Applied Physics Letters

Physical based compact model of Y-Flash memristor for neuromorphic computation

Cite as: Appl. Phys. Lett. **119**, 263504 (2021); https://doi.org/10.1063/5.0069116
Submitted: 29 August 2021 • Accepted: 08 December 2021 • Published Online: 28 December 2021

Wei Wang, Loai Danial, Eric Herbelin, et al.

AIP Publishing

Applied Physics Letters ARTICLE scitation.org/journal/apl

Physical based compact model of Y-Flash memristor for neuromorphic computation

Cite as: Appl. Phys. Lett. **119**, 263504 (2021); doi: 10.1063/5.0069116
Submitted: 29 August 2021 · Accepted: 8 December 2021 ·
Published Online: 28 December 2021

Wei Wang,[1,a)] Loai Danial,[1,b)] Eric Herbelin,[1] Barak Hoffer,[1] Batel Oved,[1] Tzofnat Greenberg-Toledo,[1] Evgeny Pikhay,[2] Yakov Roizin,[2] and Shahar Kvatinsky[1,a)]

AFFILIATIONS

[1]The Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering, Technion–Israel Institute of Technology, Haifa 3200003, Israel
[2]Tower Semiconductor, Migdal HaEmek 2310502, Israel

Note: This paper is part of the APL Special Collection on Neuromorphic Computing: From Quantum Materials to Emergent Connectivity.
[a)]**Authors to whom correspondence should be addressed:** wei.wang@campus.technion.ac.il and shahar@ee.technion.ac.il
[b)]**Current address:** Intel Corporation, IDC, Haifa, Israel.

ABSTRACT

Y-Flash memristors utilize the mature technology of single polysilicon floating gate nonvolatile memories. It can be operated in a two-terminal configuration similar to the other emerging memristive devices, e.g., resistive random-access memory and phase-change memory. Fabricated in production complementary metal-oxide-semiconductor technology, Y-Flash memristors allow excellent reproducibility reflected in high neuromorphic products yields. Working in the subthreshold region, the device can be programmed to a large number of fine-tuned intermediate states in an analog fashion and allows low readout currents (1 nA ∼ 5 μA). However, currently, there are no accurate models to describe the dynamic switching in this type of memristive device and account for multiple operational configurations. In this paper, we provide a physical-based compact model that describes Y-Flash memristor performance in both DC and AC regimes and consistently describes the dynamic program and erase operations. The model is integrated into the commercial circuit design tools and is ready to be used in applications related to neuromorphic computation.

Emerging nonvolatile memory devices (NVMs), such as resistive random-access memory (RRAM),[1] phase-change memory (PCM),[2] ferroelectric random-access memory (FeRAM),[3] and electrolyte-gate transistors,[4,5] have tunable conductance, which can faithfully emulate the plasticity of synaptic connections in an artificial neural network and, thus, are promising for the large-scale implementation of neuromorphic computations.[6–8] However, these emerging technologies are still not mature enough for production-worthy neuromorphic systems, mostly because of low yields and high non-uniformity of the memristive devices.[9–12] Conversely, NVM floating gate (FG) devices are fabricated in standard complementary metal–oxide–semiconductor (CMOS) processes. They have also been proposed as candidates for artificial synapses.[13,14] As its name suggests, the FG device has a floating gate isolated from its other terminals. The FG can be charged or discharged; thus, the threshold of the corresponding transistor can be adjusted to the desired level.

Y-Flash devices have a simpler configuration compared with conventional FG devices: The control gate is merged with drain, thus largely decreasing the device footprint, while program, erase, and readout operations are possible in two-terminal configuration.[15] Arranged in a crossbar array, the Y-Flash devices are suitable for accelerating the frequent and expensive vector-matrix multiplications (VMMs) acting as a synaptic array for multiple neuromorphic applications.[16] The Y-Flash memristor works in an analog fashion and can be operated in the subthreshold range, achieving a large number of tunable conductance states and low readout currents. Compared with other NVMs, the Y-Flash memristor shows a combination of several advantages, including fully CMOS process compatibility, low cycle-to-cycle variations, high yield, low power consumption, analog conductance tunability, self-selection, and high retention time.

The success of a neuromorphic computation system needs the co-design of the synaptic array (e.g., Y-Flash device array) and the CMOS-based peripheral circuit, i.e., artificial neurons.[17] Thus, a compact model, which can accurately simulate the synaptic characteristics of the Y-Flash device to enable the full design flow of the neuromorphic computation system, is needed. Previously, we reported a

preliminary compact model, which could capture the basic program, erase, and readout operations for a fixed configuration and operation voltage based on empirical equations.[18] However, the accurate I-V characteristics, as well as the program and erase operations at different configurations and voltage biases, could not be simulated by the preliminary model (see supplementary material, Table S1). The preliminary model did not support transient analyses either, since the effects of parasitic capacitors, as well as other parasitic effects, were not considered.

In this paper, we provide a comprehensive model with in-depth physical considerations related to the program/erase mechanism and the account for the parasitic effects. The FG voltage is determined by the analysis of the coupled capacitor network, and the I-V characteristics consider the behaviors in both the subthreshold and above-threshold regions. The program operation is modeled by the lucky hot electron injection, and the erase operation is described as a combination of band-to-band tunneling (BBT) of holes, their acceleration in the lateral field, and tunneling through the dielectric layer. The simulation results show excellent agreement with the experimental data. The model is written in Verilog-A language and is integrated into a commercial circuit design simulator to enable the co-design of a neuromorphic system with Y-Flash memristors and CMOS peripheral circuits.

