

PAPER

Optimized pulsed write schemes improve linearity and write speed for low-power organic neuromorphic devices

To cite this article: Scott T Keene *et al* 2018 *J. Phys. D: Appl. Phys.* **51** 224002

View the [article online](#) for updates and enhancements.

Related content

- [Recent progress in analog memory-based accelerators for deep learning](#)
Hsinyu Tsai, Stefano Ambrogio, Pritish Narayanan *et al.*
- [Organic synaptic devices for neuromorphic systems](#)
Jia Sun, Ying Fu and Qing Wan
- [Recent progress on fabrication of memristor and transistor-based neuromorphic devices for high signal processing speed with low power consumption](#)
Hadiyawarman, Faisal Budiman, Detiza Goldianto Octensi Hernowo *et al.*



IOP | ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Optimized pulsed write schemes improve linearity and write speed for low-power organic neuromorphic devices

Scott T Keene¹, Armantas Melianas¹, Elliot J Fuller²,
Yoeri van de Burgt³, A Alec Talin² and Alberto Salleo^{1,4}

¹ Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, United States of America

² Sandia National Laboratories, Livermore, CA 94551, United States of America

³ Microsystems and Institute for Complex Molecular Systems, Eindhoven University of Technology, 5612AJ Eindhoven, Netherlands

E-mail: asalleo@stanford.edu

Received 12 February 2018, revised 23 March 2018

Accepted for publication 16 April 2018


Published 8 May 2018



Abstract

Neuromorphic devices are becoming increasingly appealing as efficient emulators of neural networks used to model real world problems. However, no hardware to date has demonstrated the necessary high accuracy and energy efficiency gain over CMOS in both (1) training via backpropagation and (2) in read via vector matrix multiplication. Such shortcomings are due to device non-idealities, particularly asymmetric conductance tuning in response to uniform voltage pulse inputs. Here, by formulating a general circuit model for capacitive ion-exchange neuromorphic devices, we show that asymmetric nonlinearity in organic electrochemical neuromorphic devices (ENODEs) can be suppressed by an appropriately chosen write scheme. Simulations based upon our model suggest that a nonlinear write-selector could reduce the switching voltage and energy, enabling analog tuning via a continuous set of resistance states (100 states) with extremely low switching energy ($\sim 170 \text{ fJ} \cdot \mu\text{m}^{-2}$). This work clarifies the pathway to neural algorithm accelerators capable of parallelism during both read and write operations.

Keywords: neuromorphic computing, resistive memory, organic electronics, neural network, symmetric cycling, PEDOT:PSS, electrochemical organic neuromorphic device

 Supplementary material for this article is available [online](#)

(Some figures may appear in colour only in the online journal)

Introduction

Neural networks are increasingly utilized for traditionally difficult computational tasks such as image classification [1, 2], speech recognition [3, 4], and language translation [5, 6]. However, neural algorithm performance has quickly become limited by computationally demanding networks, such as in deep learning, that can require as many as 10^8 synaptic weights to represent the connection strength between neurons in a

network. One strategy is to develop dedicated hardware, such as GPU clusters, that are optimized to efficiently execute core neural algorithms [7–9]. However, any digital architecture will be inherently limited by the energy costs to transmit data between CPU and memory. This has led to interest in neuromorphic computing architectures that outperform traditional CPU and GPU hardware when implementing neural network algorithms and could provide improvements in computation speed and energy (factors of 10^2 – 10^5) [10]. One example is resistive memory devices with multi-level resistance states that emulate synaptic weights in a crossbar array, forming a

⁴ Author to whom any correspondence should be addressed.

dot product engine (DPE) [11, 12]. DPEs are capable of performing analog vector matrix multiplication (VMM), one of the most computationally intensive steps in neural algorithms, in a single compute cycle, thus greatly reducing the latency and energy of computation. Furthermore, an ideal DPE can accelerate both read and write by implementing massive parallelism, where simultaneous pulses to all of the rows and columns of the array can either read out or program conductance states in a single operation [13].

Recently, DPEs have been demonstrated to achieve accuracies of 89.9% when recognizing handwritten digits using filament forming memristors [14]. To achieve this classification accuracy, however, requires feedback programming schemes where each device in the array is iteratively read and programmed to get the desired weight, eliminating the write parallelism necessary for highly efficient analog programming [13]. To achieve acceleration for both read and write operations, a ‘blind update’ scheme (not requiring an additional read) is preferred, where symmetric voltage write-pulses of constant magnitude are applied to the array to achieve a linear change in conductance of the memristive element [15]. Therefore, a new memristive device that has a symmetric and linear response to uniform programming pulses should be implemented to avoid complex programming schemes and benefit from parallel programming in a crossbar array.

Recently, an electrochemical neuromorphic organic device (ENODE), a three-terminal memristive device that uses ionic currents to control the oxidation state of a semiconducting polymer channel, has been demonstrated [16–18]. The advantage of the ENODE architecture is the high-density of ionic doping sites in the conductive polymer which enables virtually-continuous analog conductance tuning [19]. An ENODE consists of an electrochemically active gate electrode used to drive ion exchange between an electrolyte and a doped semiconducting polymer channel (figure 1(a)). The gate voltage controls the electronic carrier concentration, thus modulating the channel conductance (G_{ENODE}). The low energetic barrier for ion migration between the electrolyte and the channel results in low minimum programming energies (e.g. 390 pJ·mm⁻²) [19]. Although the three-terminal architecture presents a challenge for device integration (see supplementary figure 1 for more details (stacks.iop.org/JPhysD/51/224002/mmedia)), the low switching energy and high number of resistance states make ENODEs a promising candidate for low-power neuromorphic computing.

