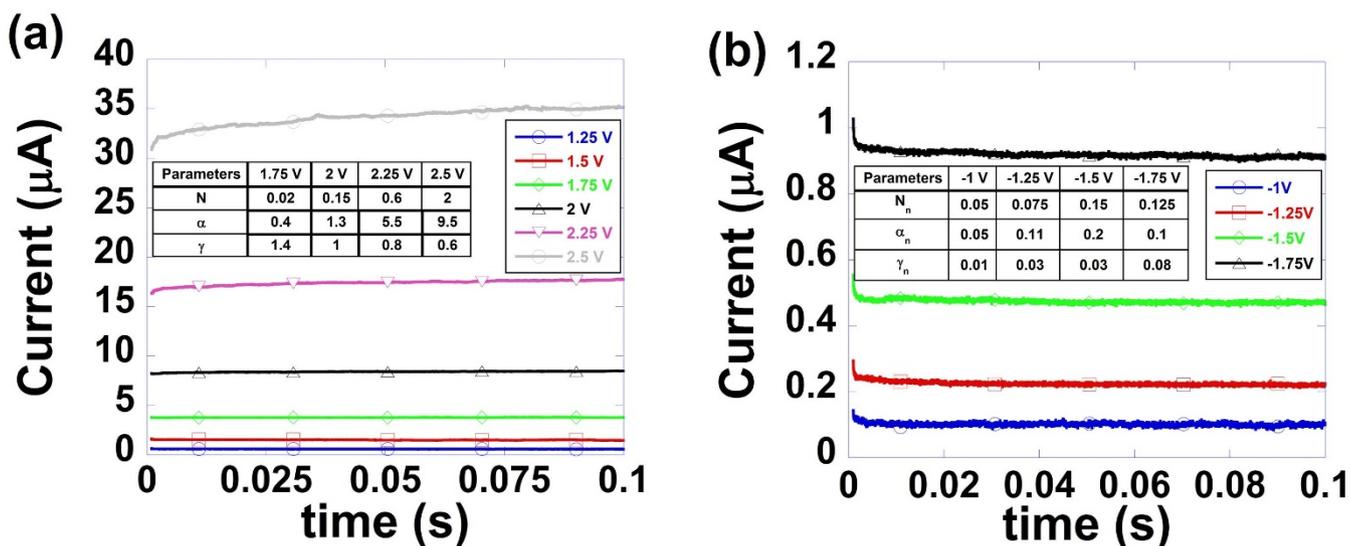


Figure 5 | (a) Subsequent constant voltage stressing for different positive biases is shown. The extracted SILC parameters are field dependent as shown in the inset (b) Several negative biases applied to device after positive excitation and the extracted parameters are shown in inset. The parameters have a weak field dependency for negative bias. Stress time was 100 ms in both plots. The biases applied are given in legends. The current increase or decrease is obtained as a function of field based on the extracted parameters from the plot.