The structure of the Y-Flash device is shown in Fig. 1(a), which consists of two transistors, i.e., a read transistor and an injection transistor, with a common FG and a common drain (D). The read transistor (channel length = 0.6 μm) has a lower threshold voltage by a different channel doping optimized for the readout, while the injection transistor has a shorter channel (0.3 μm) to enhance the hot carrier injection. The capacitance of the common D is larger than the source of read transistor (Source Read, SR) and the source of injection transistor (Source Injection, SI) to couple the FG more on the D terminal. Figures 1(b) and 1(c) show the equivalent circuit, accounting for parasitic capacitors and p-n junctions between substrate and terminals, and the symbol of the Y-Flash device, respectively. The FG is isolated from the external terminals and can store charge [Q_{FG} in Fig. 1(b)], and the substrate is always grounded. The device was fabricated using standard CMOS process flow (Tower Semiconductor 180 nm) without any additional masks. More details can be found in our previous publication.[16]

Although the Y-Flash device has three external terminals, the two sources, SR and SI, can be shortened (Nos. 1, 3, and 6 in Table I), such that the device acts as a two-terminal memristor. The two-terminal Y-Flash configuration enables easier addressing by the peripheral circuits, while the three-terminal configuration enables more flexible and efficient read, program, and erase operations (Nos. 2, 4, and 7 in Table I). Read operation is conducted by applying a low read voltage ($V_R < 2.5$ V) on the D of the device (Nos. 1 and 2 in Table I). Similar to other memristive arrays, VMM can be performed directly in a Y-Flash device array leveraging Ohm's law and Kirchhoff's current law [Fig. 1(d)]. At a high drain voltage ($V_P > 4$ V) and with one or two of the source terminals grounded, there will be hot electron injection into the FG, and the threshold voltage of both transistors will be increased, corresponding to device programming. The program will be disabled if both the SR and SI are floating (No. 5 in Table I), which can be used to deselect the devices that are not intended to be programmed in an array [Fig. 1(e)]. The erase operation (decrease in the threshold voltage, or discharge of the floating gate) can be achieved by applying a

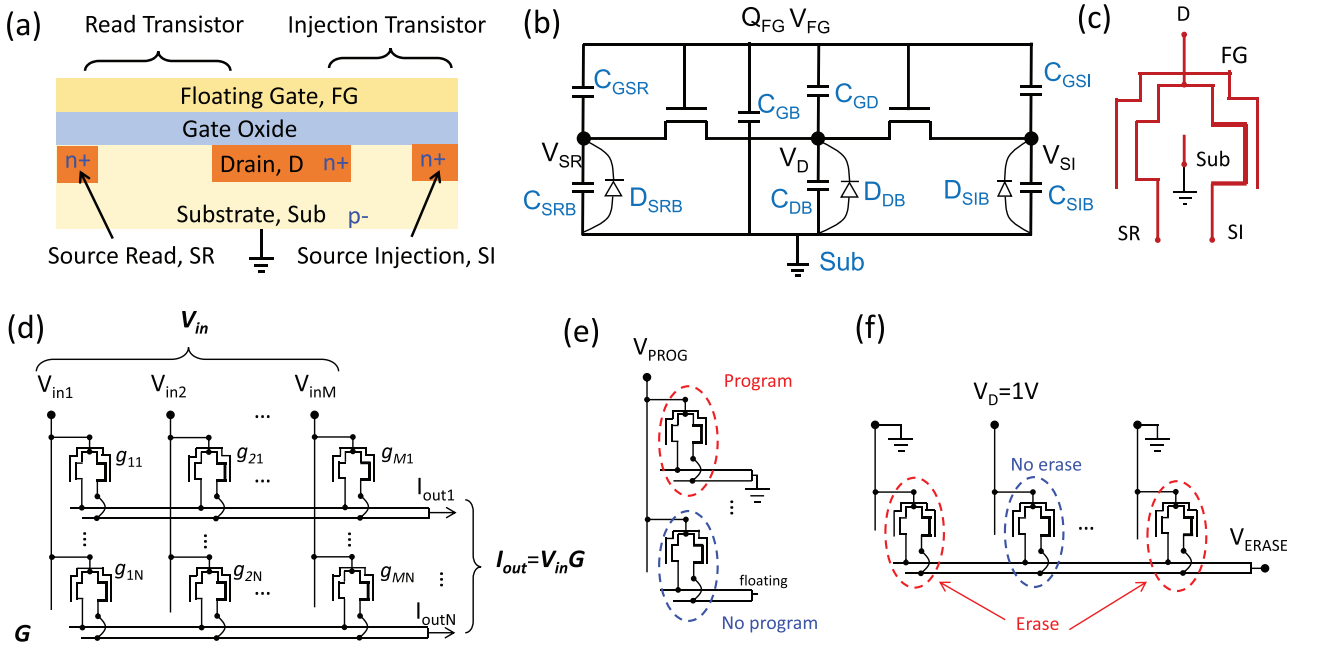

FIG. 1. (a) Schematic of the Y-Flash structure; (b) equivalent circuit accounting for parasitic capacitors, transistors, and p-n diodes; (c) the symbol of the Y-Flash device with external terminals; (d) Y-Flash array for VMM; (e) programming of the devices in an array; (f) erasing of the devices in an array.

