SAND2018-2543C

Effect of Radiation on the Classification Accuracy of a Neural Network Trained on Analog TaO_x Resistive Memory Arrays

R. B. Jacobs- Gedrim, D. R. Hughart, G. Vizkelethy, E. S. Bielejec, B. L. Vaandrager, S. E. Swanson, K. E. Knisely, J. L. Taggart, H. J. Barnaby and M. J. Marinella

Sandia National Laboratories
Albuquerque, NM 87185-1084

Abstract: The image classification accuracy of a TaO_x ReRAM based neuromorphic computing accelerator is evaluated while intentionally inducing displacement damage to the devices. An effect on classification accuracy only occurred after ~$5x10^{20}$ vacancies were produced.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.M.

Corresponding (and presenting) author:

Robin Jacobs-Gedrim, Sandia National Laboratories, PO Box 5800, Mail Stop 1084, Albuquerque, NM 87185-1084, USA; phone: 505-284-8941, email: rbjaco@sandia.gov

Contributing authors:

D. R. Hughart, G. Vizkelethy, E. S. Bielejec, B. L. Vaandrager, S. E. Swanson and M. J. Marinella are with Sandia National Laboratories, Albuquerque, NM 87185

J. L. Taggart and H. J. Barnaby, School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287-5706

Session Preference: Radiation Effects in Devices and Integrated Circuits

Presentation Preference: Oral

I. Introduction

Resistive Random Access Memory (ReRAM) is one of the leading candidates of beyond CMOS non-volatile memory technologies. A typical ReRAM structure consists of two metal terminals sandwiching a substoichiometric metal oxide layer which is dielectric in the as-deposited state. A typical device is diagramed in the inset of Fig. 1. A positive bias applied to the top electrodes can induce a soft breakdown of the dielectric material in a process known as forming. A negative bias applied to the top electrode can increase the resistance of the device in a process known as Reset which places the device in the High Resistance State (HRS). A subsequent positive can decrease the resistance of the device again in a process known as Set which places the device in the Low Resistance State (LRS). The current-voltage (I/V)

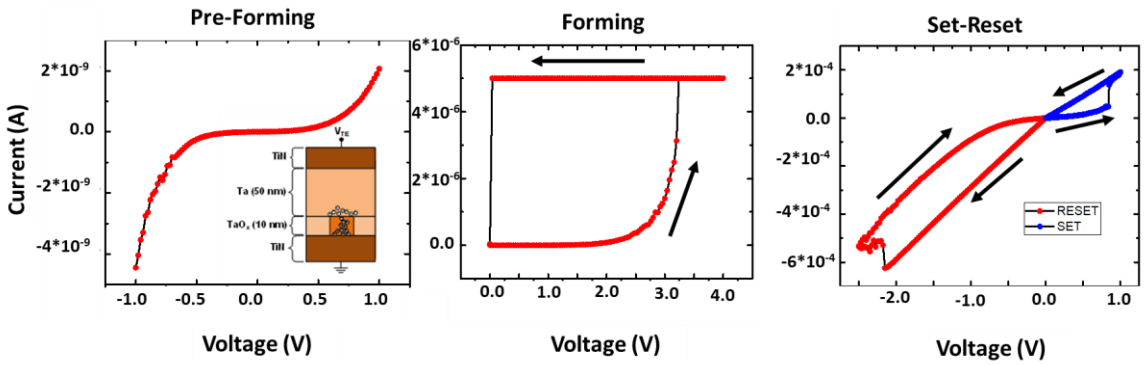

Figure 1. Current Voltage (I-V) characteristics of a ReRAM device. Left: I-V characteristic of device after fabrication and (Inset) diagram of device stack. Middle: I-V characteristics of soft breakdown process known as Forming. Right: Switching the device from high to low resistance and back again using Set and Reset operations.

characteristics of these three processes are shown in Fig. 1.