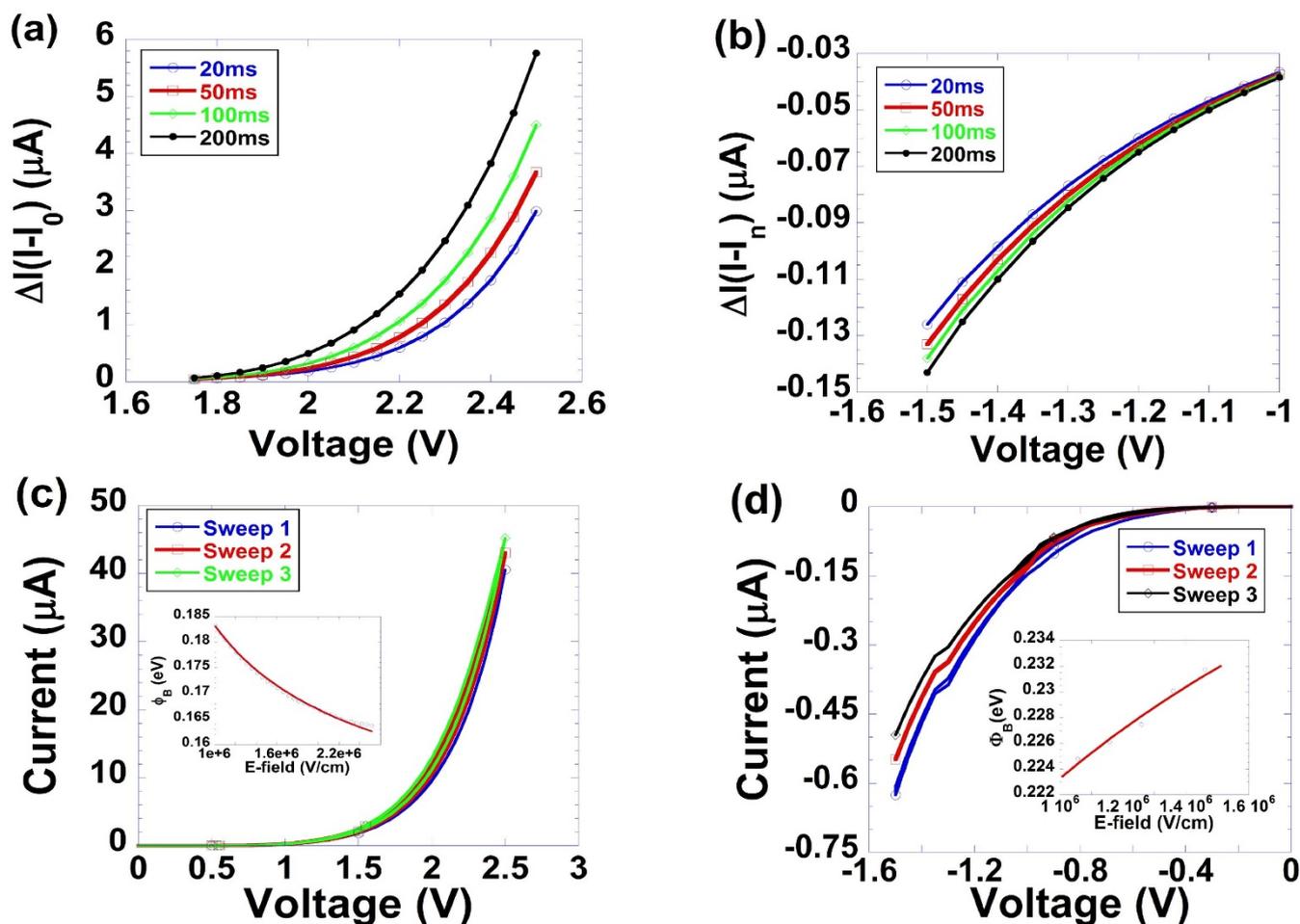


Figure 6 | (a) Excess current generated is given as a function of bias applied. Several time stamps (time for each bias point application) are used. The current at each bias point was evaluated using the field dependent parameters extracted from Figure 5. (b) Current decrease during negative hysteresis as a function of bias. Here, too several time stamps are used as shown in legend. These two plots are used to extract the dependency of trap level with E-field during the hysteresis. (c) Positive I–V hysteresis as obtained from the theoretical model. The carrier concentration increased during each bias step to simulate hysteresis. The field dependency of trap depth was also accounted for. A decay was introduced in between sweeps given by equation (4) to get overlap of hysteresis. The variation of trap depth with E-field is shown in the inset. The traps generated during stressing occupy higher energy than native traps. (d) Negative hysteresis from theoretical model. Decrease of carrier concentration and trap depth dependency were included. In subsequent sweeps the decay is only given by relaxation effects and hence very small hysteresis is observed. This follows the experimental observation of negative I–V sweeps wherein the first decrease is high. The kink in negative hysteresis is due to a combined effect of conductance decrease and trap depth increase at higher negative bias. This is consistent with the experimental observation. The trap depth variation with E-field is shown in inset. As the generated traps during positive stressing are annihilated, the trap depth increases as the conduction occurs more often through the native traps.

in the inset of Figure 5(a). Similarly, the device was stressed using negative bias ranging from -1 V to -1.75 V. Figure 5(b) shows the decrease of current during negative stressing. The parameters were again extracted from fits of the decrease to equation (3) and are presented in the inset. It is observed that for negative bias, the parameters for the stress induced reduction in current have a weak dependency on the applied bias.

From the extracted parameters, the change in current during stressing can be estimated as a function of applied bias. During a positive sweep, the excess current generated due to SILC can be obtained by incorporating the field dependency of the extracted parameters in equation (2). Likewise, the reduction in current during negative sweep can be obtained by incorporating the parameters extracted in Figure 5(b) into equation (3). Figure 6(a) and 6(b) shows this increase and decrease in current respectively as a function of bias. The time-stamp for each bias point is varied from 20 ms, 50 ms, 100 ms and 200 ms. It is evident that for larger time stamps or in effect slower sweep rates, the change in current due to stress is higher. This field dependent change in current could be added to or subtracted from the F–P equation to model the I–V hysteresis in positive

or negative bias. However, using the Φ_B from equation (1) the current was under-estimated for the positive bias while the shape of the curve did not fit well. Therefore, it was apparent that along with the increase/decrease in the density of traps in the dielectric, the Φ_B was also changing during voltage sweeps. In fact it was observed that the Φ_B decreases during positive voltage sweep. This can be explained by assuming that the traps generated during positive voltage sweeps occupy a higher energy in the dielectric than the native traps, thus lowering the average Φ_B . In an analogous manner, during negative sweeps, the Φ_B would increase as the extra defects generated get annihilated. The hypothesis was confirmed by obtaining the relation of Φ_B with E-field.

To obtain the variation of Φ_B with E-field, the following procedure was applied. A time-stamp of 100 ms was used for each bias point. For positive hysteresis, the current due to SILC was obtained for voltages equal to and above 1.75 V. This current was subtracted from the experimental hysteresis I–V to obtain the unstressed current level of the device, when there is no trap generation. The unstressed current is then used to extract Φ_B as a function of E-field using the F–P equation (1). Similarly, for the negative hysteresis, the unstressed



device current refers to the condition when there is no decrease of current due to negative stress. Therefore, the unstressed current was the sum of the experimental current and the current decrease due to stressing. For positive bias, as explained above, Φ_B was found to decrease with E-field. The best fit relation was found out as:

$$\phi_B(E) = \alpha + \beta \cdot E^{-\kappa} \quad (5)$$

where α and β are constants and κ is the power for the E-field dependency. Similarly, as explained earlier, for the negative bias Φ_B was found to increase with E-field, the relation given as:

