On the area scalability of valence-change memristors for neuromorphic computing

Ang, Diing Shenp; Zhou, Yu; Yew, Kwang Sing; Berco, Dan

2019

Ang, D. S., Zhou, Y., Yew, K. S., & Berco, D. (2019). On the area scalability of valence-change memristors for neuromorphic computing. *Applied Physics Letters*, 115(17), 173501-. doi:10.1063/1.5116270

<https://hdl.handle.net/10356/143153>

<https://doi.org/10.1063/1.5116270>

© 2019 Author(s). All rights reserved. This paper was published by AIP Publishing in *Applied Physics Letters* and is made available with permission of Author(s).

Downloaded on 26 Apr 2023 20:21:36 SGT

On the area scalability of valence-change memristors for neuromorphic computing

Cite as: Appl. Phys. Lett. **115**, 173501 (2019); doi: [10.1063/1.5116270](https://doi.org/10.1063/1.5116270)

Submitted: 23 June 2019 · Accepted: 2 October 2019 ·

Published Online: 21 October 2019

View Online

Export Citation

CrossMark

D. S. Ang,^{a)} Y. Zhou, K. S. Yew, and D. Berco

AFFILIATIONS

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

^{a)} Author to whom correspondence should be addressed: edsang@ntu.edu.sg

ABSTRACT

The ability to vary the conductance of a valence-change memristor in a continuous manner makes it a prime choice as an artificial synapse in neuromorphic systems. Because synapses are the most numerous components in the brain, exceeding the neurons by several orders of magnitude, the scalability of artificial synapses is crucial to the development of large scale neuromorphic systems but is an issue which is seldom investigated. Leveraging on the conductive atomic force microscopy method, we found that the conductance switching of nanoscale memristors ($\sim 25 \text{ nm}^2$) is abrupt in a majority of the cases examined. This behavior is contrary to the analoglike conductance modulation or plasticity typically observed in larger area memristors. The result therefore implies that plasticity may be lost when the device dimension is scaled down. The contributing factor behind the plasticity behavior of a large-area memristor was investigated by current mapping, and may be ascribed to the disruption of the plurality of conductive filaments happening at different voltages, thus yielding an apparent continuous change in conductance with voltage. The loss of plasticity in scaled memristors may pose a serious constraint to the development of large scale neuromorphic systems.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5116270>

Artificial intelligence (AI) is now recognized as the main driver of next-generation electronics.^{1,2} At the heart of AI lies the artificial neural network (ANN), which mimics the massively distributed in-memory computing capability of the human brain.³ An ANN comprises microcomputing units or nodes (akin to neurons) interconnected by artificial synapses that emulate the functions of the biological counterparts. Processing and learning (or remembering) of information occur through the simultaneous and continuous adjustment of the strength of synaptic connections throughout the network, controlled by the activity of pre- and post-synaptic neurons.³ This process creates input-specific signal propagation pathways in the network which in turn lead to a fast convergence of the result during deployment. However, artificial synapses designed using traditional CMOS devices have large footprints and power consumption,⁴ making the implementation of large-scale ANNs unfeasible. Therefore, although fundamentally different in architecture, ANNs are usually implemented as software algorithms executed on von Neumann computers. The well-known von Neumann bottleneck^{5,6} has been largely mitigated by performance improvement brought by Moore's law scaling.⁷ But with the scaling already reaching a physical limit,⁸ the continued reliance on a software approach is deemed impractical and alternative hardware solutions have been sought.^{9–11}

The success of hardware ANNs hinges on the availability of a single electronic device that could emulate a biological synapse. The demonstration of a memristor in 2008¹² provided a promising solution for a hardware artificial synapse. Besides having a simple metal/insulator/metal structure which promises scalability and high-density three-dimensional crossbar integration,¹³ a unique property of the memristor lies in the continuous and nonvolatile voltage-controlled modulation of its conductance, which could be leveraged for the emulation of the synapse's plasticity.¹⁴ Basic synaptic behaviors such as long-term potentiation/depression, paired pulse facilitation, and spike-time dependent plasticity have been widely demonstrated on memristors made of different materials.^{15–20} In a large scale ANN, synapses typically outnumber the neurons by several orders of magnitude and synapse scalability is therefore of utmost importance. Whether a memristor could retain its conductance plasticity when scaled down in size is a crucial question which is, however, seldom being addressed. In this work, we examine the scalability of the valence-change memristor, known to exhibit conductance plasticity typically in the reset characteristics.²¹ Numerous nanoscale devices ($\sim 25 \text{ nm}^2$) were tested, and it was found that a majority ($>70\%$) exhibit an abrupt conductance change during reset, in contrast to the gradual conductance change observed in the large-area counterparts. The physics underlying the

