Chai, Z, Freitas, P, Zhang, WD, Hatem, F, Degraeve, R, Clima, S, Zhang, J, Marsland, J, Fantini, A, Garbin, D, Goux, L and Kar, G

Stochastic computing based on volatile GeSe ovonic threshold switching selectors

Stochastic computing based on volatile GeSe ovonic threshold switching selectors

Zheng Chai, Pedro Freitas, Weidong Zhang, Firas Hatem, Robin Degraeve, Sergiu Clima, Jian Fu Zhang, John Marsland, Andrea Fantini, Daniele Garbin, Ludovic Goux, and Gouri Sankar Kar

***Abstract*—Stochastic computing (SC) is a special type of digital compute strategy where values are represented by the probability of 1 and 0 in stochastic bit streams, which leads to superior hardware simplicity and error-tolerance. In this paper, we propose and demonstrate SC with GeSe-based Ovonic Threshold Switching (OTS) selector devices by exploiting their probabilistic switching behavior. The stochastic bit streams generated by OTS are demonstrated with good computation accuracy in both multiplication operation and image processing circuit. Moreover, the bit distribution has been statistically studied and linked to the collective defect de/localization behavior in the chalcogenide material. Weibull distribution of the delay time supports the origin of such probabilistic switching, facilitates further optimization of the operation condition, and lays the foundation for device modelling and circuit design. Considering its other advantages such as simple structure, fast speed, and volatile nature, OTS is a promising material for implementing SC in a wide range of novel applications, such as image processors, neural networks, control systems and reliability analysis.**

***Index Terms*—selector, GeSe, stochastic computing, memristor, bit stream, random number**

I. Introduction

Stochastic computing (SC) is an approximate computing paradigm where a value is represented by the probability encoded in a stream of stochastic bits [1], [2]. It enables low-cost implementations of arithmetic operations using standard logic elements, and also provides high tolerance to soft errors. Due to these advantages, SC is particularly suitable for those applications requiring parallel-processing techniques, such as image processing [3], [4], neural networks [5]-[7]and control systems [1], [8]. Despite these benefits, the performance of SC is limited by the quality of bit streams: correlation in the bit streams could dramatically degrade the computing accuracy. **Fig. 1a** demonstrates that a multiplication operation can be simply realized with a single AND gate, but correlated input streams could cause errors and should be avoided (Fig.1b).

The natural probabilistic behavior and error tolerance of SC suits well with emerging nanotechnologies such as memristors [4], [7], [9], carbon nanotube field-effect transistors (CNTFETs) [10], and magnetic-tunnel junction (MTJ) devices [11], [12], by utilizing their probabilistic behavior associated with different mechanisms. However, the non-volatile memristors and MTJ devices need an erase (reset) operation and a separate read-out operation in each programming cycle, which increases operational complexity and energy consumption, and limits the generation frequency; the CNTFETs, although based on a revolutionary technology and have exciting potentials, still face difficulties in achieving large-scale production compatible with silicon-based CMOS.

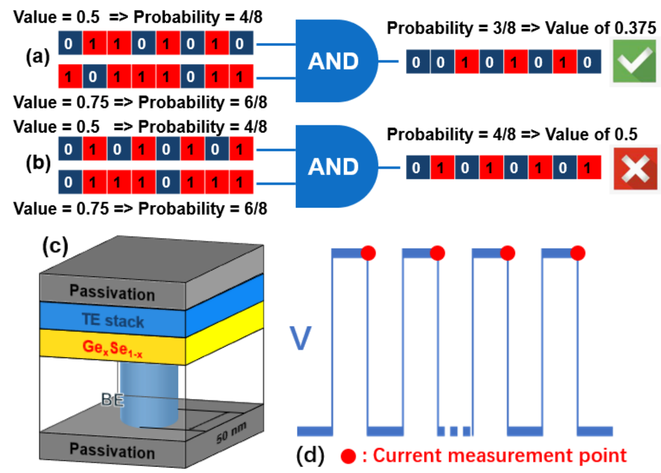

Fig.1. (a) A multiplication operation is implemented with a single AND gate. (b) Correlated input streams could cause errors. (c) Schematic of the OTS structure. Ge_xSe_{1-x} is the switching layer; device size is confined by the pillar bottom electrode (BE). (d) Schematic of the bit stream generation waveform. Current is measured at the end of each pulse.