$$\phi_B'(E) = \alpha' + \beta' \cdot |E|^{\kappa'} \quad (6)$$

The fittings for the extracted Φ_B are shown in the insets of figure 6(c) and 6(d) for positive and negative sweeps, respectively. Based on this extracted trap depth, henceforth referred to as $\Phi_B(E)$, the carrier density generated or annihilated during positive or negative stress could be extracted as a function of field. Therefore the overall F-P equation needs to be modified to reflect the variations in n_0 and $\Phi_B(E)$ as:

$$I = q\mu EA[n_0 + \Delta n^+(E, t)] \cdot \exp\left(-\frac{\phi_B(E) - \sqrt{\frac{qE}{\pi\epsilon}}}{kT}\right) \quad (7)$$

for positive bias and:

$$I = q\mu EA[n_0' - \Delta n^-(E, t)] \cdot \exp\left(-\frac{\phi_B'(E) - \sqrt{\frac{q|E|}{\pi\epsilon}}}{kT}\right) \quad (8)$$

for negative bias.

Here n_0 denotes the carrier concentration of an unstressed device, while n_0' refers to some carrier concentration after the device was positively stressed. Δn^+ and Δn^- are the changes in n_0 and n_0' respectively due to stressing, and are functions of both E-field and stress time.

Based on these equations, the hysteretic behaviour of the device was modelled as shown in figures 6(c) and 6(d) for positive and negative sweeps respectively. The overlap between subsequent loops was modelled using the relaxation effect between applications of voltage sweeps. Equation (3) was used to include a slight decay of carrier concentration when no bias was applied to the device. The kink in the negative hysteresis is due to the combined effect of increase in $\Phi_B(E)$ and reduction of current. Such a kink is also observed in the experimental data as shown in Figure 1(b). Hence, a unified model was obtained to explain the synaptic behaviour of the Mn:HfO₂ synaptic devices.

To examine the repeatability in the reconfiguration of these devices, an endurance testing on the device was performed as shown in Figure 7. A potentiating pulse of 2.5 V and the given pulse width was applied, followed by measurement of the device conductance. A depressing pulse of the same width was then applied and the conductance was again measured at 0.5 V. Clearly, a repeatable reconfiguration in conductance as a function of pulse width is evident for multiple cycles without any obvious signs of failure.

STDP and Speech Recognition. STDP is a biologically inspired learning algorithm that is typically followed in unsupervised neuromorphic learning¹⁹. In a neural-synapse aggregate, pre-synaptic action potentials (AP) are incident on the synapses that are connected to the dendrite. The dendrite sums the contribution of synaptic weights to the incoming APs and fires a post-synaptic AP once the membrane reaches a certain threshold potential. Based on the relative timing of pre- and post-synaptic AP (Δt), the

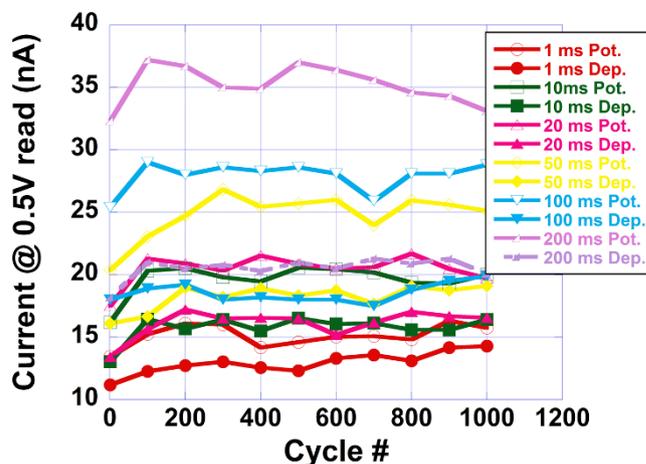


Figure 7 | Endurance for multiple pulse width potentiation and depression is shown for 1000 potentiating/depressing cycles. Here pot. refers to potentiation or applying a 2.5 V bias and dep. refers to depression or applying a -2 V bias. The pulse widths applied are as mentioned in the legend.

corresponding synapse is either potentiated or depressed²⁰. In biological systems, STDP usually occurs in the spike timing window of ± 40 ms with the highest change in synaptic plasticity occurring in the ± 10 ms range. Therefore, when a pre-synaptic AP arrives before postsynaptic depolarization (positive Δt), long-term potentiation (LTP) or the enhancement of synaptic strength occurs, whereas if the postsynaptic firing precedes the pre-synaptic arrival of AP (negative Δt), long-term depression (LTD) or weakening of synaptic strength occurs³¹. The synaptic strength due to STDP in biological systems is usually fitted to an exponentially decaying function²¹:

$$W(t) = A^+ \exp\left(-\frac{\Delta t}{\tau^+}\right) \quad (9)$$

for LTP and

$$W(t) = -A^- \exp\left(\frac{\Delta t}{\tau^-}\right) \quad (10)$$

for LTD. The plot for STDP using these equations is shown in Figure 8(a).