In addition to memory, it has been suggested that ReRAM may be used for training deep networks [1], [2]. Deep learning is a pattern recognition algorithm that is capable of outperforming traditional machine learning techniques in image recognition, autonomous vehicle, and data science applications. Training deep networks is computationally intensive, and difficult to implement in portable, embedded systems (e.g. spacecraft). Recent analysis has demonstrated that a special-purpose accelerator application specific integrated circuit (ASIC) based on analog ReRAM crossbars can improve the performance per watt by about two orders of magnitude over a digital system – potentially enabling real-time training on embedded systems [2]. The performance increase is obtained by carrying out a vector matrix multiply and weight update in one parallel step, the two most computationally intensive steps in deep network training [3]. An analog electronic vector matrix multiply is diagramed in Fig. 2. By applying bias to the left side of the crossbar, Kirkoff's laws give the current on the bottom of each column as sum of the weights in the column of the crossbar multiplied by the applied voltage amplitude.

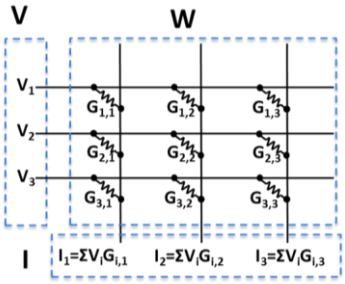

Figure 2: a vector matrix multiply may be implemented on a crossbar with ReRAM analog electronic elements.

Furthermore, a neural accelerator based on analog ReRAM may be accurate even under irradiation. Initial radiation studies for tantalum [4, 5], titanium [6, 7], and hafnium [8-10] oxide based ReRAM devices have demonstrated promising results. Previously, TaO_x ReRAM showed gradual resistance degradation only at high fluences of Si and Ta ions due to additional oxygen vacancies introduced by ion displacement damage to the switching region [4, 5] [11-13]. In the following, we investigate the effects of radiation on the classification accuracy of an analog ReRAM training accelerator for the first time. Results show that analog crossbars are resilient to radiation induced noise expected in space applications.

II. Experimental Details

TaO_x based ReRAM devices were fabricated in the Sandia National Laboratories' Silicon Microfab facility on 6" wafers. The switching stack composed of TiN-TaO_x-Ta-TiN layers is deposited using reactive sputtering, using a feedback technique described in [14]. The reduced TaO_x layer was 15 nm thick and the Ta layer was 15 nm thick, and the active region had 1.0 µm X 1.0 µm lateral dimensions. All measurements were made on a Cascade Microtech manual probe station with an Agilent B1500A Semiconductor Parameter Analyzer mainframe equipped with a B1530 Waveform Generator Fast Measurement Unit. GGB Industries Picoprobe model 40A-GS-250 were employed. The devices were irradiated using the nuclear microbeam on Sandia National Laboratories' Tandem particle accelerator. A 2.5 MeV Si ion broad beam was directed at the device array and measurements were made ex-situ. Results were collected in a lookup table for processing in CrossSim, the Sandia National Laboratories' open sourced platform for determining a device's applicability to neuromorphic computing [15].

III. Results

Devices were formed using a standard I/V sweep as seen in Fig. 1 but with a maximum voltage of 3 V and a 5 uA compliance. A series of 20 SET and RESET cycles were performed using an analog characterization scheme as shown in Fig. 3. The device was then irradiated with 2.5 MeV Si ions to a fluence of $1x10^{10}$ ions/cm^2, followed by an I/V sweep and a repetition of the cycling pulses. Another shot was then performed at an exponentially increasing fluence. Note that there is a subtle change in the analog cycling behavior of the device after the shot with fluence of 10^{14} ions / cm^2. Extracting device resistance using Ohm's law from the I/V sweeps for three devices show a noticeable change in resistance after the 10^{14} shot. These results are similar to our earlier result on binary ReRAM devices [16] where the onset of resistance change occurs near similar thresholds of calculated oxygen vacancy concentrations. Note that during the shots, the device was in the full high resistance state (HRS) as opposed to the low resistance state (LRS)

