Phosphosilicate Glass-Based Nanosecond Protonic Programmable Resistors for Analog Deep Learning

Murat Onen, Ju Li, Bilge Yildiz, Jesús A. del Alamo

Massachusetts Institute of Technology, Cambridge, MA 02139, USA

2023 7th IEEE Electron Devices Technology & Manufacturing Conference (EDTM) | 979-8-3503-3252-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/EDTM55494.2023.10102987

Abstract

Programmable resistors are widely explored devices as building blocks for analog deep learning accelerators. In this work, we demonstrate Si-compatible nanoscale protonic devices based on phosphosilicate glass (PSG) electrolyte. The conductance modulation characteristics of our devices are ideal in terms of fast operation (5 ns/pulse), high energy efficiency (~fJ/pulse), large dynamic conductance range, linearity, symmetry, reversibility, retention, and high endurance. This paper summarizes the fabrication and characteristics of these devices with particular attention to the proton dynamics during fast operation.

(Keywords: Analog deep learning, protonics, non-volatile Devices)

Introduction

Acceleratingly increasing deep learning workloads are fast becoming unmanageable for digital processors [1]. Analog hardware has attracted great interest as an alternative paradigm to provide high throughput matrix algebra using a fraction of the energy of digital hardware [2]. The building blocks of these analog architectures are programmable resistors with intrinsic physical properties that can be used to process information. For these operations to be performed accurately, devices need to satisfy a long list of strict requirements [3]. Unfortunately, none of the existing devices repurposed from memory applications matches the needs [4]. Considering that those devices were originally optimized for information *storage* purposes, they naturally have fundamental shortcomings in this new information *processing* application.

A new device family based on electrochemically controlled ion intercalation to an active channel has gained interest thanks to their promising characteristics [5-11]. The fundamental idea behind these devices is dynamic doping of a semiconductor through the application of electrical pulses (Fig. 1). Early variants of this device concept were based on mature-but-CMOS-incompatible Li^+ ion [5] while later attempts have used compatible, yet difficult to move O^{2-} based alternatives (i.e., large and heavy ion) [6]. Ultimately, H^+ (proton)-based variants have come into focus for their small size and light mass, promising fast and energy-efficient operation [7]. Recently, we have resolved the critical limitation of protonic devices: the absence of a CMOS-compatible, solid-state material system [8-10]. In this paper, we summarize the ultrafast modulation characteristics of these devices (Fig. 2) which display outstanding energy efficiency under extreme electric fields.

Device Structure and Operation

The basic operation principle of the devices is based on shuttling protons from a Pd reservoir (gate) into a WO_3 active channel through a PSG electrolyte. Each positive voltage pulse applied to the gate transfers a controlled number of protons into the channel, increasing its conductivity by an incremental amount. The application of a negative pulse reverses this process, reducing the conductivity back. Thanks to the small size of protons, this process does not mechanically stress or strain the material, ensuring high endurance. Moreover, since the modulation of these devices is through macroscopic chemical and electronic modifications, their behavior is distinctly controllable and deterministic. Finally, in the absence of programming pulses, protons remain stationary, and the device conductance remains unchanged (i.e., nonvolatile).

Results

Figure 3A shows the conductance modulation of a device with a 50 nm × 150 nm active area with 10 nm thick PSG for 1000 positive voltage pulses followed by 1000 negative ones [9]. The devices display ideal characteristics in terms of: (1) high modulation speed (5 ns/pulse), (2) nearly linear and symmetric behavior for incremental and decremental changes, (3) good retention (~10^{10}× times longer than programming pulse, Fig. 3B), (4) large dynamic range (20×), (5) optimal base resistance of 88 MΩ, and (6) high endurance without any degradation (Fig. 3C) over 10^5 pulses. Moreover, devices consumed only ~2.5 fJ/pulse which was almost exclusively capacitive loss

(i.e., CV^2), whereas the energy associated with shuttling protons was merely ~15 aJ/pulse. As a result, these protonic devices display state-of-the-art combined material, processing, and performance properties for analog deep learning applications.

In order to further establish the high-speed modulation of these devices, we have recorded the channel current with a 170 MHz bandwidth during the application of 10 copies of 5 ns pulses (Fig. 4). The results show that the conductance modulation occurs in an impulse-like fashion, without any extended equilibration period [10].

Nonetheless, capturing the dynamics during programming was not possible for 5 ns pulses, which is why we opted to study device dynamics with µs-long ones (Fig. 5, right). There are 4 distinct components expected for the channel current: (a) gate leakage, (b) electronic displacement, (c) volatile field-effect, and (d) non-volatile intercalation effect. Of these, gate leakage was found to be negligible (data not shown).