To demonstrate the possibility of implementing STDP using the proposed synaptic devices, a 2.5 V pulse for potentiation and -2 V pulse for depression were applied while the pulse-width (ω) was modulated based on Δt . When a pre-synaptic spike preceded the post-synaptic firing, a potentiating pulse would be applied, while a depressing pulse would be applied when post-synaptic firing precedes the pre-synaptic AP. Once a neuron fired, the relative timing of spike arrival (Δt) was recorded directly into the applied pulse width and polarity by the following mapping procedure. To be compatible with biological systems, a spike timing window of $\Delta t = \pm 40$ ms was chosen, where the highest change of conductance was intended when $\Delta t = \pm 10$ ms. Since these devices needed much longer ω to show appreciable changes in conductance, a relation between Δt and ω was defined such that a $\Delta t = \pm 10$ ms corresponded to $\omega = 200$ ms, $\Delta t = \pm 20$ ms corresponded to $\omega = 100$ ms, $\Delta t = \pm 30$ ms corresponded to $\omega = 50$ ms, and finally $\Delta t = \pm 40$ ms corresponded to $\omega = 20$ ms. This relation between Δt and ω can be conveniently represented by the equation (11).

$$\omega = \lambda \cdot \exp(-\kappa \cdot |\Delta t|) \quad (11)$$

where λ and κ are fitting parameters. This simple test to demonstrate STDP is schematically shown in figure 8(b). The observed potentiation (LTP) and depression (LTD) characteristic of the device, mea-

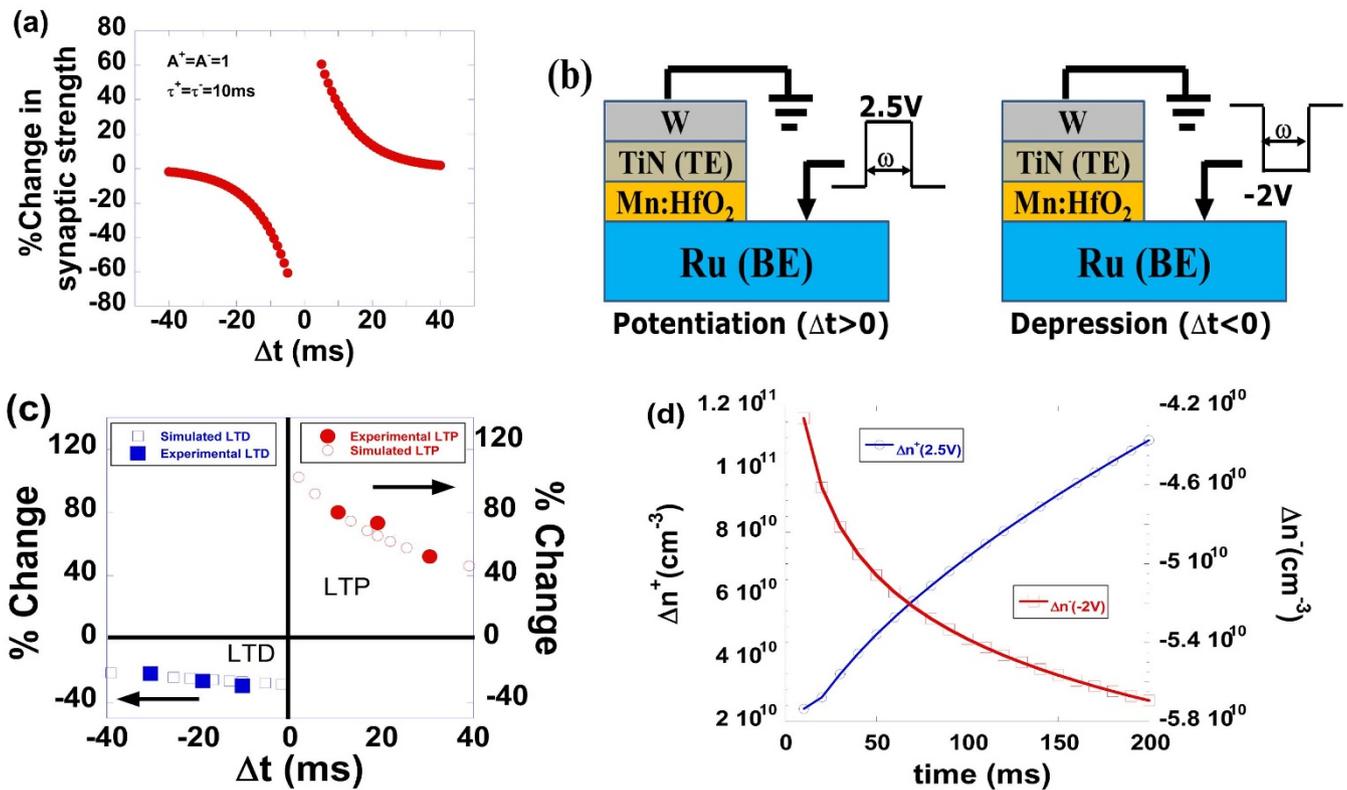


Figure 8 | STDP analysis. (a) STDP plot for a biological synapse. Equations (9) and (10) were used to generate this plot. The largest change in synaptic strength occurs in the $\Delta t = \pm 10$ ms window. The values for generating the plot are given in the inset. (b) Application of pulse during STDP. For $\Delta t > 0$, a potentiating 2.5 V pulse is applied, while for $\Delta t < 0$, a -2 V depressing pulse is applied. The pulse width is determined using equation (11). (c) STDP for theoretical and experimental comparison. Theoretical model used the dependency of carrier concentration on pulse width to calculate percentage change from Figure 8(d). Experimental data was obtained by exciting devices in their unexcited conductance levels using 2.5 V for potentiation. For LTD, already excited devices were stressed using -2 V. The percentage change in conductance is plotted as a function of Δt . The Δt for STDP is obtained from a one-to-one mapping described in the text. (d) Carrier concentration change is shown as a function of applied pulse width. This property was used in implementing STDP. Higher the pulse width, greater is the change in carrier concentration. The plot was also used to generate the theoretical STDP in 8(c).