Resistive-switching memory devices are promising candidates for next-generation memory and neuromorphic computing [13][14]. In recent years, ovonic threshold switching (OTS) chalcogenide materials, such as GeSe, GeAsTe and SiGeAsSe, have been proposed as selector devices used to suppress the sneak current in emerging memory arrays due to their favorable characteristics such as CMOS-compatibility, volatile switching, fast speed and excellent endurance [15]-[17],[27],[28]. OTS switching is an electronic process which involves defect localization/delocalization in a volatile conductive filament formed during the first-fire operation. The detailed switching mechanism has been discussed in our earlier works [17],[25]-[28].

OTS's probabilistic switching has been exploited to implement the true random number generators (TRNGs) with good randomness and stability [18]. In this paper, we propose and demonstrate another novel application: stochastic computing based on GeSe OTS selector devices. The stochastic bit streams generated by OTS have shown good computation accuracy in both multiplication operation and edge detection

This work was supported by the EPSRC of U.K. under EP/M006727/1 & EP/S000259/1). (Corresponding authors: Zheng Chai, Weidong Zhang.) The review of this letter was arranged by Editor Prof. Genquan Han.

Zheng Chai, Pedro Freitas, Weidong Zhang, Firas Hatem, Jian Fu Zhang, and John Marsland are with the Department of Electronics and Electrical Engineering, Liverpool John Moores University, L3 3AF Liverpool, U.K. (e-mail: z.chai@ljmu.ac.uk; w.zhang@ljmu.ac.uk).

Robin Degraeve, Sergiu Clima, Andrea Fantini, Daniele Garbin, Ludovic Goux, and Gouri Sankar Kar are with imec, 3001 Leuven,

circuit for image processing. Moreover, the distribution of bit in the stochastic streams has been statistically studied. Weibull distribution of OTS's delay time is used to explore the origin of such probabilistic switching. Considering OTS's other advantages such as simple structure, fast speed, and volatile nature, it is a promising material for SC implementation in a wide range of novel applications, such as image processors, neural networks, and control systems.

II. Device and Characterization

TiN/GeSe/TiN selector devices were integrated in a 300 mm process flow, using a pillar (TiN) bottom electrode which defines the device size to 50 nm **(Fig.1c)**. The 10 nm-thick amorphous GeSe chalcogenide film was achieved by room temperature physical vapor deposition (PVD) and passivated with a low-temperature BEOL process scheme [15]. Under a constant voltage pulse (**Fig. 1d**), there is a stochastic delay time (t_{delay}) before the device is switched on [19],[25]. Such stochasticity is exploited in **Fig. 1d** by applying a sequence of pulses of certain conditions and measuring the current at the end of each pulse to check whether the device has been switched on ("1") or not ("0") during this pulse. Thus, the stochastic bit stream, whose probability (of "1") represents the value, can be generated. Thanks to the volatile nature of OTS, the generation of a stochastic bit does not require a separate erase (reset) operation or a read-out operation, which could greatly simplify the circuit and improves the bit generation throughput. The fast I-V characterization was done with a Keysight B1500A semiconductor analyzer with embedded B1530A Waveform Generator/Fast Measurement Unit (WGFMU).

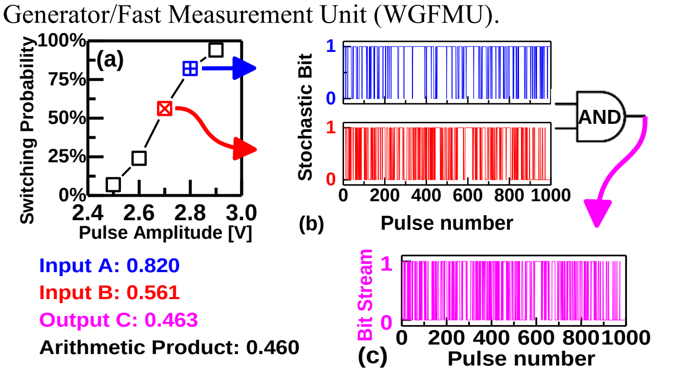

Fig.2. (a) Switching probability at different pulse amplitudes(t_{pulse} = 1μs). (b) 1000-bit stochastic streams generated at 2.7 V and 2.8 V representing the values of 0.561 and 0.820 respectively sent to an AND gate for multiplication. (c) The output bit stream after SC represents a value of 0.463, close to the arithmetic product of 0.460.