change in the conductance transition behavior with area scaling and implications of the finding are discussed. This study suggests that when extreme device area scaling is required, valence-change memristors are not suitable for ANNs which require continuous weight adjustment and alternatives such as binary ANNs^{22–24} would have to be considered.

We leveraged on the nanometer spatial resolution of the conductive atomic force microscope (C-AFM) to realize the testing of conductance switching at the nanoscale. The oxide tested was a 3.5-nm thick²⁵ HfO_x , $x \sim 1.7$ from x-ray photoelectron spectroscopy (not shown). After the sputter-deposition of a 5-nm Ti adhesion layer and a 80-nm TiN electrode on a Si substrate, the HfO_x was prepared by atomic layer deposition using tetrakis(dimethylamino)hafnium(IV) (TDMAH) as the Hf precursor and H_2O vapor as the oxidizer. The deposition temperature was 250 °C and the pressure was 0.2 Torr. A boron-doped diamond-coated Si cantilever probe in an ultrahigh vacuum (UHV) C-AFM served as the top electrode, emulating a nanoscale memristor [Fig. 1(a)]. Surface contamination during electrical testing was avoided by the UHV ambient ($\sim 3 \times 10^{-10}$ Torr). Voltage bias to the probe was supplied by using a parameter analyzer and the TiN electrode was grounded. To avoid damage to the thin oxide, the force applied to the cantilever was increased in small steps until electrical contact was made. The probe/oxide contact area was $\sim 25 \text{ nm}^2$, estimated based on an applied force of 22.8 nN and the mechanical properties of the probe.²⁶ Prior to conductance switching measurement, an electroforming step was performed to create a metastable conductive filament within the oxide layer.²⁷ Figure 1(b) depicts a typical postforming current map scan of the filament. A conductive region of $\sim 18 \text{ nm}^2$, in a good agreement with the estimated probe/oxide contact area, can be observed. After forming, the filament exhibits bipolar conductance switching [Fig. 1(c)], which is characteristic of the HfO_x valence-change memristor.²⁸ For comparison, large-area ($100 \mu\text{m}$ by $100 \mu\text{m}$) TiN/ HfO_x /TiN devices, with a HfO_x thickness of 7 nm, were also fabricated using the standard photolithography process.

The conductance set (i.e., transition from low to high conductance) of the nanoscale device is marked by a sudden increase in

FIG. 1. (a) A nanoscale HfO_x/TiN memristor with a C-AFM probe as the other electrode. (b) Current map after electroforming showing a conductive filament (bright spot) of $\sim 3 \text{ nm} \times 6 \text{ nm}$ in size, corresponding well to the estimated probe/ HfO_x contact area. (c) Current-voltage curves of the filament formed in (b); reset and set voltage sweeps were made in sequence. The arrows show the directions of the sweep.

FIG. 2. (a) Current-voltage curves from a reset voltage-sweep measurement showing (i) an abrupt reset behavior characterized by a sudden drop in current, usually by several orders of magnitude, over a single voltage step and (ii) a gradual reset behavior marked by a progressive decrease in current. The numbers denote the order and the arrows show the direction of voltage sweeps. (b) A majority of nanoscale memristors (>70%) display an abrupt reset behavior for the range of forming current compliance examined, and the percentage increases with a decrease in the compliance. A compliance current less than 50 nA gives an unstable (volatile) filament, whereas a current higher than 150 nA tends to result in a hard breakdown and is not studied.