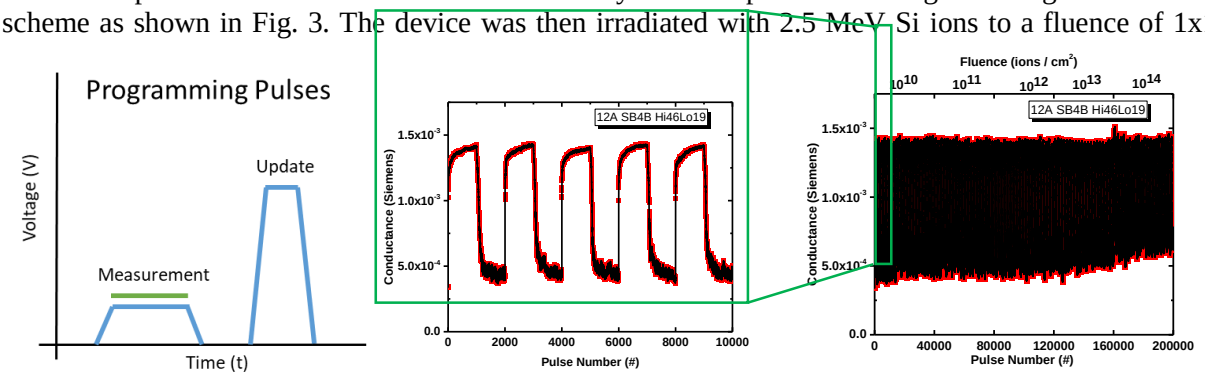

Figure 3: Operation of a TaO_x ReRAM device to simulate neural network training. Left: diagram of the programming pulses used to program the device. First, a measurement pulse is applied to the device, with a measurement during the peak of the pulse, then a voltage update is applied to the device which changes the device conductance. This programming pulse is applied 1000 times for SET and 1000 times for RESET with opposite polarity. Middle: example of one of the devices in the study during analog operation. Right: operation of the same device with ion shots every 40,000 updates. Note that there is a subtle change in the analog conductivity response after the 10^{14} fluence.

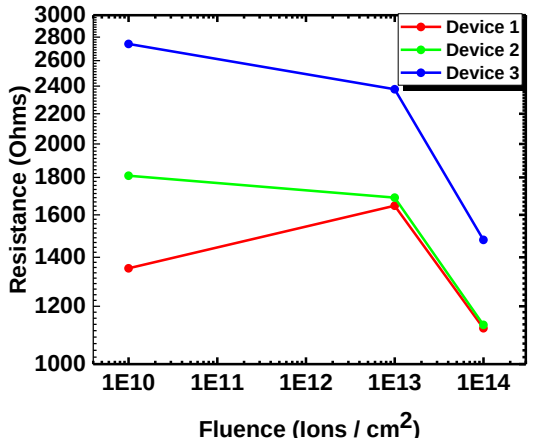

Figure 4: Resistance for three devices extracted by ohm's law from 100 mV I/V sweeps performed immediately following irradiation.

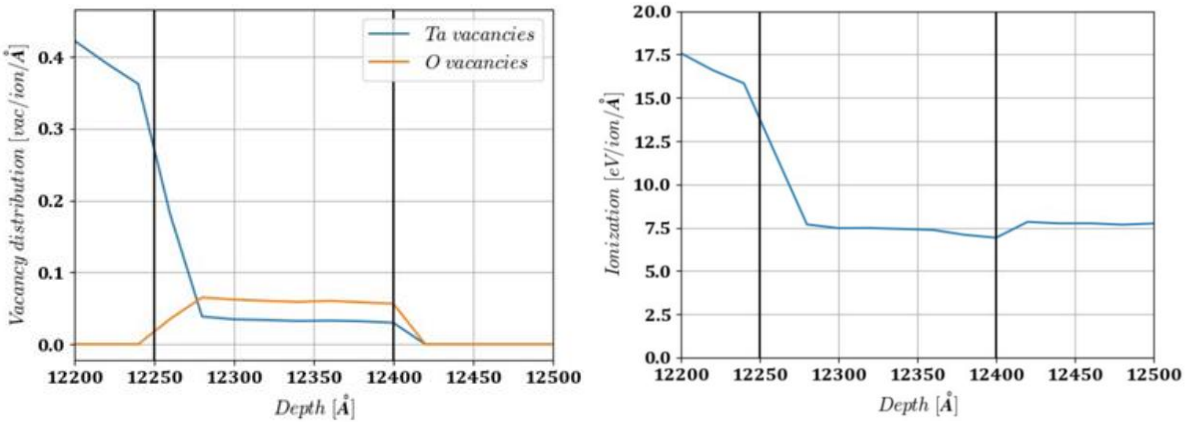

Figure 5: SRIM calculations for displacement and ionization for the TaO_x devices under 2.5 MeV Si ion irradiation. Left: vacancy production vs. depth. Total vacancy production in the TaO_x film is $5.2x10^{20}$ tantalum vacancies and $5.7x10^{20}$ oxygen vacancies. Right: Ionization vs. depth. The average ionization in the TaO_x layer given by SRIM calculations, is 7.5 eV/ion/Å. This corresponds to 208 MRad for $1x10^{14}$ ions / cm^2.

or a middle resistance state. Previous results have indicated that these devices are highly radiation tolerant [4,5] [11-13]. These devices have a thicker set of metals both above and below the switching region than those devices previously tested.

