SOLID-STATE IONICS

Nanosecond protonic programmable resistors for analog deep learning

Murat Onen[1,2]*, Nicolas Emond[2,3], Baoming Wang[2,3], Difei Zhang[1,2], Frances M. Ross[2,3], Ju Li[2,3,4]*, Bilge Yildiz[2,3,4]*, Jesús A. del Alamo[1,2]*

Nanoscale ionic programmable resistors for analog deep learning are 1000 times smaller than biological cells, but it is not yet clear how much faster they can be relative to neurons and synapses. Scaling analyses of ionic transport and charge-transfer reaction rates point to operation in the nonlinear regime, where extreme electric fields are present within the solid electrolyte and its interfaces. In this work, we generated silicon-compatible nanoscale protonic programmable resistors with highly desirable characteristics under extreme electric fields. This operation regime enabled controlled shuttling and intercalation of protons in nanoseconds at room temperature in an energy-efficient manner. The devices showed symmetric, linear, and reversible modulation characteristics with many conductance states covering a 20× dynamic range. Thus, the space-time-energy performance of the all–solid-state artificial synapses can greatly exceed that of their biological counterparts.

Aqueous ionics with mobile Na^+, K^+, Ca^{2+}, and other ions underpin biological information processing. In neurons and synapses, the action potential of magnitude ~100 mV evolves over a characteristic time scale of milliseconds, which fundamentally limits the speed of thinking and reflexes of animals. Because liquid water decomposes at voltages >1.23 V, a weak action potential is understandable. However, with human-made solid-state neurons and synapses (*1*–*3*) for analog machine learning, research efforts are no longer limited by the stability window of aqueous electrolyte. Furthermore, it is also possible to fabricate devices that are much smaller than biological neurons, by a length-scale shrinkage factor of 10^3 (from ~10 μm to ~10 nm). This possibility raises a fundamental question about how much faster we can train such "artificial synapses"—i.e., what is the ultimate speed limit to solid-state ionics–based analog deep learning? Also, when we approach that speed limit, how energy intensive would the training be, given that dissipative processes generally produce more entropy per task the faster that task is accomplished?

Fundamental ionics arguments seem to call for high voltage and small length scales—that is, an extreme programming field approach (*4*–*10*). Transport of ions (such as H^+) inside a solid electrolyte (SE) layer and a mixed ionic-electronic conductor (MIEC) conductance channel layer, as well as charge-transfer reactions at the SE/MIEC interfaces, scale monotonically and nonlinearly with the applied voltage. Therefore the speed of change in the nonvolatile state of artificial neurons should increase with increasing electric field. However, there is a dielectric breakdown limit (*11*) to solids, as well as a thermodynamic electrochemical stability window for SEs. Therefore, the programming field should be raised as high as possible, but not so high as to permanently damage the SE. In this work, we demonstrate that such an approach produced exceedingly fast nanoionic devices (5 ns), at least 10^4 times as fast as biological synapses, when the voltage was on the order of +10/−8.5 V, yielding an extremely high electric field of ~1 V/nm across the proton-conducting phosphosilicate glass (PSG) SE. Surprisingly, we also show that under appropriate operating parameters, this nonlinear ionic-electronic device was robust and reversible, operating successfully over millions of cycles. Although the layout of our three-terminal device is analogous to that of a solid-state battery, it operated 10^{10} times as fast as conventional solid-state batteries. Finally, even when approaching the ultimate speed limit of our devices, the heat generated per programming task was still favorable compared with that generated by a human synapse (~10 fJ per state).

Interest in engineering the ideal programmable resistor for analog computing applications has skyrocketed owing to increasing workloads of deep learning problems (*12*–*14*) and diminishing expectations for transistor performance improvements through size scaling (*15*). Highly optimized digital application-specific integrated circuits that run reduced precision arithmetic operations can still manage inference tasks (*16*); however, the resolution requirements for training tasks do not allow sufficient bit reduction (*17*) and make beyond-digital approaches imperative. The core idea behind analog training accelerators is to process information locally using physical device properties instead of conventional Boolean arithmetic—i.e., using Ohm's and Kirchhoff's laws for the matrix inner product (*18*) and threshold-based updating for the outer product (*19*). However, the performance benefits attained by analog training processors are dependent on a set of very strict properties: The device must be fast, energy-efficient, nonvolatile, and reversible, and it must show symmetric conductance modulation with many conductance states across a large dynamic range (*20*, *21*). Unfortunately, readily available memory technologies [e.g., phase-changing (*22*, *23*), filamentary (*24*), bulk-switching (*25*–*28*), and ferroelectric (*29*) resistive memories] do not completely satisfy these requirements. As a result, demonstrations so far have required additional control circuitry and serial operations (*22*–*24*) to correct such imperfections, at the expense of substantially diminished space-time-energy performance.

Because the core idea of neural network training is to tune the state parameters (weights) through many small incremental modifications (*12*, *13*), it is reasonable that devices originally designed for information storage purposes (i.e., memory technologies) do not perform well in this information processing application. Devices for analog deep learning (i.e., training) must be primarily optimized for state transition (i.e., modulation) rather than long-term state preservation (i.e., nonvolatility). The retention of the weight parameters is obviously important such that the information learned from previous inputs is not lost faster than the ongoing optimization. However, given that the purpose of the network is not to memorize correct labels for the training dataset but to find generalized features instead, retaining exact values for extended durations of time loses relevance. The best example for this argument can be found in digital neural network training, for which it is a common practice to simulate weight decay on values that could otherwise be retained for indefinite time (*30*). Because the properties that ensure long retention times contradict those that enable high-speed and low-energy modulation (*31*), we conclude that for training implementations the focus should be on the latter two, trading retention down to the ~100-s–to–1000-s range for nanosecond and subpicojoule performance [see the supplementary materials (SM)].