TABLE I. Operation configurations of the Y-Flash devices.

No.	Mode	D	SR	SI
1	Read	$V_R < 2.5\,V$	GND	GND
2		$V_R < 2.5\,V$	GND	Floating
3	Program	$V_P > 4\,V$	GND	GND
4		$V_P > 4\,V$	Floating	GND
5		$V_P > 4\,V$	Floating	Floating
6	Erase	Floating/GND	$V_E > 7\,V$	$V_E > 7\,V$
7		Floating/GND	Floating/GND	$V_E > 7\,V$
8		$1\,V < V_D < 2\,V$	Floating/GND	$V_E > 7\,V$

high voltage ($V_E > 7\,V$) on the SI terminal and leaving the drain terminal floating or grounded (Nos. 6 and 7 in Table I). The devices within the same row, which, however, are not intended to be erased, can be deselected by applying a certain voltage on the drain [No. 8 in Table I and Fig. 1(f)].

The I–V characteristics of the device can be modeled by combining the characteristics of the two transistors with a certain charge in the FG. The potential on the FG is not directly given, and in several operation modes, as indicated in Table I, some terminals are floating. These unknown potentials can be obtained by calculating voltages in the circuit of parasitic capacitors, p-n junctions, and equivalent resistors of the transistor channels. When all the three terminals are externally connected to voltage sources, the FG potential can be obtained by solving the equation of the fixed total charge in the floating gate,

$$\begin{aligned} Q_{FG} = {} & C_{GSR}(V_{FG} - V_{SR}) + C_{GD}(V_{FG} - V_D) \\ & + C_{GSI}(V_{FG} - V_{SI}) + C_{GB}V_{FG}, \end{aligned} \tag{1}$$

where C_{GSR}, C_{GD}, C_{GSI}, and C_{GB} are the capacitors between FG and SR, FG and D, FG and SI, and FG and substrate, respectively, and, V_{SR}, V_D, and V_{SI} are the voltages on SR, D, and SI, respectively. When one or two of the three external terminals are floating, their potentials are obtained from the circuit shown in Fig. 1(b). Note that the solution corresponding to the circuit in Fig. 1(b) can be obtained by the circuit simulator directly when the circuit is written in hardware description languages.

To model the wide range of operating voltages of the Y-Flash device, the I–V characteristics in both subthreshold and above-threshold regions should be accurately described. The source–drain current in the subthreshold region ($V_{FG} < V_{TH}$) for both transistors can be written as

$$I_{DS,sub} = I_{S0} e^{q\frac{V_{FG}-V_{TH}}{nkT}} \left(1 - e^{-q\frac{V_D - V_S}{kT}}\right), \tag{2}$$

where I_{S0} is a pre-factor, V_{TH} is the intrinsic threshold voltage, q is the elementary charge, n is the ideality factor, k is the Boltzmann constant, and T is the temperature. Note that the parameters I_{S0} and V_{TH} are different for the two transistors, and V_S is the voltage on the source terminal, which should be replaced by V_{SR} and V_{SI} for the read transistor and injection transistor, respectively. Above threshold ($V_{FG} > V_{TH}$), the source–drain current can be written as

$$I_{DS,ab} = \begin{cases} \dfrac{K}{2}(V_{FG} - V_{TH})^2, & V_{FG} - V_{TH} < V_{DS}, \\ K\left(V_{FG} - V_{TH} - \dfrac{V_{DS}}{2}\right)V_{DS}, & V_{FG} - V_{TH} > V_{DS}, \end{cases} \tag{3}$$

where $K = \frac{W}{L}\mu C_{ox}$, W is the channel width, L is the channel length, μ is the mobility, and C_{ox} is the gate oxide capacitance. Note that the V_{FG} is the function of the D voltage, as shown in the band diagram for reading operation in the supplementary material, Figs. S1(a) and S1(b). The current should be continuous between the two regions. Thus, the two currents in Eqs. (2) and (3) are extended to the full voltage range and combined by a smooth function,

$$I_{DS} = \left(\frac{1}{I_{DS,sub}^m} + \frac{1}{I_{DS,ab}^m}\right)^{-\frac{1}{m}}, \tag{4}$$

where m is a smooth factor ($m = 1$ is used in our model).

Figures 2(a) and 2(b) show the comparison between the measured DC read currents of the device and fitting lines by the model, in logarithmic and linear scales, respectively. Accounting for the parasitic capacitors in the model, the overshoot effects in the pulse measurements can be also simulated in the transient mode [Figs. 2(c) and 2(d)]. The device states can also be read with the SI or SR being floating (see the supplementary material, Fig. S2 for the model results of additional read configurations).

Note that it is only possible to operate the device with positive voltages since the substrate is always grounded and there are p-n junctions between the p-type substrate and the n+ type Ohmic drain/source contacts. Negative reading of the device, however, can be performed by grounding the D and applying a positive voltage on the SR and SI, as shown in the supplementary material, Fig. S3. The negative reading shows ultra-low current since the floating gate is coupled to the D and the transistors are turned off. The low negative reading current enables the self-selection and low sneak path current of the devices in an array.