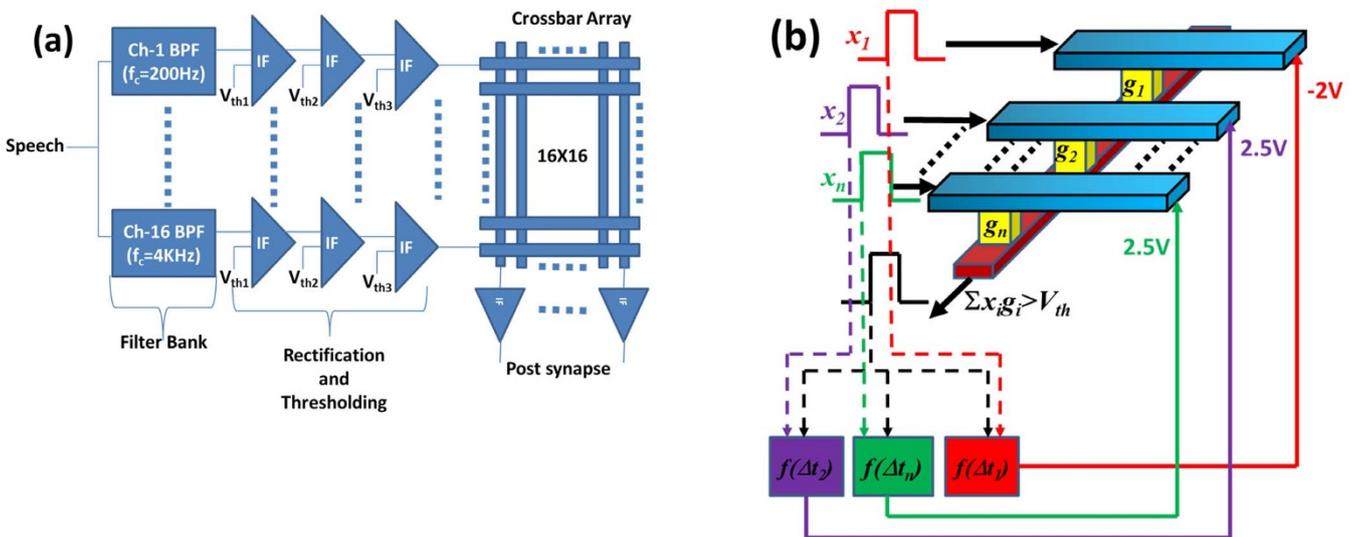


Figure 9 | (a) Block diagram showing the speech processing. The speech signal is sampled at 11.025 KHz and passed to a 16 channel band pass filter bank. The central frequency for each channel is obtained from an exponential distribution of frequencies, where for channel 1, 200 Hz is used and 4 KHz is used for channel 16. After filtering the signals are rectified and passed to thresholding stage where spikes are generated by three serial integrate and fire neurons with three different thresholds. The spikes are directly fed into the crossbar array. The post synapse fires once the threshold crosses 2.5 V. Once it fires, a feedback signal is sent to each of the synaptic device based on STDP rules as defined in the text. (b) Application of STDP is shown here. When the weighted sum of incoming signals $\sum x_i g_i > V_{th}$ the postsynaptic neuron fires an output spike and a feedback signal. The pulse width of the feedback signal is estimated by Δt and is given by equation (11). For LTP, a potentiating pulse of 2.5 V is applied as shown for devices g_2 and g_n while -2 V depressing pulse is applied for LTD as shown for g_1 .

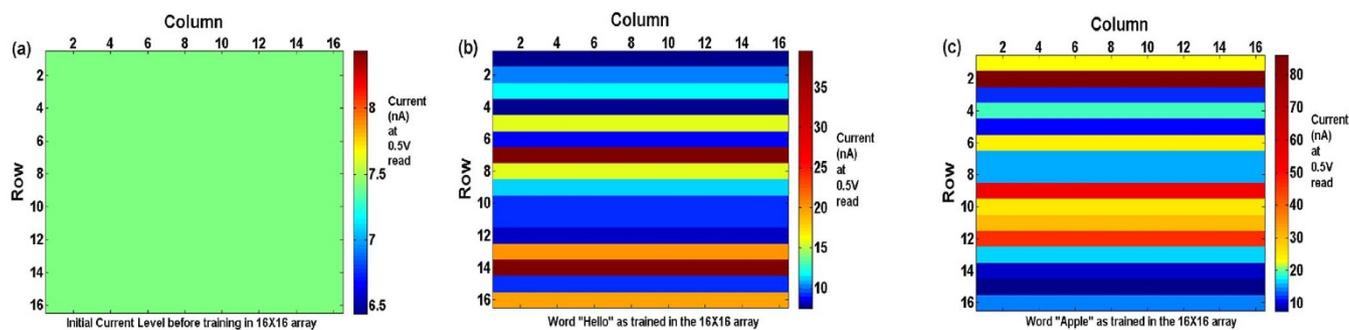


Figure 10 | Synaptic weights. (a) The initial current level of the 16×16 array of synaptic devices is shown. The conductance is initialized to 7.5 nA at 0.5 V read. (b) The word “hello” when trained on to the 16×16 crossbar array of synaptic devices results in a conductance distribution as shown in the plot. The same word was repeated 10 times to ensure the distribution is similar. (c) When the word “apple” was input into the network, the distribution of conductance is markedly different. The weight evolution is obtained from the theoretical STDP fitting of figure 8(b). The colormap shows the current read in nA at 0.5 V after the training procedure. The conductance of each synaptic device along a row is the same since they were initialized to the same conductance level and faced the same STDP training.

sured using this technique is shown in figure 8(c). The LTP was calculated by measuring the change in the conductance of the device in response to the applied potentiating pulse. LTD was calculated by measuring the change in the conductance of the potentiated device in response to the applied depressing pulse. Therefore, the timing information is stored in the device as a change of conductance.