III. Results and Discussions

Firstly, stochastic computing is demonstrated in a multiplication operation. With a fixed pulse width, OTS's switching probability is dependent on the pulse amplitude (**Fig. 2a)**. The bit streams generated by the 1,000 pulses at 2.7 V and 2.8 V are sent to an AND gate for multiplication (**Fig. 2b**). After SC, the output bit stream represents a value of 0.463 (**Fig. 2c**), very close to the arithmetic product of 0.460 and supporting the good stochasticity and uncorrelatedness of bit streams generated by OTS.

Looking into the stochastic bit streams, it is interesting to find that the segment length, i.e. the number of consecutively appearing "0" or "1" as demonstrated in **Fig. 3a**, follows exponential distribution **(Fig. 3b)**, which indicates that the stochastic generation process can be considered as a memoryless discrete-time Markov chain [20]. The mean value of the segment lengths can be named as segment length constant τ **(Fig. 4b)**. τ_0 and τ_1 are oppositely dependent on the pulse amplitude (**Fig. 4c**) and pulse width **(Fig. 4d)**.

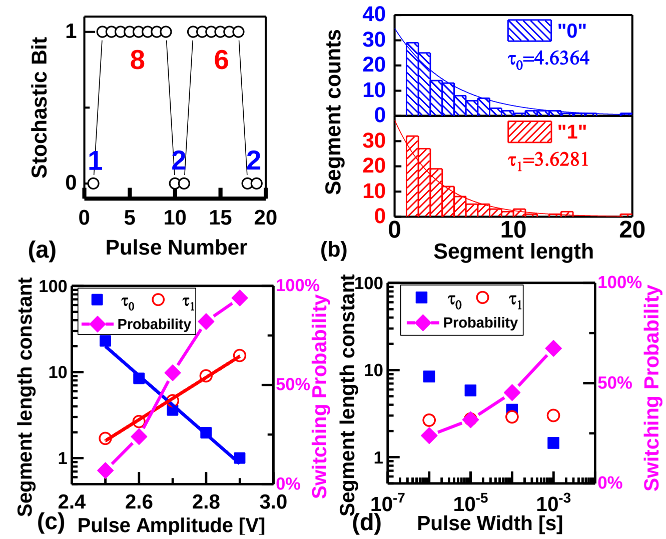

Fig.3. (a) Demonstration of "0" and "1" segments in a 20-bit stream. (b) Exponential distribution of segment length, from the 1,000-bit stream generated with V_{pulse} = 2.7 V and t_{pulse} = 1 μs. (c) Dependence of segment length constants, τ_0 and τ_1, on V_{pulse} with fixed t_{pulse} = 1 μs. The corresponding τ_0 and τ_1 of any analog probability value can be obtained via interpolation (straight lines). (d) Dependence of segment length constants on t_{pulse} with fixed V_{pulse} = 2.6 V.

Such segment length information facilitates the modelling of bit streams of any probability values and further the simulation of an OTS-based edge detection circuit in an image processing system. Such system uses an OTS array to convert pixel values into stochastic bit streams in parallel **(Fig. 4a)**, which are further processed by the Robert cross algorithm **(Fig. 4b)** to highlight significant gradients in the diagonal direction across the array. The array and circuit have been reported in [4]. In this way, image edge can be detected. **Fig. 4(c-d)** compares the edge detection result with streams of 100 and 1,000 bits respectively. Edge detection is successful and increasing the stream length can significantly improve detection quality.

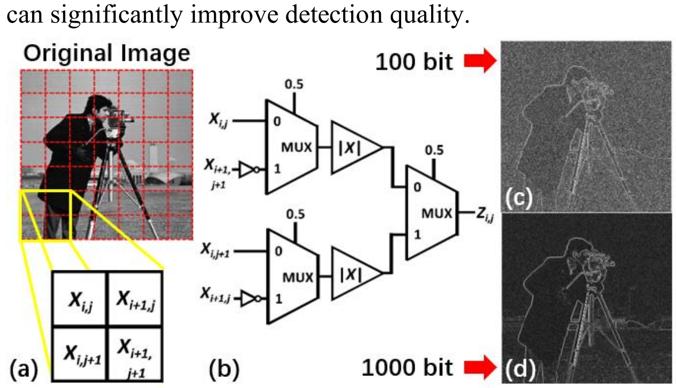

Fig.4. (a) A 256x256 image for edge detection with schematic pixel array. (b) Edge detection circuit based on Robert cross algorithm. $X_{i,j}$ is the stochastic bit stream generated by OTS, representing the pixel value at (i,j) in the array.(c-d) Edge detection results using (c) 100-bit and (d) 1000-bit stochastic streams.