Calculations using the Stopping and Range of Ions in Matter (SRIM) application [17] estimate that the vacancy production in the active region of the device to be $5.2x10^{20}$ tantalum vacancies and $5.7x10^{20}$ oxygen vacancies. A plot of the vacancy distribution with respect to depth is given in Fig. 5. These calculations are only an estimate, compound densities were calculated based on stoichiometry, and SRIM default displacement threshold energies were used. The average ionization in the TaO_x layer given by SRIM calculations, is 7.5 eV/ion/Å, which corresponds to 208 MRad for $1x10^{14}$ ions / cm^2. A plot of the average ionization with respect to depth is given in Fig. 5. The exceptionally high radiation tolerance is likely due to the small volume of the critical area which can be disrupted within the TaO_x layer. While there are some unknown aspects to the conductance switching mechanism in these devices, it has been demonstrated that conductance switching occurs within a nanoscale tantalum rich filament which is created in the forming process [12]. The switch from LRS to HRS by recombination of some of the Ta and oxygen ions is hypothesized to not entire length of this filament during RESET. The resistance of the device will only change if damage occurs within a small critical portion of the filament, which is already nanoscale in all three dimensions [12].

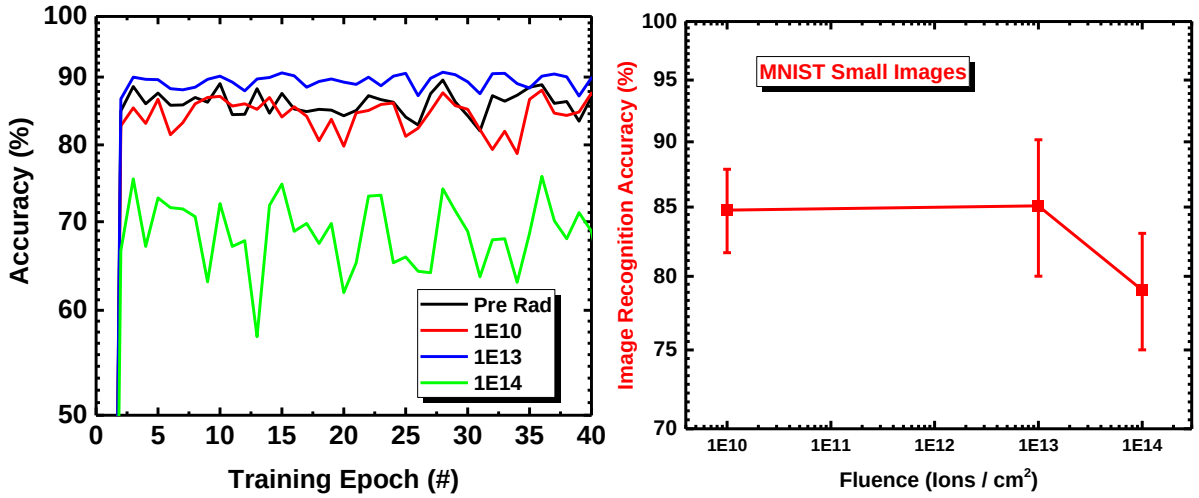

Figure 6: Accuracy of training on devices exposed to radiation. Left: Image recognition accuracy after each training update. Right: Accuracy after training for three devices taken from the maximum classification accuracy of all epochs.