Here, we analyze the performance of ENODEs when they are programmed using square voltage pulses as the programming input. We describe a straightforward ionic circuit model to describe the ENODE charge state as a function of input pulses to the gate. The circuit model is used to extract equivalent circuit parameters for aqueous-electrolyte-based ENODEs, and is expected to be applicable to all-solid-state ENODEs and other, e.g. inorganic, non-volatile electrochemical neuromorphic devices [20]. We demonstrate that asymmetric nonlinearities can be reduced by implementing appropriate programming conditions in order to utilize parallel programming schemes, while maintaining low power

consumption during write operations. The presented model focuses on individual devices, not accounting for circuit parasitics that may arise in an integrated crossbar array which could be incorporated into our model in the future. We conclude by using our proposed model to outline the critical materials parameters when designing next generation ENODE materials.

Methods

ENODE preparation

Device fabrication consists of cleaning 2 cm × 2 cm polished indium–tin oxide (ITO) coated glass slides (Xin Yang Technology LTD) with soap, acetone, methanol, and ethanol. Next, PEDOT:PSS (Hereaus, Clevios PH 1000) aqueous solution is prepared by adding 6 v/v % Ethylene Glycol (EG, Sigma Aldrich) to increase the PEDOT:PSS conductivity, 0.1 v/v % Dodecyl Benzene Sulfonic Acid (DBSA, Sigma Aldrich) as a surfactant, and 1 v/v % (3-glycidyloxypropyl) trimethoxysilane (GOPS, Sigma Aldrich) as a crosslinking agent to improve mechanical stability. PEDOT:PSS solution is spun on the ITO slide at 1000 RPM for 2 min and baked at 120 °C for 20 min. To prepare the ENODE channels, PEDOT:PSS films were placed in a sealed glass chamber and vapor-doped by heating polyethyleneimine (PEI, Sigma Aldrich) on a hotplate (Thermo Scientific) at 250 °C for 5 min. A PEI/PEDOT:PSS film and PEDOT:PSS film ($L = 11$ mm, $W = 14$ mm, $A = 154$ mm²) are placed adjacent to each other with a PDMS well (3M) defining the channel and gate of the device, respectively, and are connected with an aqueous electrolyte (100 mM NaCl).

Pulsed measurements

Pulsed measurements were carried out with a Keithley 2612B source-measure unit with custom LabView code. Devices were programmed through a limit resistor (R_{limit}) with a symmetric write-pulse amplitude (V_{pulse}) using square pulses with a programmable pulse duration (t_{pulse}), pulse delay (t_{delay}), and number of positive and negative pulses (n).

MATLAB code and model fitting procedure

All Keithley parameters are fixed to those used in experiment. ENODE parameters (capacitance C_{ENODE} , charge-transfer resistance R_{ct} , and electrolyte resistance R_{el} , see figure 1(b)) were obtained by iteratively fitting the model to the experimental data by least-squares using custom MATLAB code. The fit is global, i.e. the entire pulse cycling sequence (multiple up/down cycles, e.g. see figure 1(c)) are fitted simultaneously to determine C_{ENODE} , R_{ct} , and R_{el} for a single ENODE. The model was verified by varying V_{pulse} , t_{pulse} , and R_{limit} in experiment, and fitting the model to the experimental data, resulting in similar ENODE parameters. For our simulations, we use the average of C_{ENODE} , R_{ct} , and R_{el} device parameters for the various tested conditions.