The origin of such excellent stochasticity of OTS-based SC can be attributed to the stochastic delay time during switch-on, as demonstrated in **Fig. 5a.** It is further found that in a wider pulse, the delay time measured within the measurement range (10 ns ~ 50 μs) follows Weibull distributions at different pulse amplitudes **(Fig. 5b)**. It is well known that the time-dependent dielectric breakdown (TDDB) follows the Weibull distribution and is triggered by the formation of a filamentary conductive percolation path [23][25]. The Weibull distribution of t_{delay} in OTS can be attributed to a volatile filamentary formation process, therefore, induced by a different mechanism such as electronic defect de/localization [25-27].

The dependence of OTS switching probability on both pulse amplitude and width is further investigated. In **Fig. 5b**, t_{delay} is reduced at higher pulse amplitude, as the Weibull distribution parallel shifts to left. This demonstrates that a pulse with either higher amplitude or longer width can increase the switching probability measured at the end of the pulse. The details of the Weibull plot can be found in [25] [26]. The Weibull parameters, α and β, in **Fig. 6a**, are extracted from **Fig. 5b** and can be fitted well linearly with pulse amplitude. Based on this observation, the switching probability at a wide range of pulse conditions, i.e. amplitude and/or width, can be simulated, as shown in the heatmap of **Fig. 6c,** where the scattered colored dots are experimental switching probability measured by applying 1,000 pulses at the corresponding conditions. The good agreement supports that whilst the switching of OTS is stochastic, the switching probability can be precisely controlled by either tuning the pulse amplitude or width in a wide range, which provides further flexibility for its SC application.

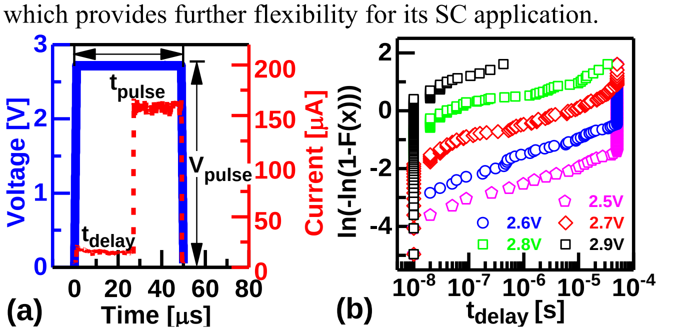

Fig.5. (a) Current of OTS during a constant voltage pulse (50 μs, 2.7V), with probabilistic the delay time (t_{delay}) before switch-on. (b) Weibull plot of t_{delay} under various pulse amplitudes. The t_{delay} is limited by the measurement resolution (10 ns) and pulse width of 50 μs

A practical OTS-based SC system will be challenged by a range of reliability issues, such as switching probability drift induced by cycling and device-to-device (D2D) variability. The endurance of GeSe OTS has been significantly improved by introducing a recovery scheme which prolongs the endurance to >10^{11} [17], while SiGeAsTe OTS has demonstrated a stable endurance performance >10^{11} [27], which is limited by the measurement instrument. The D2D variability is within a few tens [27] or hundreds of millivolts [18]. Whilst SC has demonstrated good robustness against reliability issues thanks to its error-tolerate nature [2], these issues can be further migrated by solutions at the peripheral circuitry level, such as a real-time switching probability monitor circuit utilizing counters/comparators to adjust and map the input pulse conditions accordingly. In addition, several novel programing methods have been proposed to improve the reliability of OTS selectors, such as controlling the fall time/operation current [28], or using the "1/2V" scheme [29]. The detailed circuit design is out of the scope of this paper, as it aims to timely propose the concept of OTS-based SC. Additionally, the switch-off process of OTS is also a probabilistic process but at a much faster speed [25], [26], which could be further exploited in future work to enhance the performance of OTS based SC.