In the following measurements and simulations, we only grounded the SI terminal in the program mode [Fig. 3(a)]. The program operation (injections of the electron to the FG) is conducted by the lucky hot carriers. We use a simplified model[19] where the maximum injection into the FG is at the point where the FG potential is equal to the channel potential[19,20] [Fig. 3(b) and supplementary material, Figs. S1(c) and S1(d)],

$$I_{G,inj} = -I_{DS} P_0 e^{-\frac{V_\alpha}{V_{FG}}}, \tag{5}$$

where I_{DS} is the source–drain current determined by Eq. (4), P_0 is the probability of the hot electrons being emitted to the FG, and V_α is a fitting parameter. Note that Fowler-Nordheim tunneling usually happens in a higher voltage range (>10 V),[21] which is not considered in the current model.

To perform erase, we apply voltage on the SI terminals and leave the SR terminal floating, as shown in Fig. 3(f). The erase is performed by the injection of holes to the FG, which were generated by BBT in the source/channel junctions and accelerated in the lateral field [Fig. 3(g) and supplementary material, Figs. S3(e) and S3(f)],[22,23]

$$I_{G,tunnel} = \xi(V_{FG} - V_{bi})^2 e^{-\frac{\beta}{V_{FG}-V_{bi}}}, \tag{6}$$

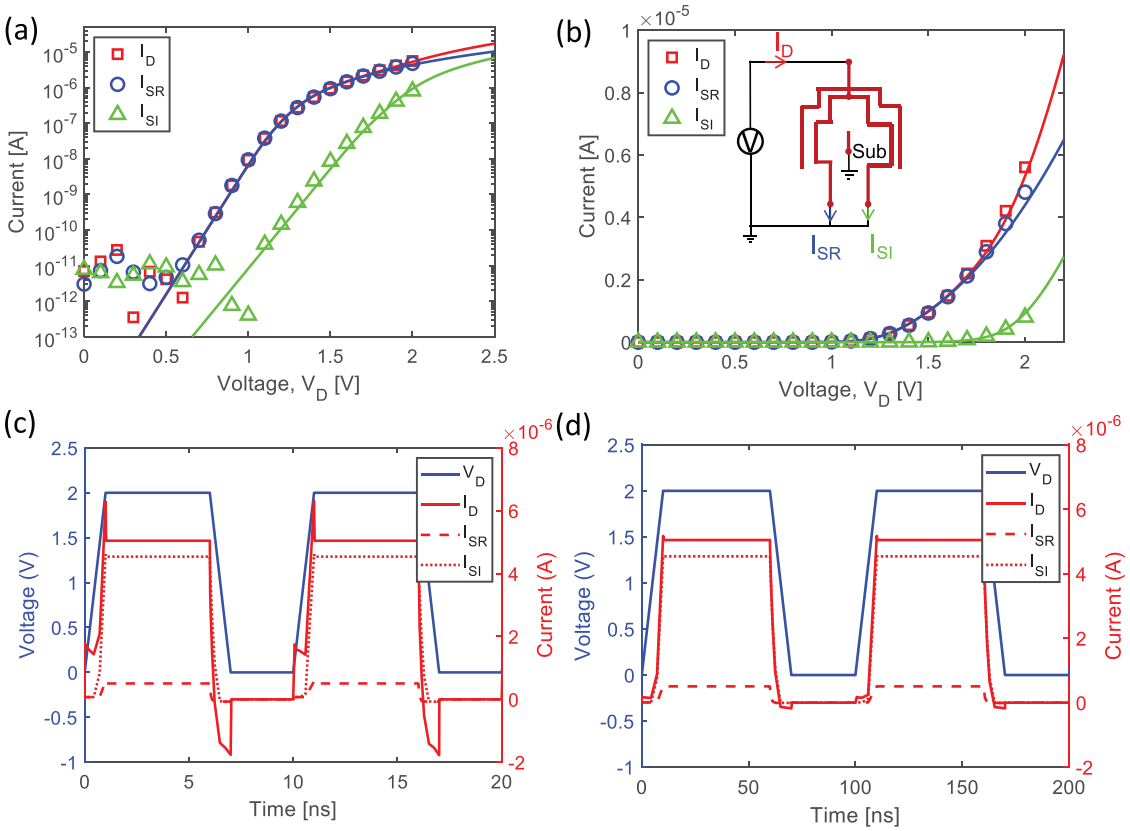

FIG. 2. Comparison of the measured and modeled DC I-V characteristics of the Y-Flash device in read mode for a pristine (non-programmed) device ($Q_{FG} = 0$): (a) in logarithmic scale; (b) in linear scale (inset: schematic of the read operation). Pulse reading of the model implemented in circuit simulator: (c) pulse width 5 ns, rise time 1 ns; (d) pulse width 50 ns, rise time 10 ns.

where β and ξ are fitting parameters reflecting hole injection efficiency into FG and V_{bi} is the potential at the interface with silicon at the point where injection of holes takes place.

The change in the charge on the FG can be modeled by combining the two contributes of gate current,

$$\frac{dQ_{FG}}{dt} = I_{G,inj} + I_{G,tunnel}, \tag{7}$$

where t is the real-time. We assume that the as-fabricated device initially has no charge on the floating gate.