Next, to model the LTP and LTD behaviour of the device, it was necessary to obtain the change of n_0 (Δn) when potentiating or depressing pulses were applied. Figure 8(d) shows Δn of the device as a function of pulse width for potentiating (2.5 V) and depressing (−2 V) pulse. The increase or decrease of current due to the applied pulse width was evaluated first using equations (2) and (3). Δn was then extracted from that current using an average $\Phi_B(E) = 0.19$ eV to account for bias dependent change in the trap depth. Once Δn as a function of applied pulse width was obtained, it was expressed as a percentage change to get the theoretical LTP and LTD values. The pulse width was mapped to Δt using equation (11). This data is plotted in Fig. 8(c).

To demonstrate the application of these devices in neuromorphic learning, a 16×16 crossbar array of Mn:HfO₂ synaptic memory devices was simulated in a neuron-synapse framework to demonstrate STDP algorithm based speech recording. The schematic is shown in Fig. 9(a). A voice was sampled at 11.025 KHz and recorded for duration of 1 s. To emulate the filtering process of the cochlea, the data was passed through a band-pass filter bank. In this simulation, a central frequency (f_c) ranging from 200 Hz to 4 KHz was used to include most of the audible human range. The frequencies were distributed in a log scale and each channel corresponded to one frequency level. The filter bandwidths were chosen following the work of Moore and Glassberg and were defined by^{22,23}:

$$\Omega = 108 \times f_c + 24.7 \quad (12)$$

where, f_c is in KHz. The filtered signals through each channel were rectified. The rectified speech was processed through three integrate and fire (IF) neurons with three different thresholds levels. Each of the IF neurons would fire a spike once the summation crossed their respective thresholds. These spikes were then directly input to the

crossbar array with each cross-point consisting of Mn:HfO₂ synaptic devices connected between input and output neurons.

Training of the synaptic array in the simulation was performed as follows. The devices were all initialized to their unexcited conductance levels. When input neurons spike, it sends out 1 V pulses of 1 ms width as the AP. A counter was used to keep track of the time of arrival of each AP. The incoming currents from different rows were summed along a column of the crossbar and fed to the output neuron. The total current through a given column j is given by:

$$I_j = \sum_i g_{ij} V_i \quad (13)$$

where I_j is the current summation of the j^{th} column, V_i is the magnitude of incoming spike AP for the i^{th} row, and g_{ij} is the synaptic conductance of Mn:HfO₂ device at the cross-point of i^{th} row and j^{th} column. The summed current was used to charge a capacitor of 20 pF for the postsynaptic IF neurom circuit in j^{th} column. Once the potential of the j^{th} column reached a threshold voltage of 2.5 V, a post-synaptic spike was fired. The time of postsynaptic firing was noted and the capacitor was reset to 0 V. The arrival of pre-synaptic spikes was paused temporarily and the Δt was obtained from the pre- and post-synaptic firing instants for each of the rows. Figure 9(b) shows an example of this implementation across one column. The conductance of each synapse in that column was then modified based on the relation shown in figure 8(c). It is worth noting that the pre-synaptic spikes were chosen to be 1 V since it ensured the devices are not affected by the incoming pulses due to the fact that any significant change in device conductance occurs only when the applied bias is > 1.75 V. At the same time the speech was sampled at 11.025 KHz, which meant that each sample of the 1 s recording was of 90 μ s duration. Therefore it must be ensured that when the feedback pulses are applied during STDP, the incoming spikes are paused temporarily until the end of feedback. Hence for practical implementation of such a system, a timer based on a global clock is required which can help keep track of the pre- and postsynaptic firing instants. Once the post-synapse fires, the input pulses are paused by activating delay circuits at the input. The stored instances of pre- and postsynaptic

Table 1 | Comparison of biological synapse with Mn doped HfO₂ synaptic device

Parameter	Biological synapse ²⁷	Mn:HfO ₂ specifications
Postsynaptic potential	<100mV	1V (dielectric thickness =4nm)
Postsynaptic current	<1nA	2.41pA (Device area = 400 nm ²)
Energy consumed per bit	~1fJ	482fJ (200ms pulse)
LTP percentage change	<=100%	56% (for 200ms pulse)

Table II | Comparison of VLSI synapse with Mn doped HfO₂ synapse

Parameter	VLSI synapse ²⁴	Mn:HfO ₂ specifications
Programming voltage	0.5 V	1V (dielectric thickness = 4nm)
Postsynaptic current	100nA	2.41 pA (Device area = 400 nm ²)
Power required for learning	54 μW	2.41 pW
Energy required for learning	540fJ (switching time 10ns)	482fJ (200ms pulse)
Area of synaptic component	0.092 μm ² (SRAM cell)	0.0016 μm ² (20nm)

firings are used to estimate the feedback pulse width and magnitude for each of the synaptic devices from equation (11). The above implementation is based on a synchronous learning scheme. In this scheme, the requirement for keeping track of the precise firing times of neurons can add significant overhead in terms of circuit requirement, which needs to be further studied. To alleviate the requirement of additional circuitry, implementation of asynchronous STDP based on the back-propagation of post-synaptic spikes has been proposed^{24,25}. Since the device conductance can be modulated using both pulse width and amplitude, such asynchronous STDP can also be implemented by capitalizing on the appropriate overlap between pre-and post-synaptic spikes. However, additional circuitry may be needed in this scheme for the desired spike-profile design, which is currently been studied^{24,25}.