To isolate the different components, we first applied a 3 V pulse before device protonation (red trace in Fig. 5, left). Under this condition, only electronic displacement current was observed during the rising and falling edges of the programming pulse. These results were expected since there are no mobile protons within the device.

Repeating the same programming pulse after the protonation of the stack (i.e., forming gas, FG, treatment), still does not show an intercalation signal (no residual increase in the channel current after the pulse comes to an end). However, the signal (blue trace) exhibits a distinct new component on top of the previously observed displacement currents. This component evolves (and disappears) with the pulse voltage, leaving no residual change behind (i.e., volatile). The nature of this additional current is a proton-induced field-effect in the n-type WO_3 channel. This is due to an increase in the electron concentration in reaction to the pile-up of positively-charged protons within the PSG layer adjacent to the WO_3 interface. In other words, the protons within the electrolyte lead to a negative threshold voltage shift.

Ultimately, non-volatile proton intercalation is only observed when the amplitude of the programming pulse is increased (4.5 V). The evolution of this effect manifests itself as a distinct current rise during the pulse plateau, which ultimately results in a permanent current enhancement after the pulse ends.

Conclusion

In summary, we have demonstrated CMOS-compatible nanosecond programmable protonic resistors with outstanding energy efficiency and near ideal modulation characteristics. Future work will focus on optimizing the PSG/WO_3 interface to reduce the programming voltage.

Acknowledgements

This work was supported by MIT-IBM Watson AI Lab and MIT Quest for Intelligence.

References

[1] E. Strubell, et al., “Energy and Policy Considerations for Deep Learning in NLP”, ACL, (2020).

[2] M. A. Zidan, et al., “The future of electronics based on memristive systems”, Nat. Electron. 1, 22–29 (2018).

[3] T. Gokmen and Y. Vlasov, “Acceleration of deep neural network training with resistive cross-point devices: Design considerations”, Front. Neurosci. 10, 333 (2016).

[4] M. Onen, et al., “Neural Network Training with Asymmetric Crosspoint Elements”, Frontiers in Artificial Intelligence, (2022).

[5] E. J. Fuller, et al., “Li-ion synaptic transistor for low power analog computing”, Adv. Mater. 29, 1604310 (2017).

[6] S. Kim, et al., “Metal-oxide based CMOS-compatible ECRAM for Deep Learning Accelerator”, IEEE Int. Electron Devices Meet., 847–850 (2019).

[7] X. Yao, et al., “Protonic solid-state electrochemical synapse for physical neural networks”, Nat. Commun. 11, 1–10 (2020).

[8] M. Onen, et al., “CMOS-compatible protonic programmable resistor based on phosphosilicate glass electrolyte for analog deep learning”, Nano Lett., 21, 6111–6116 (2021).

[9] M. Onen, et al., “Nanosecond Protonic Programmable Resistors for Analog Deep Learning”, Science, 6605, 539-543, (2022).

[10] M. Onen, et al., “Dynamics of PSG-Based Nanosecond Protonic Programmable Resistors for Analog Deep Learning”, IEDM, p. 38 (2022).

[11] M. Huang, et al., “Electrochemical Ionic Synapses: Progress and Perspectives”, Adv. Mater., (2022).

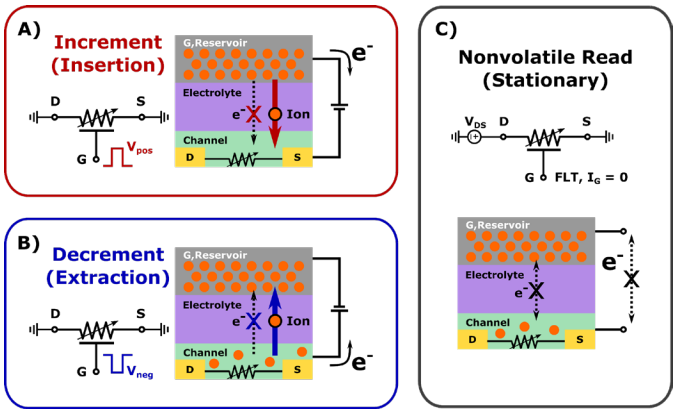

Fig. 1. Operation of a programmable resistor based on ion intercalation under **(A)** positive pulse, **(B)** negative pulse, and **(C)** no pulse conditions.