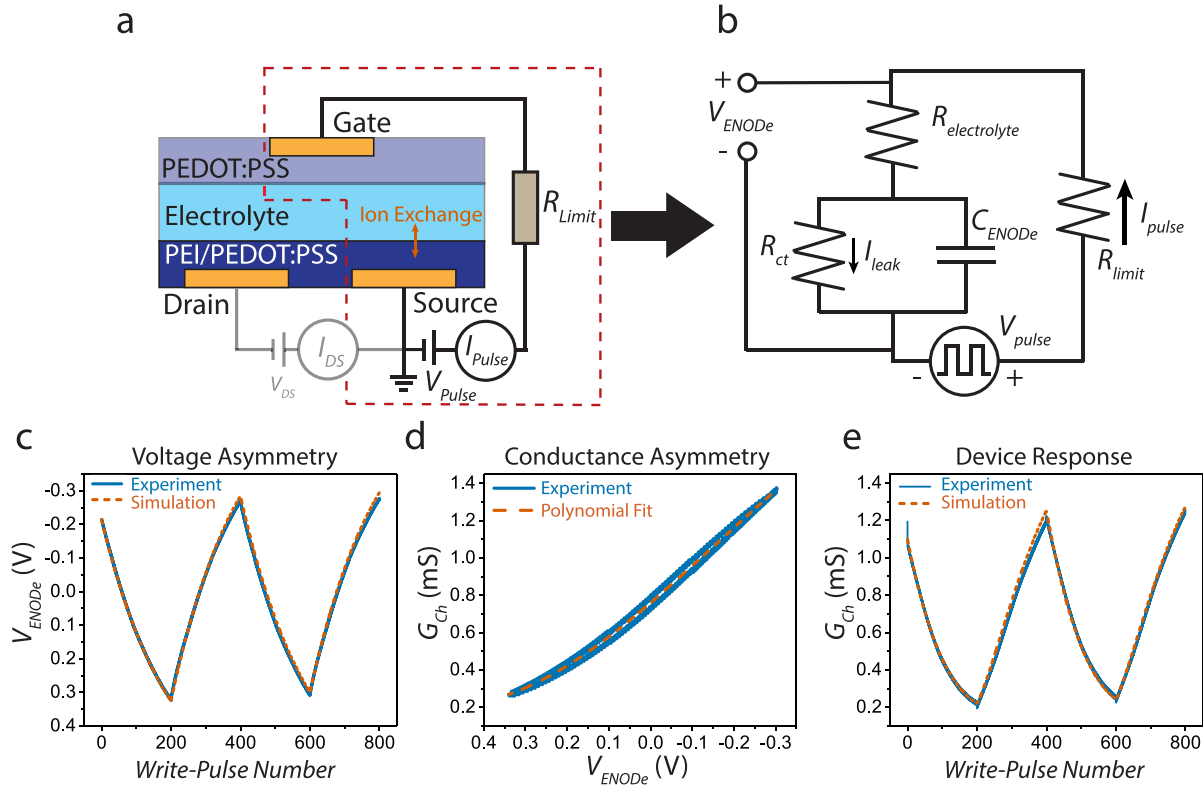


Figure 1. ENODE circuit model: (a) schematic of the ENODE measurement setup and (b) equivalent circuit model. (c) First, the circuit elements C_{ENODE} , R_{el} , and R_{ct} are obtained by fitting the experimentally determined voltage across the ENODE V_{ENODE} under pulsed conditions, then (d) the relationship between V_{ENODE} and conductance G_{ENODE} is fit using a 3rd order polynomial. (e) The simulated V is then input into the polynomial describing G_{ENODE} versus V_{ENODE} to reproduce the conductance response of the ENODE channel to input pulses.

Results and discussion

Equivalent circuit model

Figure 1(a) shows a schematic of an ENODE with a limit resistor ($R_{limit} = 1 \text{ M}\Omega$) to prevent discharge between programming pulses. Figure 1(b) shows the corresponding circuit model used in the simulations, where R_{el} represents the electrolyte resistance, C_{ENODE} represents the mutual capacitance between the gate and channel, and R_{ct} represents the equivalent resistance describing faradaic current from undesired redox reactions at the channel/electrolyte interface. Figure 1(c) shows that our model can successfully reproduce experimental device cycling data spanning hundreds of potentiation/depotentiation (up/down) write pulses. The dataset in figure 1(c) corresponds to a measurement using a series of 100 square pulses with pulse amplitude $V_{pulse} = 2 \text{ V}$, pulse duration $t_{pulse} = 1 \text{ s}$, and delay between pulses $t_{delay} = 1 \text{ s}$. We verified our model by varying the experimental conditions V_{pulse} , t_{pulse} , and/or R_{limit} , and in all cases we could successfully fit the experimental data.

In the circuit simulation, the voltage across the ENODE (V_{ENODE}), i.e. the voltage between the gate and source, is computed by integrating the current flowing into/out of C_{ENODE} during continued pulsing. To fit simulations to measured data, we measure the gate current (I_{GS}) during off pulses ($V_{app} = 0 \text{ V}$) and compute V_{ENODE} using Ohm's Law:

$$V_{ENODE} = I_{GS} * R_{limit}. \quad (1)$$

The result is plotted in figure 1(c). The simulation is iteratively fit to experiment by a least-squares method to accurately obtain the fit parameters (C_{ENODE} , R_{el} , and R_{ct}) describing the ENODE equivalent circuit. We perform this procedure for varied experimental parameters (V_{pulses} , t_{pulses} , R_{limit}) and take their average to obtain $C_{ENODE} = 4.06 \pm 0.24 \text{ }\mu\text{F}\cdot\text{mm}^{-2}$, $R_{el} = 228 \pm 62 \text{ }\Omega$, and $R_{ct} = 274 \pm 87 \text{ M}\Omega\cdot\text{mm}^2$ for a PEDOT/PEI:PSS ENODE. We will later discuss how these channel properties affect ENODE write speeds and switching energy.

The modeling illustrates two sources of asymmetric non-linearity (v) when cycling ENODEs with a uniform write pulse input: (1) voltage asymmetry (figure 1(c)) and (2) channel conductance asymmetry (figure 1(d)). Voltage asymmetry is the result of the voltage division between the limit resistor and the parallel RC component of the ENODE; as C_{ENODE} is charged, the voltage across R_{limit} , and therefore the programming current (I_{pulse}), decreases. This decrease leads to reduced charging of C_{ENODE} with successive programming pulses, giving rise to the tapered shape of the V_{ENODE} versus pulse number as shown in figure 1(c). The conductance asymmetry is an ENODE materials property: as holes in the organic channel are removed, further de-doping of the channel becomes increasingly difficult, possibly related to the shape of the material's density of states [21] or changes in the charge