current, just like the large-area counterpart.²⁹ However, the reset (i.e., transition from a high to low conductance) behavior is different. Unlike the typical gradual decrease in conductance observed on large-area devices²⁹ [cf. Fig. 3(a)], the reset of the nanoscale device is abrupt, as can be seen from the sudden drop in current [Fig. 1(c)]. Figure 2(a)(i) shows another example of an abrupt reset obtained at a different location on the HfO_x . Numerous locations (~ 30) were measured for each forming current compliance. A majority show an abrupt reset,

FIG. 3. (a) Typical set/reset characteristics of a $\text{TiN}/\text{HfO}_x/\text{TiN}$ memristor device, showing gradual conductance reset. Current maps obtained by scanning tunneling microscopy measurement (b) before; (c) after electroforming; (d) a zoom-in view of the area marked by the green box in (c); (e) same region as (d) after resetting some of the filaments individually (dashed circles A–D). (f) Percentage of filaments probed which show abrupt reset, gradual reset, and no reset (numbers in white). The percentages shown in red are for filaments that can be reset. (g) and (h) Distribution of probe voltage and current (just before reset), respectively, for filaments showing abrupt reset.

as summarized in Fig. 2(b) for compliance currents ranging from 50 to 130 nA. Only a minority show a gradual reset behavior [Fig. 2(a)(ii)]. The forming current is limited to this range because for a compliance current below 50 nA, the filament formed is volatile (i.e., it resets the moment the forming is completed), whereas a compliance current higher than 150 nA leads to a filament which cannot be reset. The percentage of abrupt reset cases can be observed to increase with a decrease in the forming current compliance, from $\sim 64\%$ at 150 nA to $\sim 86\%$ at 50 nA. The results imply that the conductance plasticity of the HfO_2 memristor may be lost when its dimension is scaled down. The loss would inevitably limit the scalability of memristors in synaptic applications, especially in spiking neural networks where analog weight tuning is mandatory for the functioning of the network.³⁰

On the other hand, large-area $\text{TiN}/\text{HfO}_x/\text{TiN}$ devices exhibit typical gradual conductance reset²⁹ [Fig. 3(a)]. To probe the difference, the current mapping scan was first performed on numerous randomly selected regions of a pristine device. Because of the top TiN electrode, the scan was performed in the noncontact tunneling mode, where the probe was separated from the TiN by a vacuum gap ($\sim 5 \text{ \AA}$).³¹ The voltage recorded therefore included the voltage drop across this vacuum gap. After the current mapping scan, the device was transferred to a probe station for electroforming. After the electroforming, the device was reloaded into the C-AFM for the current mapping scan again. Before forming, the current maps obtained are uniform; Fig. 3(b) shows an example. In contrast, multiple bright shades denoting high leakage-current regions within the HfO_x can be seen after forming [Fig. 3(c)]. A closer examination of these bright shades (e.g., those within the green box) reveals numerous conductive filaments ranging from several to tens of nanometers [Fig. 3(d)]. An attempt to reset these filaments was made by locating the probe over each of these filaments. A majority ($\sim 62\%$) of those which show a conductance reset [e.g., filament A which no longer appears in Fig. 3(e) after reset] were found to exhibit an abrupt behavior like the nanoscale memristor [Figs. 1(c) and 2(a)(i)]. The general agreement between the percentages of abrupt and gradual reset [Fig. 3(f)] with those obtained from the C-AFM measurement [Fig. 2(b)] should be noted. Moreover, the reset voltage and current vary broadly [Figs. 3(g) and 3(h)]. Such a broad variation may be ascribed to the stochastic nature of filament formation.³²

From these results, one may infer that the disruption of filaments occurring at different voltages during the reset sweep contribute to the gradual current decrease or conductance reset observed in the large-area device. Because the reset voltage measured in Fig. 3(g) includes the voltage drop across the vacuum gap, the current-voltage (I-V) curves of the filaments probed cannot be directly added to give an I-V curve representing a larger area device. To provide support for our inference, we also examine the reset voltage and current distributions for the nanoscale memristors measured by the C-AFM probe and similar broad distributions are observed [Figs. 4(a) and 4(b)]. For a given forming current compliance, we could qualitatively reproduce the gradual reset characteristic of a large-area device by summing up the I-V curves of the individual filaments [e.g., Fig. 4(c) for a 90 nA forming compliance].