To enable precise control of the states of the Y-Flash device, the voltage pulses with precise time width are employed to program and erase the devices [Figs. 3(a) and 3(f)]. The program pulse in Fig. 3(a) has an amplitude of 5 V, a pulse width of 4 ms, and the rise and fall times of 10 μs. After the application of each program pulse, a sweep voltage from 0 to 2 V with the configuration shown in Fig. 2(b) inset was applied to measure the state of the device. Figure 3(c) shows the readout I–V curves before and after each of the consecutive program pulses. (Only currents on the SR are plotted for simplicity.) The circles are experimentally measured data, and the lines are simulation results. The values of the parameters of the model are given in Table II. The readout currents at 2 V are collected to dedicate the states of the devices as a function of accumulated program time in Fig. 3(d), comparing the experimental data and simulation results. Figure 3(e) shows the charge on the FG extracted from the simulation, as a function of accumulated program time. Note that it takes nine pulses to program the device from the low resistance state (LRS) of readout current being 5 μA to high resistance (HRS) states of the readout current approximating 1 nA; thus, ten discrete conductance states are measured. However, the program pulse width can be lowered to achieve more conductance states. More than 1000 analog conductance states could be obtained when using program pulses of 10 μs [supplementary material, Fig. S4(a)]. The longer pulses used here are mainly to accelerate the measurement of a full program cycle from LRS to HRS.

Similarly, the erase operation is performed by alternatively applying an erase pulse, as illustrated in Fig. 3(c), and performing the readout operation. Voltage pulses with an amplitude of 8 V and a width of 200 μs are used for the erase operation. Figure 3(h) shows the readout I–V curves during the erase operation. The read currents at 2 V are presented in Fig. 3(i) to illustrate the states of the devices as a function of the accumulated erase time. Figure 3(j) shows the charge on the FG as a function of the erase time. Similar to the program operation, more analog conductance states can be achieved by lowering the pulse width of the erase operation [supplementary material, Fig. S4(b)].

The reading current evolution curves in Figs. 3(d) and 3(i) mimic the long-term depression (LTD) and long-term potentiation (LTP) of a

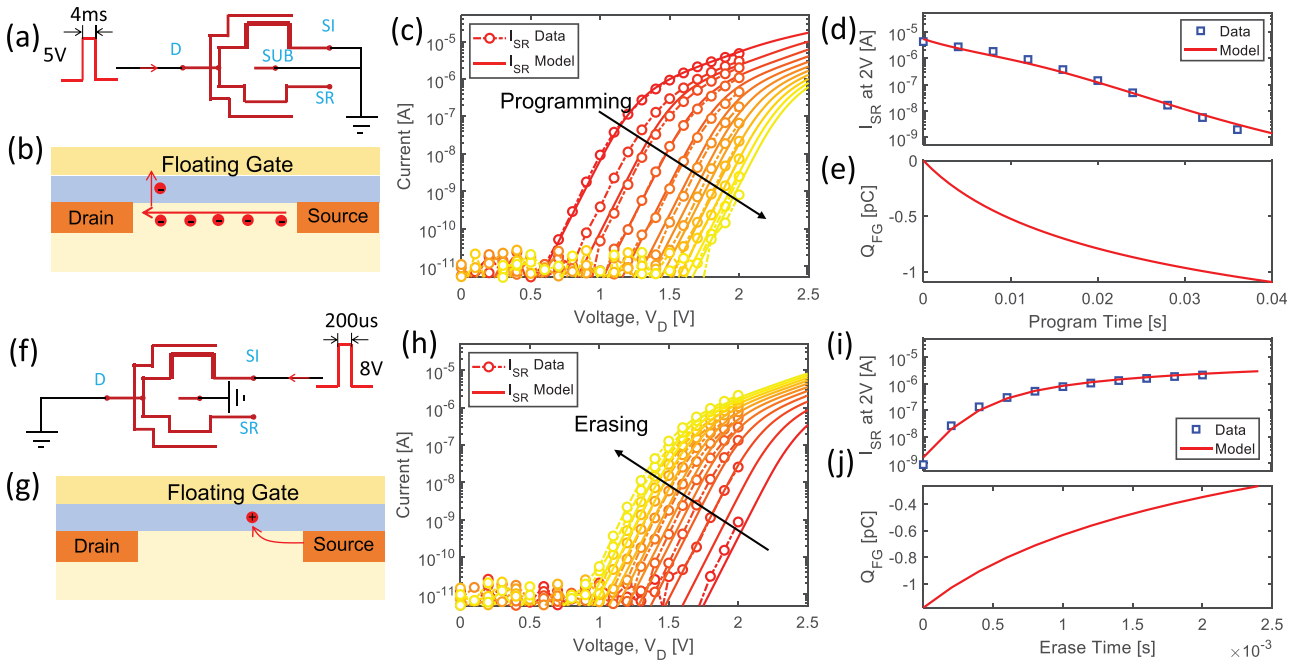

FIG. 3. (a) Schematic of the program operation; (b) mechanism of hot carrier injection into the FG in program operation; (c) readout I-V characteristics for subsequent programming pulses; (d) readout current at 2 V as a function of accumulated program time; (e) FG charge as a function of the accumulated program time; (f) schematic of the erase operation; (g) mechanisms of FG discharge in the erase operation; (h) readout of the device for subsequent erase pulses; (i) readout current at 2 V as the function of accumulated erase time; (j) FG charge as a function of the accumulated erase time.