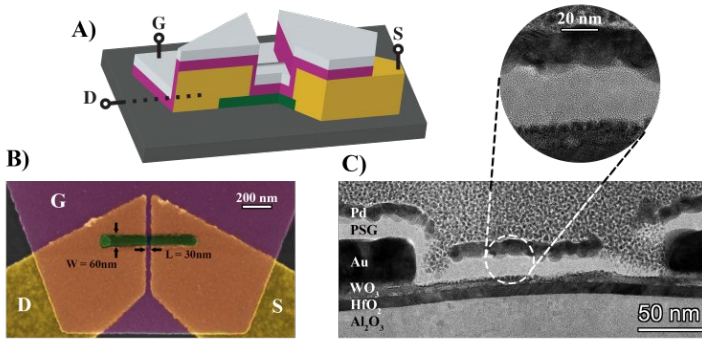

Fig. 2. **(A)** 3-D illustration, **(B)** False-colored top-view SEM image, and **(C)** TEM cross-section image of a protonic programmable resistor. Figure adapted from [9].

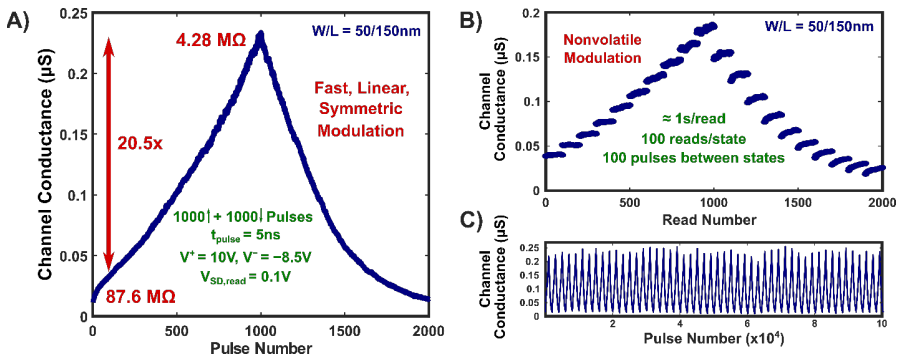

Fig. 3 (A) Modulation performance of a 50 nm × 150 nm active area protonic device showing fast (5 ns/pulse), nearly linear, and symmetric characteristics. **(B)** Retention, and **(C)** endurance behavior of the protonic device. Figure adapted from [9].

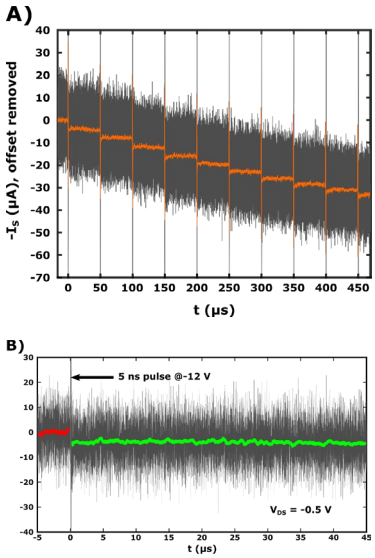

Fig. 4. Modulation behavior of a 30 × 5000 nm active area device with d_{PSG}=10 nm for -12 V, 5 ns gate pulses under V_{DS}=-0.5 V captured at 170 MHz. **(A)** Linear, gradual deprotonation of the channel over 10 pulses. **(B)** I_S(t) before, during, and after the first pulse on a finer time scale, clearly displaying impulse-like fast modulation characteristics. Figure adapted from [10].

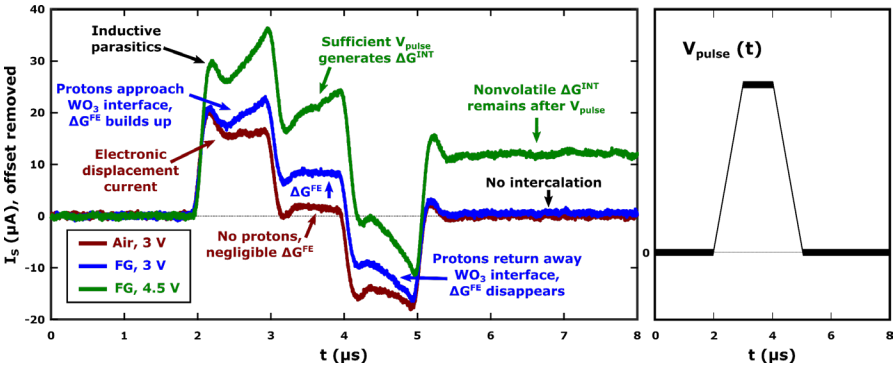

Fig. 5. I_s (left) under positive gate voltage pulse drive (right). Negligible ΔG^{FE} in the absence of protons (i.e., *before* FG treatment, red). *After* FG treatment (blue, green), a volatile ΔG^{FE} is observed. When V_{pulse} is sufficiently high (green), a nonvolatile ΔG^{INT} is generated. If V_{pulse} is small enough (blue), no intercalation takes place. Figure adapted from [10].