To further probe the reason behind the predominantly abrupt reset behavior of the nanoscale memristors, a physics-based simulation study on filament formation in HfO_x under voltage stressing was performed. The algorithm used was earlier described in detail in Refs. 33

FIG. 4. Distribution of (a) reset voltage and (b) reset current (just before reset) of nanoscale memristors formed at various current compliance settings (as indicated). (c) Current-voltage (I-V) curves obtained by summing together the individual I-V curves of nanoscale memristors formed at 90 nA compliance reproduce the gradual conductance reset behavior seen in the large-area device.

and 34 and is summarized here. Initially, the 3.5-nm HfO_x was randomly populated with oxygen vacancy traps. Increased tunneling (described by the Wentzel-Kramers-Brillouin approximation) at the trap site caused local heating and the rise in temperature enhanced the field-induced migration of oxygen towards the anode, leading eventually to the formation of a vacancy filament. The simulation was allowed to progress in steps towards a final state, defined by a preset maximum current or the forming current compliance. At each step, the local potential, temperature, vacancy density, and electrical and thermal conductivities were calculated. The electrical conductivity was described using the Arrhenius model, and the local potential, temperature, and current were determined by the Poisson, charge continuity, and steady-state Fourier equations.³⁵ The thermal conductivity was assumed to vary linearly between $0.5 \text{ W m}^{-1} \text{ K}^{-1}$ (representative of phonon-based conduction in insulators) and $23 \text{ W m}^{-1} \text{ K}^{-1}$ (typical of electronic conduction in metals).³⁵ The change in the local oxygen density was described by the drift-diffusion equation.³⁵

From the simulation study, most of the filaments formed under low current compliances are found to have a “weak link,” characterized by a relatively low density of vacancy defects. Figure 5 shows, for example, the vacancy distribution across the HfO_x for a filament formed under a positive and negative probe voltage. The weak links are highlighted by the dashed circles. The part of the filament with a high vacancy density originates from pre-existing defects. During forming, increased local conduction and heating promote the migration of oxygen towards the anode, leading to new defect formation in pristine oxide regions. The forming process is stopped when a percolation of defects is formed across the HfO_2 , which results in the current exceeding the current compliance. As can be seen from the change in the vacancy distribution during reset, the weak link is the part that is disrupted first, due to its low defect density. Once this occurs, an abrupt decrease in current is seen.

In a large area device, multiple filaments, such as those shown in Fig. 5, are formed, each contributing a small part towards the overall current increase. As the current increase through each filament cannot be restricted, filaments that were formed earlier but with the total

FIG. 5. Examples of simulated spatial distribution of oxygen vacancy defects in a 3.5-nm HfO_2 memristor after positive (top) and negative (bottom) electroforming. The filaments show a weak link characterized by a low vacancy density (dashed circle). Also shown are changes to the vacancy distribution in the filament during reset, where rupture always first occurs at the weak link.

current still below the compliance, would further evolve, with more defects generated at the weak links leading to the formation of more uniform filaments. The reset of these filaments would be gradual, as it involves progressive filament dissolution and thinning as opposed to the sudden “disconnection” caused by the disruption of weak links. This may account for the relatively high percentage of the gradual filament reset cases seen in the large area device [$\sim 37\%$, Fig. 3(f)]. For the nanoscale devices, a higher percentage of gradual reset cases is obtained with a higher forming current compliance. A non-negligible percentage of nonresettable filaments is also observed in the large-area device [Fig. 3(f)]. These are likely to be filaments which had evolved to a more advanced stage of oxide breakdown that could not be reversed due to the depletion of oxygen. From the C-AFM study, we found that filaments formed at a compliance current larger 150 nA were not resettable. The lack of ability to limit the current through individual filaments formed in the large-area device could explain the substantial portion of nonresettable filaments formed.