biological synapse, respectively. By properly engineering the pre-synaptic pulse in the D and the post-synaptic pulse in the SI, the spike-timing-dependent plasticity (STDP) can also be emulated by the device.[16]

As indicated in Table I, the Y-Flash device can be programmed and erased in multiple configurations (for instance, SR is floating or shortened to SI), all of which can be simulated by the current model (see the supplementary material, Figs. S5 and S6 for more program/erase modes). It can be seen clearly that the three-terminal configuration is more efficient than the two-terminal one. This is largely due to the lowering of FG to channel voltage by the coupling of the FG to the SR when SR is shortened to SI. Varing the amplitude of the program or erase pulses, the program and erase operations will result in different program and erase curves as a function of program time and erase time, respectively, which can also be accurately captured by the model (see the supplementary material, Fig. S7). The complete model has been implemented in Verilog-A (supplementary material, Table S3) and enables the circuit design and simulation in the Cadence Virtuoso tool (supplementary material, Fig. S8, all model equations are summarized in the supplementary material, Table S2).

TABLE II. Parameters and their values in the model.

Parameters	Read transistor	Injection transistor
C_{GD}	1.0 fF	
C_{GB}	0.24 fF	
C_{DB}	0.64 fF	
C_{GSR}, C_{GSI}	49 aF	48 aF
C_{SRB}, C_{SIB}	32 aF	32 aF
V_{TH}	0.82 V	1.34 V
I_{S0}	40 nA	80 nA
K	1.9×10^{-5}A/V^2	3.8×10^{-5}A/V^2
n	1.7	2.21
P_0	⋯	3.8×10^{-5}
V_α	⋯	20 V
σ_{V_α}	⋯	0.8 V
β	⋯	10 V
σ_β	⋯	0.8 V
V_{bi}	⋯	5.5 V
ξ	⋯	3.9×10^{-12}A/V^2

To illustrate the cycling performance of the Y-Flash memristor, the program/erase cycling test was conducted. We program the device ($V_P = 5$ V) from the high conductive states ($I_{SR} > 2\ \mu$A at 2 V) to low conductive states ($I_{SR} < 1$ nA at 2 V) by a set of program pulses and then erase the device ($V_E = 8$ V) backward. After 100 cycles, no performance degradation was detected. Figure 4 shows experimental data for ten cycles, together with the corresponding simulation results by the compact model.

We further investigated the device-to-device variations of the Y-Flash memristor. We measured 96 devices in an array (12×8) of the devices. The conductance traces during the program and erase operations as a function of program time and erase time are shown in Figs. 5(a) and 5(b), respectively. The colored lines are typical traces, and the gray lines are all the traces of 96 devices. There do exist

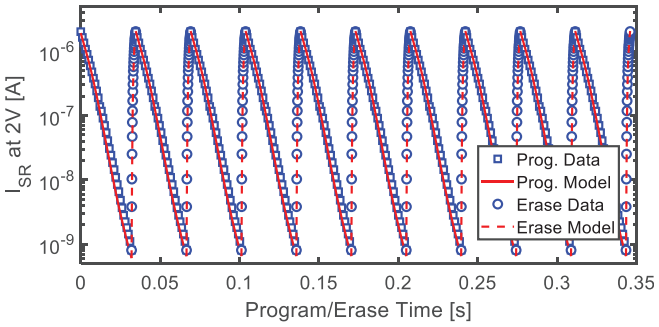

FIG. 4. Cycling of the program and erase operations showing excellent cycle-to-cycle uniformity.

device-to-device variations; however, all the devices work normally, illustrating the high yield of the fabricated devices. The statistical results of the total program time and total erase time are shown in Figs. 5(c) and 5(d), respectively, both showing lognormal distributions. The device-to-device variation can be modeled by adding a Gaussian-type variation to the parameters in the exponents of the gate currents for the program and erase operations in Eqs. (6) and (7), that is,

$$V_{\alpha,d2d} \in N(V_\alpha, \sigma^2_{V_\alpha}) \tag{8}$$

and

$$\beta_{d2d} \in N(\beta, \sigma^2_\beta), \tag{9}$$

where σ_{V_α} and σ_β are the standard deviations of the parameter V_α and β in Eqs. (6) and (7), respectively. Supplementary material Fig. S9 shows the simulated conductance traces of the program and erase cycling as a function of program and erase times. The statistical results of the program time and erase time by the simulation are also shown in Figs. 5(c) and 5(d), respectively.

The high yield guaranteed by the production CMOS fabrication flow and high uniformity of analog switching demonstrated here promise the large-scale implementation of neuromorphic computations. However, learning from recent memristive neuromorphic research, we predict that two issues are limiting the performance of large scale neuromorphic systems based on Y-Flash memristor devices: (i) the non-Ohmic readout behavior (note that the read currents in Fig. 2(a) are exponentially increasing with the applied voltage; thus, the conductance is not constant); (ii) non-linear weight updates for the identical program and erase pulses. [The read currents at 2 V or conductance at 2 V are exponentially dependent on the program/erase time in Figs. 3(d) and 3(i).]

The non-Ohmic behavior prevents the direct implementation of VMM on the Y-Flash memristor array shown in Fig. 1(d) for analog input voltage vector. This issue can be solved by representing the analog input in the pulse width or number of pulses with a fixed read voltage,[24] or by conducting the VMM in a small signal domain.[16] In another aspect, in binary neural networks, with the input voltage vector being binarized,[25,26] the non-Ohmic reading issue would not exist.