The initial synaptic weights before training are shown in Figure 10(a). The current levels at 0.5 V read are shown on the adjacent colour map. A unified colour refers to a constant conductance level for all the devices in the array. Figures 10(b) and (c) show the weight distribution of the synaptic array at the end of the simulation for words “apple” and “hello”, respectively. A clear distinction can be observed in the pattern of conductance levels for these two words. It is interesting to note that the current level along each row of the crossbar is equal as inferred from the colour map. Such a pattern was expected as each of the synaptic elements along a particular row undergoes the same STDP learning since the conductance change and conductance initialization for the synaptic devices were fixed. The impacts of device to device variability and statistical variation in STDP have been examined in a previous work which can be incorporated leading to a more diversified map²⁶.

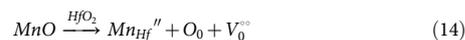
The energy requirements of the synaptic device in the face of biological synapses were also evaluated. The energy consumption for transmission of 1 bit of information across a biological synapse is around 1 fJ²⁷. The excitatory post synaptic current is less than 1 nA while the postsynaptic potential is <100 mV. The size of these devices is 100 × 100 μm². Scaling down to a 400 nm² node would require a current density < (1 nA/400 nm² = 250 A/cm²) for matching a biological synapse. It is apparent from the proposed model that the current due to SILC mechanism is dependent on field and pulse width and independent of trap depth and initial trap density. For scaling down the operating voltages it must be ensured that the electric field is still in the same range. Hence, for an operating voltage of 1 V, the dielectric needs to be scaled to 4 nm. Therefore, the optimum parameters for a 20 nm × 20 nm sized synaptic device operating at 1 V are given in Table I. Here the trap density and trap depth have been assumed to be 5 × 10¹¹ cm⁻³ and 0.19 eV respectively. It must be noted that there is a limit to dielectric thickness scaling since for very thin films F-P emission would cease to be the dominant conduction mechanism and direct tunnelling would tend to take over. Hence the lower limit of thickness was kept to be 4 nm.

A comparison can also be made for the energy and area requirements of the proposed synaptic device with a 22 nm node VLSI synapse based on existing technology as shown in Table II. A ~10× improvement in area is obtained if the synapse circuitry using SRAM cell is replaced by the proposed synaptic devices. The power requirement for programming the device is also significantly low as a reduction of 10⁶ times is obtained, while for a switching time of 10 ns

for the SRAM cell, the energy requirements are comparable with the device. However, since the devices are slow, the overall energy consumption can be lowered further by designing better materials where defect generation and annihilation is a faster process and can occur at much lower fields. The future work in this area would include hardware implementation of the proposed approach and benchmarking against other technologies.

Methods

Device fabrication. The synaptic devices were fabricated using RF magnetron sputtering system. 3 nm of Ti was deposited as the adhesion layer on a 2” p-Si substrate. This was followed by a 100 nm layer of Ru that was deposited as the BE. The 9.93 nm thin switching layer of Mn doped HfO₂ was deposited by co-sputtering of Mn and Hf in an Argon and Oxygen environment. The sputtering power of Mn to Hf was in the ratio of 1 : 5. The Ar:O₂ gas ratio was 4 : 1. The layer was deposited at a substrate temperature of 300 °C. Mn introduces doubly ionized oxygen vacancies and negatively charged defects in HfO₂²⁸ as given by:



The TiN TE was deposited by reactive sputtering of Ti in an Argon and Nitrogen environment. The Ar:N₂ ratio was chosen as 1 : 1 and the substrate temperature was 300 °C. The TE layer was 20 nm thick. A 70 nm thick W capping layer was deposited finally.

Electrical characterization. The device characterization was done in a Lakeshore probe-station under a chamber pressure of 7.5e-5 Torr. The bias was applied to the BE while the TE was always grounded as shown in figure 1(a). 100 μm × 100 μm sized devices were used for characterization. Cryogenic testing was done on 200 μm × 200 μm sized devices. The simulation was performed using MATLAB.