From the above discussion, a relevant question would then be: “At what device area would the transition from a smooth to an abrupt reset occur?” We attempt a possible projection as outlined below. From the measurements on large-area devices, the average leakage current density just before forming is found to be $\sim 500 \text{ A/cm}^2$. C-AFM measurements on both the large-area device and uncapped oxide sample have shown that the order of current where the abrupt reset behavior of a single filament becomes dominant is $\sim 100 \text{ nA}$. For the current jump during forming to be discernable, we choose the leakage current prior to forming to be an order lower than 100 nA (i.e., 10 nA). This gives an estimated area of $\sim 2 \times 10^3 \text{ nm}^2$. In our projection, we have assumed a single filament. Since “discreteness” in the reset response should begin to manifest in a device with a few filaments, plasticity loss may occur at a larger area than that estimated above.

Conductance switching study at the nanoscale dimension by the C-AFM method has provided useful insights into the area scalability

potential of the HfO_2 valence-change memristor for artificial synapse application. It is found that area scaling would eventually result in a loss of the gradual conductance reset seen in large-area devices. This may be ascribed to a corresponding decrease in the forming current compliance which leads to incomplete formation of the filament, marked by regions of relatively low defect density that always rupture first during reset causing an abrupt change in conductance. Although this study has only involved selected oxide thicknesses, the conclusion may also hold for other oxide thicknesses. An important benefit of area scaling is a reduction in the operating current, which is determined by the filament resistance. For a given oxide thickness, the filament’s resistance varies inversely with its diameter. With a decrease in the operating current, the probability of forming a thin filament with a weak link (thus an abrupt reset behavior) should increase. Increasing the forming current could help raise the percentage of gradual reset cases due to a more uniform filament generation, but such cases remain as minority. Moreover, the number of nonresettable filaments increases significantly above a certain compliance level ($\sim 150 \text{ nA}$ for this study) due to the depletion of oxygen in the nanoscale oxide volume. It is also shown that the gradual conductance reset typically observed in larger area devices is a consequence of many individual filaments resetting at different voltages, thus producing a collective, analog-like conductance change. The loss of gradual conductance reset may pose a serious constraint on the scalability of the valence-change memristor for synaptic applications. If extreme device area scaling is required, alternatives such as binary ANNs^{22–24} would need to be considered.

The authors are grateful to the Singapore Ministry of Education for the funding support under Grant Nos. MOE2016-T2-1-102 and MOE2016-T2-2-102.

REFERENCES

- ¹B.-H. Li, B.-C. Hou, W.-T. Yu, X.-B. Lu, and C.-W. Yang, *Front. Inf. Technol. Electron. Eng.* **18**, 86 (2017).
- ²Y. Zang, F. Zhang, C.-A. Di, and D. Zhu, *Mater. Horiz.* **2**, 140 (2015).
- ³I. A. Basheer and M. Hajmeer, *J. Microbiol. Methods* **43**, 3 (2000).
- ⁴L. Q. Zhu, C. J. Wan, L. Q. Guo, Y. Shi, and Q. Wan, *Nat. Commun.* **5**, 3158 (2014).
- ⁵G. J. Myers, *Advances in Computer Architecture* (John Wiley & Sons, Inc., 1982).
- ⁶F. Yazdanpanah, C. Alvarez-Martinez, D. Jimenez-Gonzalez, and Y. Etsion, *IEEE Trans. Parallel Distrib. Syst.* **25**, 1489 (2014).
- ⁷R. R. Schaller, *IEEE Spectrum* **34**, 52 (1997).
- ⁸L. Beckhout, *IEEE Micro* **37**, 4 (2017).
- ⁹S. K. Esser, R. Appuswamy, P. Merolla, J. V. Arthur, and D. S. Modha, *Adv. Neural Inf. Process. Syst.* **28**, 1 (2015).
- ¹⁰S. Furber, *J. Neural Eng.* **13**, 051001 (2016).
- ¹¹Z. Gingsl, L. B. Kish, and S. P. Khatri, *Fluctuation Noise Lett.* **9**, 403 (2010).
- ¹²A. Krogh, *Nat. Biotechnol.* **26**, 195 (2008).
- ¹³D. Chabi, Z. Wang, C. Bennett, J.-O. Klein, and W. Zhao, *IEEE Trans. Nanotechnol.* **14**, 954 (2015).
- ¹⁴G. C. Adam, B. D. Hoskins, M. Prezioso, F. Merrikh-Bayat, B. Chakrabarti, and D. B. Strukov, *IEEE Trans. Electron Devices* **64**, 312 (2017).
- ¹⁵H. Mostafa, C. Mayr, and G. Indiveri, in 2016 IEEE International Symposium on Circuits and Systems (ISCAS) (2016), Vol. 926.
- ¹⁶Y. Babacan and F. Kaçar, *AEU-Int. J. Electron. Commun.* **73**, 16 (2017).
- ¹⁷N. Du, M. Kiani, C. G. Mayr, T. You, D. Bürger, I. Skorupa, O. G. Schmidt, and H. Schmidt, *Front. Neurosci.* **9**, 227 (2015).
- ¹⁸M. Prezioso, F. M. Bayat, B. Hoskins, K. Likharev, and D. Strukov, *Sci. Rep.* **6**, 21331 (2016).