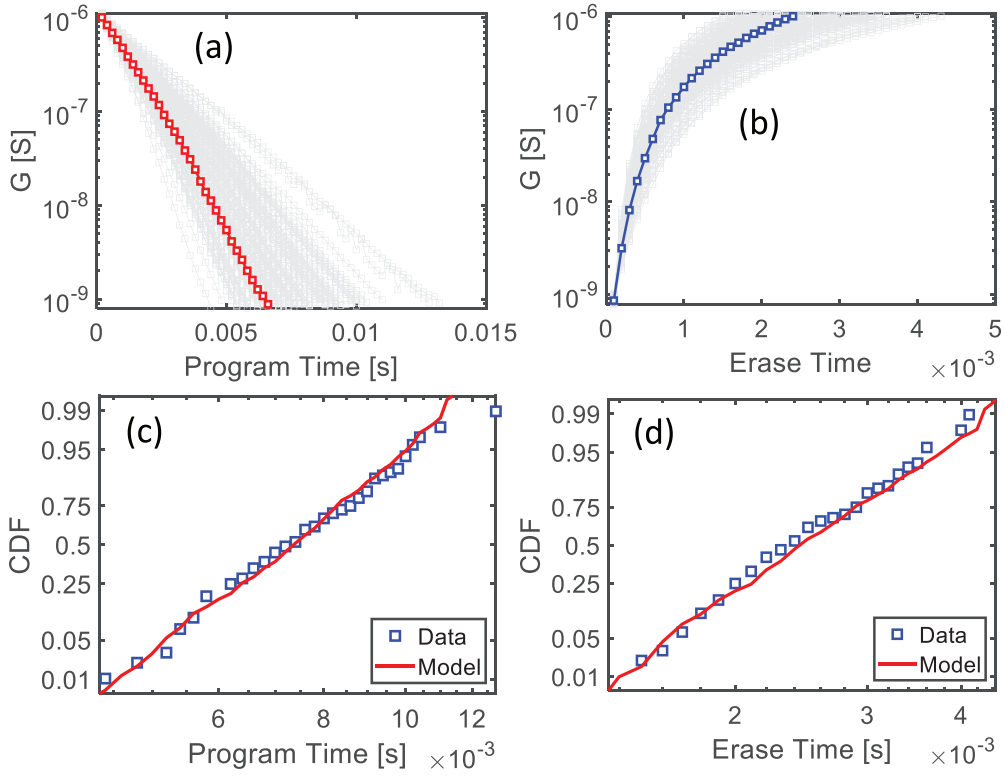

FIG. 5. Device-to-device variations and modeling. Device-to-device variations for (a) the pulsed program operations ($V_P = 5$ V, pulse width 200 μs), and (b) the pulsed erase operations ($V_E = 8$ V, pulse width 100 μs); statistical total (c) program time and (d) erase time from the experiments and model results.

The nonlinear weight update issue prevents precise weight tuning and is the major source of accuracy loss for online training of the mainstream deep neural networks.[27] This issue can be solved by a write-and-verify-read (close-loop write)[28] method, where multiple write operations are needed to tune the conductance to the targeted value with verifying readout operations. An alternative method is to utilize hybrid synaptic cells, incorporating a linear volatile part (for instance, a capacitor) to linearly update weights for each epoch and periodically transfer the volatile weights to the memristor part.[29]

Novel solutions from the perspective of neural network structures and learning algorithms are needed to more efficiently use the Y-Flash memristor in a practical neuromorphic system. The success of these solutions needs precise simulations accounting for the exact electrical behavior of the Y-Flash memristors. The model developed here enables such precise simulations.

In summary, we develop an accurate compact model for the Y-Flash memristor. The model can work in various operational configurations since the missing voltage at floating terminals, including the FG, can be determined by analyzing the capacitor net. The I-V characteristics of the Y-Flash memristor can be accurately reproduced in both subthreshold and above-threshold regions. The switching behaviors (program and erase with voltage pulses) are modeled using equations that account for hot electron injection and band-to-band hot hole injection into the FG. The simulated memristors are fabricated in the mass production CMOS process flows and demonstrate excellent reproducibility of parameters, thus making them promising for various neuromorphic applications.

See the supplementary material for more experimental data, modeling results, and the source code of the model.