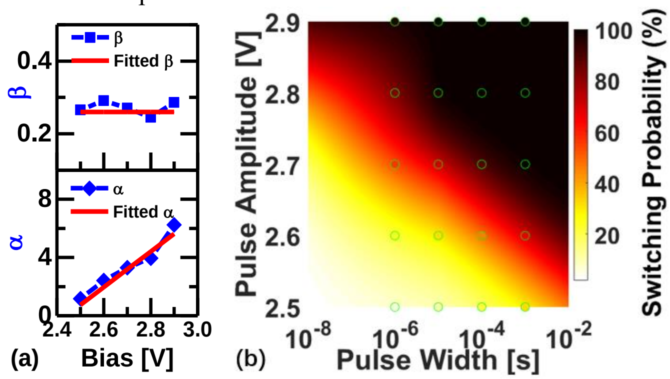

Fig.6. (a) Extracted Weibull parameters at different pulse amplitudes. (b) Simulated dependence of switching probabilities on pulse amplitude and width. Experimental data (scatters "o" with filled colours) indicates the measured probability.

IV. Conclusions

In this paper, we propose and demonstrate stochastic computing with GeSe-based OTS selector device by exploiting its probabilistic switching behavior. The high-quality stochastic bit streams generated by OTS lead to good computation accuracy in both multiplication operation and edge detection circuit for image processing. Moreover, the bit distribution in the stochastic streams has been statistically studied and linked to the collective defect localization/delocalization behavior in the chalcogenide material. Weibull distribution of the delay time supports the filamentary origin of such probabilistic switching, facilitates optimization of operation conditions, and lays the foundation for device modelling and circuit design. The simple structure, fast speed, and volatile nature of OTS make it promise for stochastic computing in a wide range of novel applications, such as image processors and neural networks.

Acknowledgements

The authors would like to thank Dr. Yudi Zhao and Dr. Peng Huang at Peking University for their valuable assistance.

References

[1] B.R. Gaines, Stochastic Computing Systems, *Advances in Information Systems Science*, Boston, MA, Springer, pp. 37-172, 1969, DOI: 10.1007/978-1-4899-5841-9_2

[2] A. Alaghi, The Logic of Random Pulses: Stochastic Computing, University of Michigan, Ph.D Dissertation, 2015

[3] P. Li, D.J. Lilja, W. Qian, K. Bazargan and M. D. Riedel, Computation on Stochastic Bit Streams: Digital Image Processing Case Studies, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol.

22, no. 3, pp. 449 - 462, Apr. 2014 DOI: 10.1109/TVLSI.2013.2247429

[4] Y. Zhao, W. Shen, P. Huang, W. Xu, M. Fan, X. Liu and J. Kang, A Physics-based Model of RRAM Probabilistic Switching for Generating Stable and Accurate Stochastic Bit-streams, *in IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2019 DOI: 10.1109/IEDM19573.2019.8993559

[5] J. Yu, K. Kim, J. Lee and K. Choi, Accurate and Efficient Stochastic Computing Hardware for Convolutional Neural Networks, *IEEE 35th International Conference on Computer Design*, Boston, MA, USA, Nov. 2017, DOI: DOI: 10.1109/ICCD.2017.24

[6] A. Ardakani, F. Leduc-Primeau, N. Onizawa, T. Hanyu and W.J. Gross, VLSI Implementation of Deep Neural Network Using Integral Stochastic Computing, *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 25, no. 10, pp. 2688 - 2699, Feb. 2017, DOI: 10.1109/TVLSI.2017.2654298

[7] M. Suri, D. Querlioz, O. Bichler, G. Palma, E. Vianello, D. Vuillaume, C. Gamrat, and B. DeSalvo, Bio-Inspired Stochastic Computing Using Binary CBRAM Synapses, *IEEE Trans. Electron Devices*,vol. 60, no. 7, pp. 2402-2409, July 2013, DOI: 10.1109/TED.2013.2263000

[8] D. Zhang and H. Li, A Stochastic-Based FPGA Controller for an Induction Motor Drive With Integrated Neural Network Algorithms, *IEEE Trans. Ind. Electron.*, vol. 55, no. 2, pp. 551-561, Feb. 2008, DOI: 10.1109/TIE.2007.911946