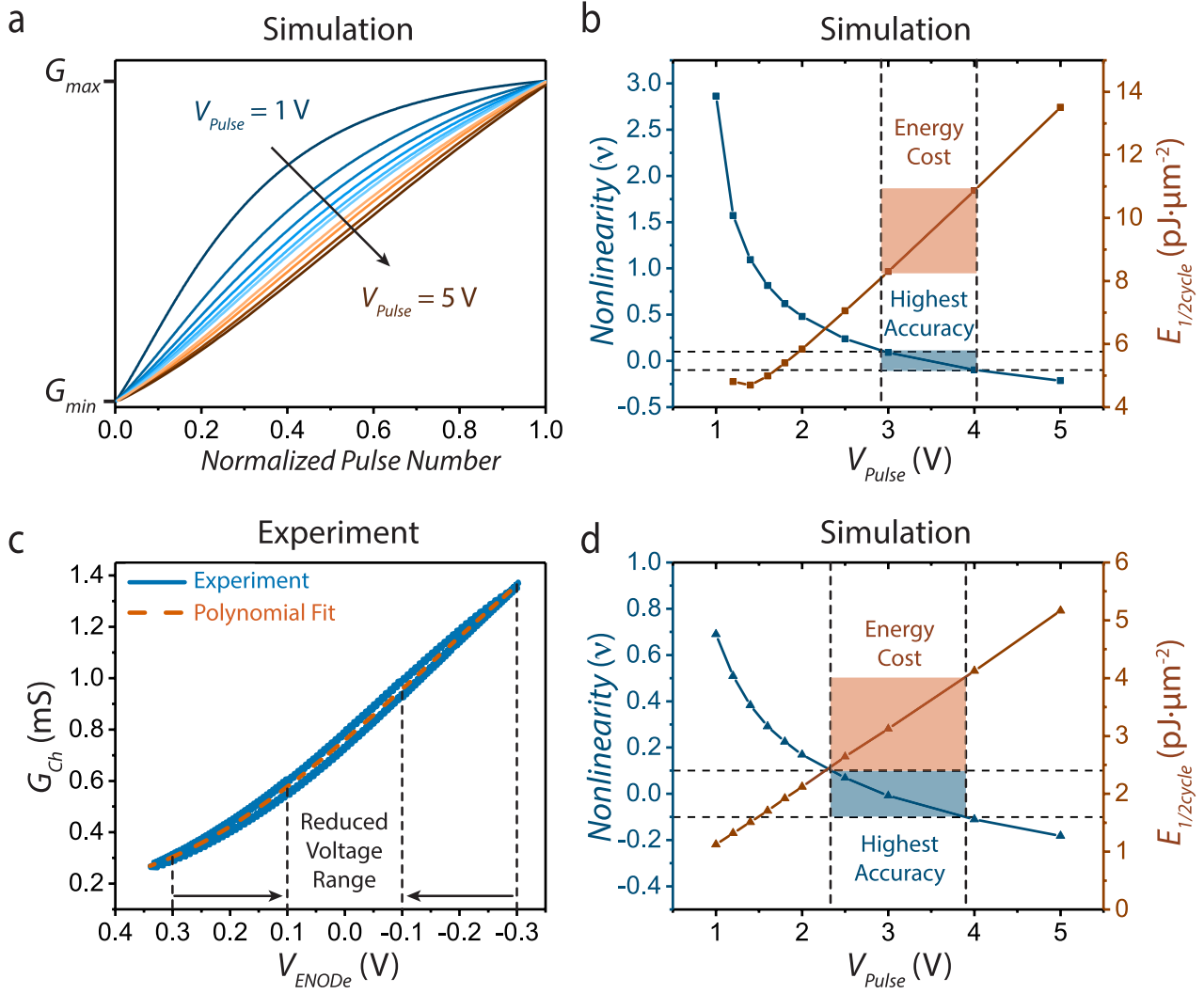


Figure 2. Minimizing ENODE asymmetric nonlinearity (v): (a) conductance versus pulse number for ENODEs programmed with increasing pulse amplitude V_{pulse} . (b) asymmetric nonlinearity v and half-cycle energy $E_{1/2cycle}$ versus V_{pulse} . (c) By limiting the conductance range utilized, (d) $E_{1/2cycle}$ and V_{pulse} can be reduced at the cost of a smaller synaptic weight range. Dashed lines in (b) and (d) indicate the desired linearity for a neural algorithm accelerator ($-0.1 < v < 0.1$) [23].

carrier mobility [22]. The combination of the voltage asymmetry and conductance asymmetry leads to the ‘shark-fin’ shape seen in figure 1(e), which is undesirable when applying a blind-update programming scheme [23]. In the next section, we describe how nonlinearity can be minimized by tuning the write pulse and external circuitry.

Minimizing asymmetric nonlinearity by increasing pulse amplitude

To quantify the linearity of the simulated G_{ENODE} versus write pulse number response plotted in figure 2(a), we fit our data according to the definition of asymmetric nonlinearity from [23, 24] for each V_{pulse} :

$$G = G_1 (1 - e^{-vP}) + G_{min} \quad (2)$$

$$\text{where } G_1 = \frac{G_{max} - G_{min}}{1 - e^{-v}} \quad (3)$$

where G is the device conductance as a function of the normalized number of programming pulses (P), G_{max} and G_{min} are the maximum and minimum conductance state, respectively, and v is the parameter characterizing asymmetric nonlinearity (v closer to 0 is more linear). The absolute number of pulses is normalized to fit the form of equations (2) and (3) describing v . Although increasing V_{pulse} reduces nonlinearity (figure 2(a)), it also increases the energy cost. To quantify both effects and find the optimum write scheme, we compute the energy cost per cycle for each programming scheme using:

$$E_{switch} = \int_0^{t_{pulse}} f * \frac{V_{pulse}^2}{R_{Limit}} * dt \approx f * \frac{V_{pulse}^2}{R_{Limit}} * t_{pulse} \quad (4)$$

$$E_{1/2cycle} = \sum_1^n E_{switch} = f * \frac{V_{pulse}^2}{R_{limit}} * t_{pulse} * n \quad (5)$$

where $E_{1/2cycle}$ is the total switching energy to span the desired conductance range of the ENODE, E_{switch} is the energy to