- Gray, C. M. & Singer, W. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *P. Natl. Acad. Sci. USA* **86**, 1698–1702 (1989).
- Masquelier, T. & Thorpe, S. J. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Comput Biol* **3** (2007).
- Borkar, S. & Chien, A. M. The Future of Microprocessors. *Commun. ACM* **54**, 67–77 (2011).
- Modha, D. S., Ananthanarayanan, R., Esser, S. K., Ndirango, A., Sherbondy, A. J. & Singh, R. Cognitive computing. *Commun. ACM* **54**, 62–71 (2011).
- Guizzo, E. IBM’s Watson Jeopardy computer shuts down humans in final game. *IEEE Spectrum* (2011).
- Mead, C. & Ismail, M. *Analog VLSI Implementation Of Neural Systems* [Mead, C. & Ismail, M. (eds)] (Springer, USA, 1989).
- Indiveri, G., Chicca, E. & Douglas, R. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Trans. Neural Networks* **17**, 211–221 (2006).
- Chicca, E. *et al.* A VLSI recurrent network of integrate-and-fire neurons connected by plastic synapses with long-term memory. *IEEE Trans. Neural Networks* **14**, 1297–1307 (2003).
- Fusi, S., Annunziato, M., Badoni, D., Salamon, A. & Amit, D. J. Spike-driven synaptic plasticity: theory, simulation, VLSI implementation. *Neural Comput.* **12**, 2227–2258 (2000).
- Chang, T., Jo, S. H. & Lu, W. Short-term memory to long-term memory transition in a nanoscale memristor. *ACS nano* **5**, 7669–7676 (2011).
- Jo, S. H., Chang, T., Ebong, I., Bhadviya, B. B., Mazumder, P. & Lu, W. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
- Yu, S., Wu, Y., Jeyasingh, R., Kuzum, D. & Wong, H. P. An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. *IEEE Trans. Electron Devices* **58**, 2729–2737 (2011).
- Kuzum, D., Jeyasingh, R. G., Lee, B. & Wong, H. S. P. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Letters* **12**, 2179–2186 (2011).
- Evangelou, E. K., Rahman, M. S. & Dimoulas, A. Correlation of charge buildup and stress-induced leakage current in cerium oxide films grown on Ge (100) substrates. *IEEE Trans. Electron Dev.* **56**, 399–407 (2009).



15. Zhu, W. J., Ma, T. P., Tamagawa, T., Kim, J. & Di, Y. Current transport in metal/hafnium oxide/silicon structure. *IEEE Elec. Dev. Lett.* **23**, 97–99 (2002).
16. Long, B. M., Mandal, S., Livecchi, J. & Jha, R. Effects of Mg-doping on HfO₂-based ReRAM device switching characteristics. *IEEE Elec. Dev. Lett.* **34**, 1247–1249 (2013).
17. Yu, S., Guan, X. & Wong, H. S. P. Conduction mechanism of TiN/HfOx/Pt resistive switching memory: A trap-assisted-tunneling model. *Appl. Phys. Lett.* **99**, 063507 (2011).
18. Jonscher, A. K. Dielectric relaxation in solids. *J Phys. D Appl. Phys.* **32**, R57 (1999).
19. Bi, G. Q. & Poo, M. M. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *J. Neurosci.* **18**, 10464–10472 (1998).
20. Bi, G. Q. & Poo, M. M. Synaptic modification by correlated activity: Hebb's postulate revisited. *Annu. Rev. Neurosci.* **24**, 139–166 (2001).
21. Song, S., Miller, K. D. & Abbott, L. F. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nat. neurosci.* **3**, 919–926 (2000).
22. Glasberg, B. R. & Moore, B. C. Derivation of auditory filter shapes from notched-noise data. *Hearing Res.* **47**, 103–138 (1990).
23. Loisel, S., Rouat, J., Pressnitzer, D. & Thorpe, S. Exploration of rank order coding with spiking neural networks for speech recognition. *IEEE IJCNN* **4**, 2076–2080 (2005).
24. Rajendran, B. *et al.* Specifications of nanoscale devices and circuits for neuromorphic computational systems. *IEEE Trans. Electron Dev.* **60**, 246–253 (2013).
25. Zamarreño-Ramos, C. *et al.* On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Front. Neurosci.* **5**, 1–22 (2011).
26. Mandal, S., Long, B., El-Amin, A. & Jha, R. Doped HfO₂ based nanoelectronic memristive devices for self-learning neural circuits and architecture. *2013 IEEE/ACM NANOARCH* 13–18 (2013).
27. Attwell, D. & Laughlin, S. B. An energy budget for signaling in the grey matter of the brain. *J. Cerebr. Blood F. Met.* **21**, 1133–1145 (2001).
28. Tuller, H. L. & Bishop, S. R. Point defects in oxides: tailoring materials through defect engineering. *Annu. Rev. Mater. Res.* **41**, 369–398 (2011).
29. Yeagan, J. R. & Taylor, H. L. The Poole-Frenkel effect with compensation present. *J. of Appl. Phys.* **39**, 5600–5604 (2003).
30. Jeong, D. S., Park, H. B. & Seong Hwang, C. Reasons for obtaining an optical dielectric constant from the Poole-Frenkel conduction behavior of atomic-layer-deposited HfO₂ films. *Appl. Phys. Lett.* **86**, 072903–072903 (2005).
31. Jeong, D. S., Kim, I., Ziegler, M. & Kohlstedt, H. Towards artificial neurons and synapses: a materials point of view. *RSC Advances* **3**, 3169–3183 (2013).

Acknowledgments

This work was supported by NSF BRIGE Grant No. 1125743, NSF CAREER Grant No. 1254271, and IBM Faculty Award. The authors would like to thank Dr. Mark Ritter and his group at IBM TJ Watson Research Center.

Author contributions

S.M. fabricated the device, performed electrical characterization and data analysis and simulated the speech processing. A.E. aided in device physics and modelling of IV characteristics. K.A. is an undergraduate researcher who participated in circuit assembly and testing. B.R. provided data analysis, critical comments and feedback. R.J. directed the project which involved planning of experiments, testing and analysis.

Additional information

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mandal, S., El-Amin, A., Alexander, K., Rajendran, B. & Jha, R. Novel synaptic memory device for neuromorphic computing. *Sci. Rep.* **4**, 5333; DOI:10.1038/srep05333 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>