Sandia's CrossSim code was used to evaluate the effect of radiation on training on an analog ReRAM crossbar [14], following the procedure in [2]. CrossSim takes a statistical device conductance dataset (Fig. 3) and converts it into a lookup table. It uses this data to simulate the behavior of the backpropagation of error neural algorithm training network weights based on analog ReRAM cells. A standardized dataset of handwritten digit images known as MNIST, was used to train the network [18]. CrossSim created a dataset from the cycles following each shot and was used to simulate a plot of classification accuracy (correctly identifying the handwritten digit) versus the neural network iteration (epoch) for a sample device (Fig. 6). While the device largely continues to switch after the shot with 10^{14} ions / cm^2 fluence, subtle changes to the analog characteristics result in a somewhat lower accuracy after training than in the non-irradiated device. After a large dose of radiation, the neural network would also need to be retrained, but the fact that it can be trained online again is an advantage in continuous learning applications. The very high fluence levels required to impact the accuracy after training of a neuromorphic accelerator based on TaO_x RRAM devices implies applicability of this technology to embedded systems in radiation environments, such as spacecraft environments.

The full paper will include additional measurements which have been already preformed on three similar devices with in-line 4 kohm resistors, as well as three control samples which were exposed to a similar environment during tests, sans radiation. Data from three devices of each sample type held in the low resistance state during irradiation, and a TID test consisting of three devices of each type in ^{60}Co γ-ray exposure which has already been performed will be included. Furthermore, the classification accuracy following training for these devices will be evaluated with additional training datasets.

IV. Summary

The effect of displacement damage on neural network accuracy after training on an analog-crossbar based on TaO_x ReRAM array is investigated. It is found that a neural network training accelerator based on this technology is robust to ion displacement damage, requiring a shot of 2.5 MeV Si ions with fluence of 10^{14} ions / cm^2 to show a significant effect on the classification accuracy after training. At this dose, the device conductance is found to shift. The devices were still functionally switching, and although subtle changes to their analog conductance response were observed, neural training was still possible. SRIM calculations confirm that displacement damage is occurring in the oxide. It appears that ReRAM may be more robust to displacement damage when used as an analog neural network weight than as a digital memory, due to the inherent fault tolerance of these algorithms. This study gives a promising direction for embedded pattern recognition systems for space applications.

V. References

1. Y. LeCun, et al. *Nature.*, vol. 521, no. 7553, pp. 436-444, May 2015.
2. M. J. Marinella, et al., *IEEE JETCAS*., early access, doi: 10.1109/JETCAS.2018.2796379
3. S. Agarwal, et al. *Front. Neurosci.*, vol. 9, pp. 484, Jan. 2016.
4. M. J. Marinella, et al., *IEEE Trans. Nucl. Sci.*, vol. 59, pp. 2987-2994, Dec. 2012.
5. D. R. Hughart, et al., *IEEE Trans. Nucl. Sci.*, vol. 60, pp. 4512-4519, Dec. 2013.
6. W.M. Tong, et al., *IEEE Trans. Nucl. Sci.*, vol. 57, pp. 1640, June 2010.
7. H. J. Barnaby, et al., *IEEE Trans. Nucl. Sci.*, vol. 58, pp. 2838, Dec. 2011.
8. Y. Wang, et al., *Elec. Dev. Lett.*, vol 31, no. 12, pp. 1470-1472, Dec. 2010.
9. X. He, et al., *IEEE Trans. Nucl. Sci.*, vol. 59, no. 5, pp. 2550-2555, Oct. 2012.
10. J. S. Bi, et al., *IEEE Trans. Nucl. Sci.*, vol. 60, pp. 4540-4546, Dec. 2013.
11. P. R. Mickel, et al., *Appl. Phys. Lett.*, vol. 102**,** 223502 (2013).
12. F. Miao, et al., *Adv. Mater.*, 23, pp. 5633-5640 2011.
13. J. J. Yang, et al., *Nat. Nanotech.*, vol. 8, 13-24 (2013).
14. J. E. Stevens, et al., *Vac. Sci. Technol., A*, vol. 32, no. 2, 021501, Oct. 2013.
15. S. A. Agarwal, *CrossSim 2017 [Online]*. Available: https://cross-sim.sandia.gov/
16. D. R. Hugart, et al., *IEEE Trans. Nucl. Sci.*, vol 60, no. 6, 4512-4519, Dec. 2013.
17. J. F. Ziegler, *SRIM-2012 [Online]*. Available: www.srim.org
18. Y. LeCun et al. *AT&T Labs [Online]* Available: http://yann.lecun.com/exdb/mnist