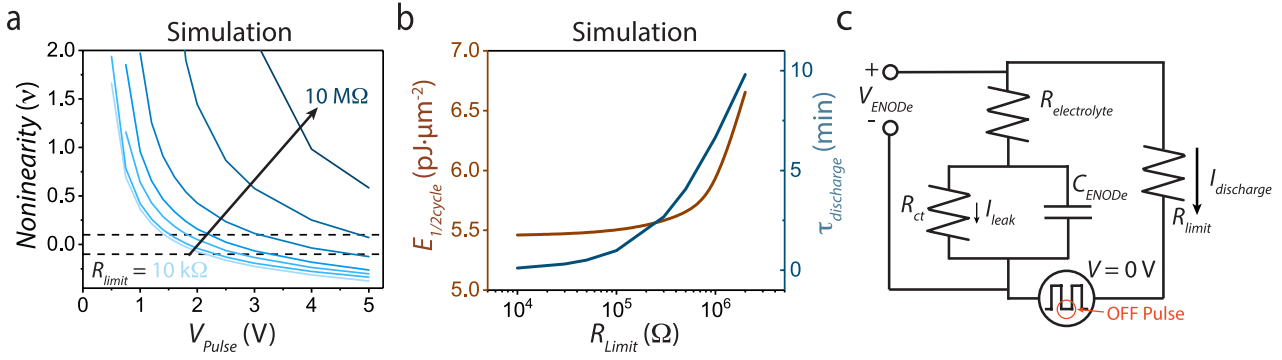


Figure 3. Simulating varied limit resistance R_{limit} : (a) nonlinearity as a function of pulse amplitude V_{pulse} for ENODEs with increasing R_{limit} (from light blue to dark blue) when cycled between $\pm 0.3 \text{ V}$, (b) half-cycle energy consumption $E_{1/2 \text{ cycle}}$ as a function of R_{limit} for a fixed $V_{pulse} = 3 \text{ V}$. (c) Schematic showing discharge pathway for low R_{limit} values between programming pulses.

switch the device to an adjacent state, n is the number of states that span the dynamic range of the ENODE, and f is the fraction of I_{pulse} traveling through C_{ENODE} . Figure 2(b) compares the v and $E_{1/2 \text{ cycle}}$ as a function of V_{pulse} . The dashed lines highlight the region that satisfies the required linearity for high accuracy neural algorithm accelerator ($-0.1 < v < 0.1$) [23]. The ENODEs presented here can operate in the high accuracy range, albeit at a relatively high voltage. More concretely, for an ENODE cycled between $\pm 0.3 \text{ V}$, the minimum V_{pulse} to achieve high accuracy in an array is 3 V with $E_{1/2 \text{ cycle}} = 7.5 \text{ pJ} \cdot \mu\text{m}^{-2}$. It would be desirable to further reduce both the applied voltage and the energy consumption.

One strategy to reduce the applied voltage and therefore energy cost is to reduce the voltage range accessed during programming, as indicated in figures 2(c) and (d). This reduces V_{pulse} and $E_{1/2 \text{ cycle}}$ to ca. 2.4 V and $2.5 \text{ pJ} \cdot \mu\text{m}^{-2}$, respectively. However, this comes at the cost of the dynamic range; when using the full programming range (ca. $\pm 0.3 \text{ V}$), ENODEs have an G_{max}/G_{min} of ca. 4.5, whereas when the operating conductance range is reduced to $\pm 0.1 \text{ V}$, the G_{max}/G_{min} ratio is only 1.7. Therefore, although the ENODEs presented here meet the stringent linearity criteria using $\sim 2.4 \text{ V}$ write pulses, this leads to a relatively low G_{max}/G_{min} ratio, setting a low noise tolerance for an integrated array. Thus, there is a trade-off between high noise tolerance (full range) and low power (reduced range).

Reducing limit resistance for high speed and low power

In order to retain the charge state across the ENODE, and therefore memory state, we insert an R_{limit} to impede discharge of the device through the programming circuit. However, a high R_{limit} leads to inefficient writing by both increasing t_{switch} and E_{switch} , resulting in increased v (figure 3(a)) and increased $E_{1/2 \text{ cycle}}$ (figure 3(b), orange trace). Therefore, R_{limit} values should be increased to minimize self-discharge only by as much as necessary for the desired application. In contrast, low R_{limit} values have several benefits, as outlined below.

It is evident from the simulation that lower R_{limit} reduces the optimal V_{pulse} to achieve the desired linearity (dashed horizontal lines in figure 3(a)). To explain this behavior, first

we consider how the fractional charging current, $(I_{pulse} \cdot f)$, depends on R_{limit} :

$$I_{leak} = \frac{V_{ENODE}}{R_{ct}} \quad \text{and} \quad I_{pulse} = \frac{V_{pulse}}{R_{limit}} \quad (6) \text{ and } (7)$$

$$f = \frac{I_{pulse} - I_{leak}}{I_{pulse}} = 1 - \frac{V_{ENODE} * R_{limit}}{V_{pulse} * R_{ct}}. \quad (8)$$

As R_{limit} is decreased, f increases, resulting in increased linearity in the ENODE response to square voltage pulse inputs. Additionally, the increase in f also leads to decreased energy costs for programming due to decreased total current, given by equations (4) and (5). Lowering R_{limit} also increases the switching speed of the ENODE. When $I_{leak} \ll I_{pulse}$ ($f \sim 1$), the equivalent switching time can be estimated as a charging RC circuit as follows:

$$\begin{aligned} t_{switch} &= \frac{1}{n} * C_{ENODE} \frac{(V_{max} - V_{min})}{I_{pulse}} \\ &= \frac{1}{n} * R_{limit} * C_{ENODE} \frac{(V_{max} - V_{min})}{V_{pulse}} \end{aligned} \quad (9)$$

where V_{max} and V_{min} correspond to the maximum and minimum programmed voltage across C_{ENODE} , respectively.