[9] S. Gaba, P. Knag, Z. Zhang, and W. Lu, Memristive devices for stochastic computing, *IEEE International Symposium on Circuits and Systems (ISCAS),* Melbourne VIC, Australia, July 2014, DOI: 10.1109/ISCAS.2014.6865703

[10] M.M. Shulaker, G. Hills, N. Patil, H. Wei, H.-Y. Chen, H.-S.P. Wong and S. Mitra, Carbon nanotube computer, *Nature*, vol. 501, pp. 526–530, 2013, DOI: 10.1038/nature12502

[11] Y. Lv and J.-P. Wang, A single magnetic-tunnel-junction stochastic computing unit, *in IEDM Tech. Dig.,* San Francisco, CA, USA, Jan. 2017, DOI: 10.1109/IEDM.2017.8268504

[12] H. Lee, A. Lee, F. Ebrahimi, P.K. Amiri and K.L. Wang, Analog to Stochastic Bit Stream Converter Utilizing Voltage-Assisted Spin Hall Effect, *IEEE Electron Device Lett.*, vol. 38, no. 9, pp. 1343-1346, Sept. 2017, DOI: 10.1109/LED.2017.2730844

[13] T.-Y. Wang, J.-L. Meng, M.-Y. Rao, Z.-Y. He, L. Chen, H. Zhu, Q.-Q. Sun, S.-J. Ding, W.-Z. Bao, P. Zhou, and D.W. Zhang, Three-Dimensional Nanoscale Flexible Memristor Networks with Ultralow Power for Information Transmission and Processing Application, Nano Lett. vol. 20, pp. 4111−4120, 2020, DOI: 10.1021/acs.nanolett.9b05271

[14] T.-Y. Wang, J.L. Meng, Z.-Y. He, L. Chen, H. Zhu, Q.Q. Sun, S.J. Ding, P. Zhou, D.W. Zhang, Wearable Heterosynapses: Ultralow Power Wearable Heterosynapse with Photoelectric Synergistic Modulation, vol. 7, no. 8, April 2020, DOI: 10.1002/advs.202070041

[15] B. Govoreanu, G.L. Donadio, K. Opsomer, W. Devulder, V.V. Afanas' ev, T. Witters, S. Clima, N.S. Avasarala, A. Redolfi, S. Kundu, O. Richard, D. Tsvetanova, G. Pourtois, C. Detavemie, L. Goux and G. S. Kar, Thermally stable integrated Se-based OTS selectors with >20 MA/cm^2 current drive, >3.10^3 half-bias nonlinearity, tunable threshold voltage and excellent endurance, *in VLSI Symp. Tech. Dig.*, Kyoto, Japan, Aug. 2017, DOI: 10.23919/VLSIT.2017.7998207

[16] N.S. Avasarala, G. L. Donadio, T. Witters, K. Opsomer, B. Govoreanu, A. Fantini, S. Clima, H. Oh, S. Kundu, W. Devulder, M. H. van der Veen, J. Van Houdt, M. Heyns, L. Goux and G. S. Kar, Half-threshold bias Ioffreduction down to nA range of thermally and electrically stable high-performance integrated OTS selector, obtained by Se enrichment and N-doping of thin GeSe layers, *in VLSI Symp. Tech. Dig.,* Honolulu, HI, USA, June 2018, DOI: 10.1109/VLSIT.2018.8510680

[17] F. Hatem, Z. Chai, W. Zhang, A. Fantini, R. Degraeve, S. Clima, D. Garbin, J. Robertson, Y. Guo, J.F. Zhang, J. Marsland, P. Freitas, L. Goux, and G.S. Kar, Endurance improvement of more than five orders in Ge_xSe_{1-x} OTS selectors by using a novel refreshing program scheme, *in IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2019, DOI: 10.1109/IEDM19573.2019.8993448

[18] Z. Chai , W. Shao , W. Zhang , J. Brown, R. Degraeve, F.D. Salim , S. Clima , F. Hatem , J. F. Zhang , P. Freitas , J. Marsland, A. Fantini, D. Garbin , L. Goux, and G.S. Kar, GeSe-Based Ovonic Threshold Switching Volatile True Random Number Generator, *IEEE Electron Device Lett.*, vol. 41, no. 2, pp. 228-231, Feb. 2020, DOI: 10.1109/LED.2019.2960947