- ¹⁹X. Zhang, S. Liu, X. Zhao, F. Wu, Q. Wu, W. Wang, R. Cao, Y. Fang, H. Lv, and S. Long, *IEEE Electron Device Lett.* **38**, 1208 (2017).
- ²⁰S. Ambrogio, S. Balatti, F. Nardi, S. Facchinetti, and D. Ielmini, *Nanotechnology* **24**, 384012 (2013).
- ²¹X. Yan, J. Zhao, S. Liu, Z. Zhou, Q. Liu, J. Chen, and X. Y. Liu, *Adv. Funct. Mater.* **28**, 1705320 (2018).
- ²²M. Rastegari, V. Ordonez, J. Redmon, and A. Farhadi, *Lect. Notes Comput. Sci.* **9908**, 525 (2016).
- ²³I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, *Adv. Neural Inf. Process. Syst.* **29**, 1 (2016).
- ²⁴M. Bocquet, T. Hirtzlin, J.-O. Klein, E. Nowak, E. Vianello, J.-M. Portal, and D. Querlioz, in 2018 International Electron Devices Meeting (IEDM) (2018), p. 20.6.1.
- ²⁵We surveyed a range of oxide thicknesses and found that a 3.5–4 nm thick oxide gives the highest percentage of resettable filaments for the C-AFM based nanoscale memristor testing. For thicker oxides, a hard breakdown tends to occur. Because of the higher forming voltage and finite measurement delay, a large surge of current beyond the compliance current setting could occur before the voltage is aborted. For thinner oxides, the larger direct tunneling leakage current would require the forming to be carried out at a higher compliance current setting. This also reduces the percentage of resettable filaments.
- ²⁶F. Nardi, D. Deleruyelle, S. Spiga, C. Muller, B. Bouteille, and D. Ielmini, *J. Appl. Phys.* **112**, 064310 (2012).
- ²⁷M. Buckwell, L. Montesi, S. Hudziak, A. Mehonic, and A. J. Kenyon, *Nanoscale* **7**, 18030 (2015).
- ²⁸A. S. Sokolov, Y.-R. Jeon, S. Kim, B. Ku, D. Lim, H. Han, M. G. Chae, J. Lee, B. G. Ha, and C. Choi, *Appl. Surf. Sci.* **434**, 822 (2018).
- ²⁹F. M. Puglisi, P. Pavan, A. Padovani, and L. Larcher, *Solid-State Electron.* **102**, 69 (2014).
- ³⁰J. Schemmel, A. Grubl, K. Meier, and E. Mueller, in 2006 IEEE International Joint Conference on Neural Network Proceedings (2006), p. 1.
- ³¹K. S. Yew, D. S. Ang, L. J. Tang, and J. Pan, *J. Vac. Sci. Technol., B* **34**, 011205 (2016).
- ³²S. Yu, X. Guan, and H.-S. P. Wong, in 2011 International Electron Devices Meeting (2011), p. 17.3.1.
- ³³D. Berco and T.-Y. Tseng, *J. Comput. Electron.* **15**, 577 (2016).
- ³⁴D. Berco and T.-Y. Tseng, *Appl. Phys. Lett.* **107**, 253504 (2015).
- ³⁵D. Berco, Y. Zhou, S. R. Gollu, P. S. Kalaga, A. Kole, M. Hassan, and D. S. Ang, *ACS Nano* **12**, 5946 (2018).