Although decreasing R_{limit} decreases t_{switch} , v , V_{pulse} , and E_{switch} and thus improves device performance considerably, decreased R_{limit} also results in rapid decay of ENODE memory, eliminating the non-volatile aspect of the device (figure 3(b), blue trace). The characteristic discharge time, $\tau_{discharge}$, is given by the following equation:

$$\tau_{discharge} = R_{limit} * C_{ENODE}. \quad (10)$$

For the case when the discharge current ($I_{discharge}$) through the external circuit dominates. For the resistances used here, $\tau_{discharge}$ ranges from ca. 6 s (for $R_{limit} = 10 \text{ k}\Omega$) to ca. 6000 s ($R_{limit} = 10 \text{ M}\Omega$). This is a crucial drawback, as the ENODE state must be maintained in order to do useful computations after training. In order to be limited by materials performance rather than the programming circuit, the discharge time $\tau_{discharge}$ should be greater than the discharge time due to parasitic reactions such as oxidation. In the next section,

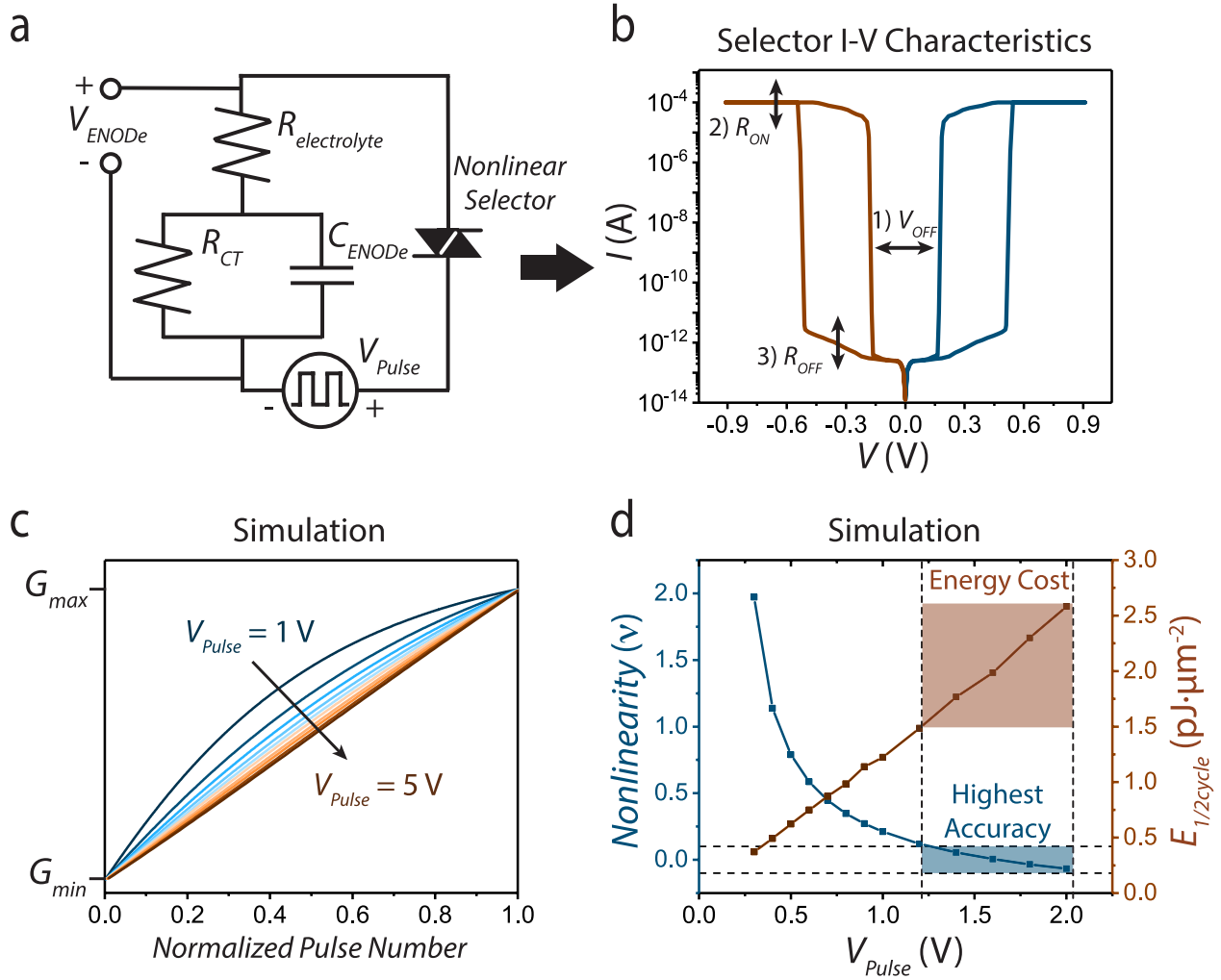


Figure 4. Device simulation with a nonlinear selector: (a) circuit diagram of a nonlinear selector device integrated with an ENODE, (b) IV characteristics of a high performance nonlinear selector from [25] showing (1) turn-off voltage V_{OFF} , (2) ON resistance R_{ON} , and (3) OFF resistance R_{OFF} , (c) conductance G_{ENODE} versus normalized pulse number for ENODEs programmed through a selector device with a reduced voltage range (± 0.15 V), (d) nonlinearity and energy cost of an ENODE programmed through a nonlinear selector with 10^8 ON/OFF ratio. The non-linear selector enables linearity at lower write pulse amplitudes.

we discuss how volatility can be minimized by integrating a nonlinear selector with the ENODE gate.