[19] W. Czubatyj and S.J. Hudgens, Thin-Film Ovonic Threshold Switch: Its Operation and Application in Modern Integrated Circuits, *Electron. Mater. Lett.*, vol. 8, no. 2, pp. 157-167, April 2012, DOI: 10.1007/s13391-012-2040-z

[20] Samuel Karlin, A First Course in Stochastic Processes, Cambridge, MA, Academic Press, 1968, DOI: 10.1016/C2013-0-12346-X

[21] T. Grasser, Stochastic charge trapping in oxides: From random telegraph noise to bias, *Microelectron Reliab*, vol. 52, no. 1, pp. 39-70, Jan. 2011, DOI: 10.1016/j.microrel.2011.09.002

[22] M.J. Kirton and M.J. Uren, Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency $(1/f)$ noise, *Adv. Phys*, vol. 38, no. 4, pp. 367-468, Nov. 1989, DOI: 10.1080/00018738900101122

[23] E. Y. Wu and R.-P. Vollertsen, On the Weibull Shape Factor of Intrinsic Breakdown of Dielectric Films and Its Accurate Experimental Determination-Part I: Theory Methodology Experimental Techniques, *IEEE Trans. Electron Devices*, vol. 49, no. 12, pp. 2131-2140, Dec. 2002, DOI: 10.1109/TED.2002.805612

[24] M. Wimmer and M. Salinga, The gradual nature of threshold switching, *New J. Phys.*, vol. 16, p. 113044, Nov. 2014, DOI: 10.1088/1367-2630/16/11/113044

[25] Z. Chai, W. Zhang, R. Degraeve, S. Clima, F. Hatem, J. F. Zhang, P. Freitas, J. Marsland, A. Fantini, D. Garbin, L. Goux and G.S. Kar, Evidence of filamentary switching and relaxation mechanisms in Ge_xSe_{1-x} OTS selectors, *in VLSI Symp. Tech. Dig.*, Kyoto, Japan, June 2019, DOI: 10.23919/VLSIT.2019.8776566

[26] Z. Chai, W. Zhang, R. Degraeve, S. Clima, F. Hatem, J.F. Zhang, P. Freitas, J. Marsland, A. Fantini, D. Garbin, L. Goux, and G.S. Kar, Dependence of switching probability on operation conditions in Ge_xSe_{1-x} OTS selectors, *IEEE Electron Device Lett.*,vol. 40, no. 8, pp. 1269-1272, Aug. 2019, DOI: 10.1109/LED.2019.2924270

[27] D. Garbin, M. Pakala, A. Cockburn, C. Detavernier, R. Delhougne, L. Goux, G.S. Kar, W. Devulder, R. Degraeve, G.L. Donadio, S. Clima, K. Opsomer, A. Fantini, D. Cellier, W.G. Kim, Composition Optimization and Device Understanding of Si-Ge-As-Te Ovonic Threshold Switch Selector with Excellent Endurance, *in IEDM Tech. Dig.*, San Francisco, CA, USA, Dec. 2019, DOI: 10.1109/IEDM19573.2019.8993547

[28] S. Kabuyanagi, D. Garbin, A. Fantini, S. Clima, R. Degraeve, G.L. Donadio, W. Devulder, R. Delhougne, D. Cellier, A. Cockburn, W.G. Kim, M. Pakala, M. Suzuki, L. Goux and G.S. Kar, Understanding of Tunable Selector Performance in Si-Ge-As-Se OTS Devices by Extended Percolation Cluster Model Considering Operation Scheme and Material Design, *in VLSI Symp. Tech. Dig.*, Virtual conference, June 2020

[29] N. Gong, W. Chien, Y. Chou, C. Yeh, N. Li, H. Cheng, C. Cheng, I. Kuo, C. Yang, R. Bruce, A.Ray, L.Gignac, Y. Lin, C. Miller, T. Perri, W. Kim, L. Buzi, H. Utomo, F. Carta, E. Lai, H. Ho, H. Lung, and M. BrightSky, A No-Verification Multi-Level-Cell (MLC) Operation in Cross-Point OTS-PCM, *in VLSI Symp. Tech. Dig.*, Virtual conference, June 2020