This work was supported by the European Research Council through the European Union's Horizon 2020 Research and Innovation Programe under Grant No. 757259. W. Wang was supported in part at the Technion by the Aly Kaufman Fellowship.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]J. J. Yang, D. B. Strukov, and D. R. Stewart, Nat. Nanotechnol. **8**, 13 (2013).
[2]W. Zhang, R. Mazzarello, M. Wuttig, and E. Ma, Nat. Rev. Mater. **4**, 150 (2019).
[3]B. Max, M. Hoffmann, H. Mulaosmanovic, S. Slesazeck, and T. Mikolajick, ACS Appl. Electron. Mater. **2**, 4023 (2020).
[4]Y. Li, J. Lu, D. Shang, Q. Liu, S. Wu, Z. Wu, X. Zhang, J. Yang, Z. Wang, H. Lv, and M. Liu, Adv. Mater. **32**, 2003018 (2020).
[5]Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. A. Talin, and A. Salleo, Nat. Mater. **16**, 414 (2017).
[6]J. Tang, F. Yuan, X. Shen, Z. Wang, M. Rao, Y. He, Y. Sun, X. Li, W. Zhang, Y. Li, B. Gao, H. Qian, G. Bi, S. Song, J. J. Yang, and H. Wu, Adv. Mater. **31**, 1902761 (2019).
[7]G. W. Burr, R. M. Shelby, A. Sebastian, S. Kim, S. Kim, S. Sidler, K. Virwani, M. Ishii, P. Narayanan, A. Fumarola, L. L. Sanches, I. Boybat, M. L. Gallo, K. Moon, J. Woo, H. Hwang, and Y. Leblebici, Adv. Phys. X **2**, 89 (2017).
[8]D. Ielmini and H.-S. P. Wong, Nat. Electron. **1**, 333 (2018).
[9]T. Gokmen and Y. Vlasov, Front. Neurosci. **10**, 333 (2016).
[10]P.-Y. Chen, X. Peng, and S. Yu, in *2017 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2017), pp. 6.1.1–6.1.4.
[11]R. Dittmann and J. P. Strachan, APL Mater. **7**, 110903 (2019).
[12]K. Berggren, Q. Xia, K. K. Likharev, D. B. Strukov, H. Jiang, T. Mikolajick, D. Querlioz, M. Salinga, J. R. Erickson, S. Pi, F. Xiong, P. Lin, C. Li, Y. Chen, S. Xiong, B. D. Hoskins, M. W. Daniels, A. Madhavan, J. A. Liddle, J. J. McClelland, Y. Yang, J. Rupp, S. S. Nonnenmann, K.-T. Cheng, N. Gong, M. A. Lastras-Montaño, A. A. Talin, A. Salleo, B. J. Shastri, T. F. de Lima, P. Prucnal, A. N. Tait, Y. Shen, H. Meng, C. Roques-Carmes, Z. Cheng, H. Bhaskaran, D. Jariwala, H. Wang, J. M. Shainline, K. Segall, J. J. Yang, K. Roy, S. Datta, and A. Raychowdhury, Nanotechnology **32**, 012002 (2021).
[13]M. Ziegler, M. Oberländer, D. Schroeder, W. H. Krautschneider, and H. Kohlstedt, Appl. Phys. Lett. **101**, 263504 (2012).
[14]G. Malavena, M. Filippi, A. S. Spinelli, and C. Monzio Compagnoni, IEEE Trans. Electron Devices **66**, 4727 (2019).
[15]Y. Roizin and E. Pikhay, "Memristor using parallel asymmetrical transistors having shared floating gate and diode," U.S. patent 9514818 (December 6, 2016).
[16]L. Danial, E. Pikhay, E. Herbelin, N. Wainstein, V. Gupta, N. Wald, Y. Roizin, R. Daniel, and S. Kvatinsky, Nat. Electron. **2**, 596 (2019).
[17]W. Wang, W. Song, P. Yao, Y. Li, J. Van Nostrand, Q. Qiu, D. Ielmini, and J. J. Yang, iScience **23**, 101809 (2020).
[18]L. Danial, V. Gupta, E. Pikhay, Y. Roizin, and S. Kvatinsky, in *2020 Design, Automation and Test in Europe Conference and Exhibition (DATE)* (IEEE, 2020), pp. 472–477.
[19]C. Diorio, P. Hasler, and B. A. Minch, IEEE Trans. Electron Devices **43**, 1972–1980 (1996).
[20]S. Tam, P.-K. Ko, and C. Hu, IEEE Trans. Electron Devices **31**, 1116 (1984).
[21]P. Pavan, R. Bez, P. Olivo, and E. Zanoni, Proc. IEEE **85**, 1248 (1997).
[22]K. Yoshikawa, S. Mori, E. Sakagami, Y. Ohshima, Y. Kaneko, and N. Arai, in *Technical Digest—International Electron Devices Meeting* (IEEE, 1990), pp. 577–580.
[23]D. Ielmini, A. Ghetti, A. S. Spinelli, and A. Visconti, IEEE Trans. Electron Devices **53**, 668 (2006).
[24]P. Yao, H. Wu, B. Gao, S. B. Eryilmaz, X. Huang, W. Zhang, Q. Zhang, N. Deng, L. Shi, H.-S. P. Wong, and H. Qian, Nat. Commun. **8**, 15199 (2017).
[25]T. Hirtzlin, M. Bocquet, B. Penkovsky, J.-O. Klein, E. Nowak, E. Vianello, J.-M. Portal, and D. Querlioz, Front. Neurosci. **13**, 1383 (2020).
[26]T. G. Toledo, B. Perach, D. Soudry, and S. Kvatinsky, "Mtj-based hardware synapse design for quantized deep neural networks," arXiv:1912.12636 (2019).
[27]G. W. Burr, R. M. Shelby, S. Sidler, C. D. Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, and H. Hwang, IEEE Trans. Electron Devices **62**, 3498 (2015).
[28]P. Yao, H. Wu, B. Gao, J. Tang, Q. Zhang, W. Zhang, J. J. Yang, and H. Qian, Nature **577**, 641 (2020).
[29]S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, D. Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, and G. W. Burr, Nature **558**, 60 (2018).