Nonlinear selector improves linearity and reduces switching energy

In order to minimize the effect of ENODE discharge between pulses, a nonlinear selector device can be inserted between the ENODE source and gate (figures 4(a) and (b)). The nonlinear selector acts as an electronic switch which uses voltage threshold filament formation between two metal electrodes to switch between an OFF state (no filaments) and ON state (filament forming a conductive bridge between electrodes) [25]. There are three important characteristics that determine the performance of a nonlinear selector coupled to an ENODE (figure 4(b)): (1) the turn-off voltage (V_{OFF}), (2) the ON resistance (R_{ON}), and (3) the OFF resistance R_{OFF} . V_{OFF} determines the programming voltage range, and therefore G_{min}/G_{max} that can be used; if the ENODE charges beyond V_{ON} , it will

discharge through the selector until it reaches the V_{OFF} and the selector resistance increases. R_{ON} determines the write V_{pulse} and E_{switch} required to achieve linear G_{ENODE} versus pulse number as illustrated in the previous section. R_{OFF} determines the ENODE discharge rate by inserting it in place of R_{limit} in equation (10). Here, we insert the characteristics from [25] (OFF resistance of $10^{13} \Omega$ and an ON/OFF ratio of 10^8 , slope of $< 1 \text{ mV} \cdot \text{dec}^{-1}$, and turn ON/OFF time of 75/250 ns) into our simulation to demonstrate the effectiveness of the nonlinear selector when coupled with an ENODE (figures 4(c) and (d)).

In the simulation (figures 4(c) and (d)), V_{ENODE} is limited to ± 0.15 V to match the approximate V_{OFF} of the selector [25]. The required V_{pulse} to achieve sufficient v is reduced to 1.25 V (versus 3 V) primarily because of the reduced R_{ON} of the selector ($\sim 10^5 \Omega$) compared to R_{limit} ($10^6 \Omega$) used in the previous sections. The limited voltage range of ± 0.15 V also contributes to the reduced V_{pulse} requirement. Because the total charge (and therefore integrated current) required

to change the ENODE state is determined by C_{ENODE} , E_{switch} will decrease even after inserting the selector device due to the lowered V_{pulse} . Additionally, the selector device does not have the drawback of rapid discharge, or loss of state, associated with low R_{limit} values due to the high R_{OFF} ($10^{13} \Omega$). The simulation shows $E_{1/2\ cycle}$ as low as $17\text{ pJ} \cdot \mu\text{m}^{-2}$ can be achieved with desired linearity $v < 0.1$, which corresponds to $E_{switch} = 170\text{ fJ} \cdot \mu\text{m}^{-2}$ for an ENODE with 100 conductance states.

Inserting a selector in place of R_{limit} eliminates the tradeoff between switching energy and retention time that is problematic for some memristor devices. The nonlinear selector minimizes the effect of circuit discharge between programming pulses, which would be detrimental during both training and operation of the crossbar array. This allows the state-retention of the ENODE to be materials-limited rather than circuit-limited. In addition, the selector provides a low load during programming, allowing the ENODE to maintain previously reported low energy consumption [19]. Furthermore, because of the ENODE's three terminal architecture, the selector is decoupled from the channel (only couples the source-gate circuit) and thus does not distort the read signal, as is the case for two terminal memristors [14]. Thus, the selector is an ideal alternative to R_{limit} ; the resistance is maximized to retain the programmed state but minimized during write to avoid high V_{pulse} and E_{switch} , enabling low-power operation.

Conclusion

In this work, we present a general model to describe the linearity of ENODE response to square programming pulses with arbitrary amplitude V_{pulse} to determine optimal write schemes to utilize parallel read-write operations. In our analysis, we find a tradeoff between the maximum/minimum conductance ratio G_{max}/G_{min} and energy consumption due to the linearity requirement. Additionally, we demonstrate the need for a nonlinear selector device, which minimizes resistance during programming to reduce switching time t_{pulse} , voltage V_{pulse} , and energy E_{switch} , and maximizes resistance during read to minimize leakage of the ENODE charge. Finally, we project linear scaling of E_{switch} and t_{switch} by decreasing the ENODE volume, and therefore device capacitance C_{ENODE} . Using PEDOT/PEI:PSS as a model channel material with a nonlinear selector device, we project $E_{switch} = 170\text{ fJ}$ and $t_{switch} = 500\text{ ns}$ using $V_{pulse} = 1.3\text{ V}$ for a $1\text{ }\mu\text{m} \times 1\text{ }\mu\text{m} \times 200\text{ nm}$ (length \times width \times thickness) channel with 100 conductance states. The predicted t_{switch} is estimated based on the switching speed of the selector and the drift velocity of protons in PEDOT:PSS rather than the RC charging time of the circuit, which would give a $t_{switch} < 1\text{ ns}$ (see supplementary text 1). We chose to project to a feature size of $1\text{ }\mu\text{m}$ because it is the resolution limit for our current photolithographic process [26]. Our future works aim to investigate the switching speed limitations for further scaled devices.

We also use the insights of the model to identify the critical materials properties for ENODE channels. We find that the write speed is improved by reducing C_{ENODE} . This can be achieved by scaling down the channel dimensions, e.g. by standard

lithography, and by selecting low capacitance organic channels. To increase state retention, parasitic reactions leading to loss of state must be minimized. State retention may be improved by eliminating electron transfer pathways between the electrolyte and the channel, such as by encapsulating in an inert environment to avoid oxidation. Finally, we find that the critical parameter when selecting channel materials is the G_{Ch} versus V_g response. An ideal material will have a broad linear region with a steep slope in order to maximize the G_{max}/G_{min} ratio within a minimal V window. Our future work aims to expand the scope of the model to include solid-state electrolytes and to identify and leverage the structure-property relationships of organic semiconductors to design next generation ENODE materials.

Acknowledgments

STK was supported by an Office of Technology Licensing Stanford Graduate Fellowship. AM gratefully acknowledges support from the Knut and Alice Wallenberg Foundation (KAW 2016.0494) for Postdoctoral Research at Stanford University. EJJ and AAT were supported by Nanostructures for Electrical Energy Storage (NEES), an Energy Frontier Research Center (EFRC) funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under award number DESC0001160 and by Sandia's Laboratory-Directed Research and Development program. 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 US Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the US Department of Energy or the United States Government. AS acknowledges financial support from the National Science Foundation, E2CDA Award #1739795.

ORCID iDs

Scott T Keene  <https://orcid.org/0000-0002-6635-670X>
 Armantas Melianas  <https://orcid.org/0000-0002-3443-0987>
 Yoeri van de Burgt  <https://orcid.org/0000-0003-3472-0148>

References

- [1] Esteva A et al 2017 Dermatologist-level classification of skin cancer with deep neural networks *Nature* **542** 115–8
- [2] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks *Proc. 25th Int. Conf. Neural Information Processing Systems* vol 1 pp 1097–105
- [3] Graves A, Mohamed A-R and Hinton G 2013 Speech recognition with deep recurrent neural networks 2013 *IEEE Int. Conf. Acoustics Speech and Signal Processing* (<https://doi.org/10.1109/ICASSP.2013.6638947>)
- [4] Xiong W et al 2017 The microsoft 2016 conversational speech recognition system 2017 *IEEE Int. Conf. Acoustics Speech and Signal Processing* (<https://doi.org/10.1109/ICASSP.2017.7953159>)

- [5] Johnson M *et al* 2017 Google's multilingual neural machine translation system: enabling zero-shot translation (arXiv:1611.04558v2)
- [6] Cho K *et al* 2014 Learning phrase representations using RNN encoder-decoder for statistical machine translation (arXiv:1406.1078v3)
- [7] Chetlur S *et al* 2017 cuDNN: efficient primitives for deep learning (arXiv:1410.0759v3)
- [8] Cong G, Kingsbury B, Gosh S, Saon G and Zhou F 2017 Accelerating deep neural network learning for speech recognition on a cluster of GPUs *Proc. Machine Learning HPC Environment* (<https://doi.org/10.1145/3146347.3146351>)
- [9] Jouppi N P *et al* 2017 In-datacenter performance of a tensor processing unit *44th Annual Int. Symp. on Computer Architecture* pp 1–12
- [10] Taha T M, Hasan R, Yakopcic C and McLean M R 2013 Exploring the design space of specialized multicore neural processors *Proc. Int. Joint Conf. Neural Networks* pp 2633–40
- [11] Hu M *et al* 2016 Dot-product engine for neuromorphic computing: programming 1T1M crossbar to accelerate matrix-vector multiplication *53rd ACM/EDAC/IEEE, Design Automation Conf. (DAC)* (<https://doi.org/10.1145/2897937.2898010>)
- [12] Burr G W *et al* 2017 Neuromorphic computing using non-volatile memory *Adv. Phys. X* **2** 89–124
- [13] Agarwal S *et al* 2016 Energy scaling advantages of resistive memory crossbar based computation and its application to sparse coding *Frontiers Neurosci.* **9** 484
- [14] Hu M *et al* 2018 Memristor-based analog computation and neural network classification with a dot product engine *Adv. Mater.* **30** 1705914
- [15] Jacobs-Gedrim R B *et al* 2017 Impact of linearity and write noise of analog resistive memory devices in a neural algorithm accelerator *2017 IEEE Int. Conf. on Rebooting Computing (ICRC)* (IEEE) pp 1–10
- [16] Gkoupidenis P, Schaefer N, Garlan B and Malliaras G G 2015 Neuromorphic functions in PEDOT:PSS organic electrochemical transistors *Adv. Mater.* **27** 7176–80
- [17] Gkoupidenis P, Schaefer N, Strakosas X, Fairfield J A and Malliaras G G 2015 Synaptic plasticity functions in an organic electrochemical transistor *Appl. Phys. Lett.* **107** 263302
- [18] Qian C *et al* 2016 Artificial synapses based on in-plane gate organic electrochemical transistors *ACS Appl. Mater. Interfaces* **8** 26169–75
- [19] van de Burgt Y *et al* 2017 A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing *Nat. Mater.* **16** 414–8
- [20] Fuller E J *et al* 2017 Li-ion synaptic transistor for low power analog computing *Adv. Mater.* **29** 1–8
- [21] Friedlein J T *et al* 2017 Influence of disorder on transfer characteristics of organic electrochemical transistors *Appl. Phys. Lett.* **111** 023301
- [22] Tybrandt K, Zozoulenko I V and Berggren M 2017 Chemical potential–electric double layer coupling in conjugated polymer-polyelectrolyte blends *Sci. Adv.* **3** eaao3659
- [23] Agarwal S *et al* 2016 Resistive memory device requirements for a neural algorithm accelerator *Proc. Int. Joint Conf. Neural Networks (October 2016)* pp 929–38
- [24] Chen P-Y *et al* 2015 Mitigating effects of non-ideal synaptic device characteristics for on-chip learning *Proc. IEEE/ACM Int. Conf. Computer Design* pp 194–9
- [25] Midya R *et al* 2017 Anatomy of Ag/hafnia-based selectors with 10^{10} nonlinearity *Adv. Mater.* **29** 1604457
- [26] DeFranco J A, Schmidt B S, Lipson M and Malliaras G G 2006 Photolithographic patterning of organic electronic materials *Org. Electron.* **7** 22–8