Any memristive technology that aims to realize practical analog processors must contain a high density of devices built on top of standard complementary metal-oxide semiconductor (CMOS) logic and therefore must rely on CMOS-compatible materials and involve back-end-of-line–compatible processes.

[1]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. [2]MIT-IBM Watson AI Lab, 75 Binney St., Cambridge, MA 02142, USA. [3]Department of Materials Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA. [4]Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA 02139, USA.
*Corresponding author. Email: monen@mit.edu (M.O.); alamo@mit.edu (J.A.d.A.); byildiz@mit.edu (B.Y.); liju@mit.edu (J.L.)

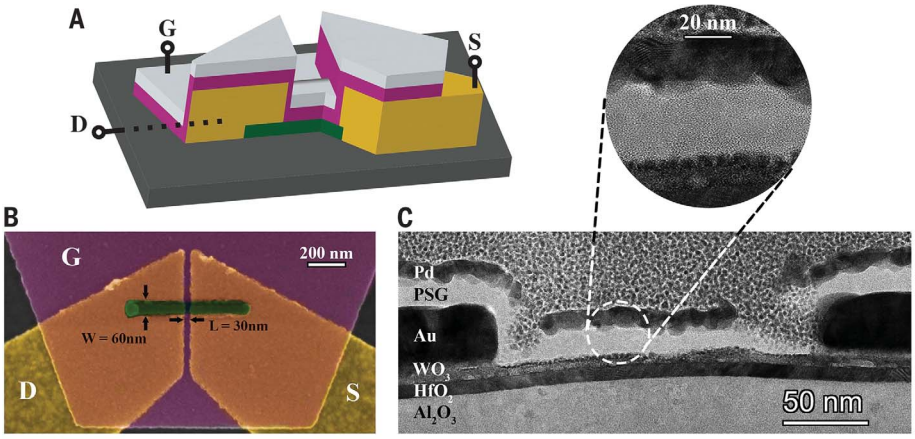

Fig. 1. Nanosecond protonic programmable resistors. (**A**) Three-dimensional illustration of the protonic programmable resistors studied in this work. Au (yellow), WO_3 (green), PSG (magenta), and Pd (gray) layers are indicated. As a result of an engineered sidewall, the Pd layer that overlaps with the channel electrodes is isolated from the remainder of the gate electrode. G, gate; D, drain; S, source. (**B**) False-colored top-view SEM image of a fabricated device with a 30-nm–by–60-nm channel. (**C**) Transmission electron microscopy (TEM) cross-sectional image of a protonic programmable resistor that had previously been extensively modulated in the ultrafast regime.

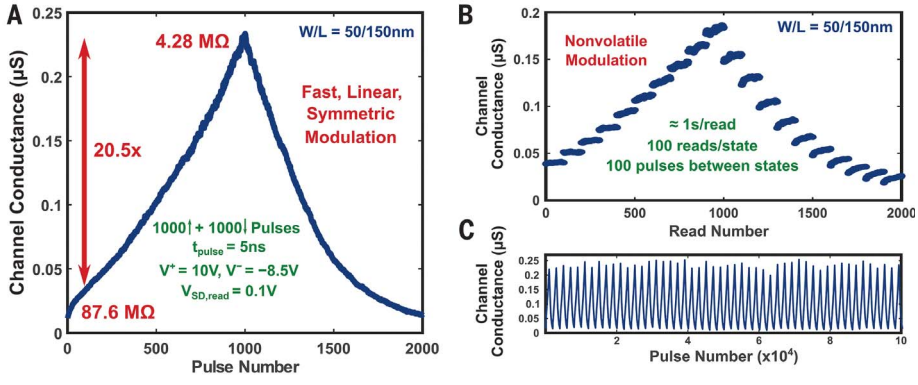

Fig. 2. Ultrafast and energy-efficient modulation characteristics of protonic programmable resistors. (**A**) Modulation performance of a 50-nm–by–150-nm protonic device with 10-nm PSG, showing fast (5 ns per pulse), nearly linear, and symmetric characteristics. W, width; L, length. (**B**) Retention behavior of the protonic device for ≈100 s at different conductance levels over the full dynamic range. (**C**) Endurance characterization of the protonic device, displaying nondegrading modulation over 10^5 pulses conducted over 30 hours.

Moreover, device operation should not depend on unconventional environmental conditions such as high temperature or humidity. With these concepts in mind, we developed an integration-friendly technology with inorganic solid materials that are native to conventional Si processing: WO_3 as the active channel material, nanoporous PSG as the protonic SE, and Pd as the hydrogen reservoir and controlling gate. The basic operating principle of the device relies on modulating the channel conductance via the electrochemically controlled intercalation of protons (chosen as the ion with the smallest radius and lightest mass) into WO_3 (*32*, *33*).

The rationale for choosing these materials is as follows: WO_3 is known for its conductance modulation and electrochromism upon ion intercalation (*25*, *32*–*34*) with well-established dynamics (*32*, *35*). For analog deep learning applications, the channel material must also have high base resistivity (*19*), proton insertion rate, and sensitivity. We found that superior modulation characteristics result from reactive sputtering of WO_3 at room temperature, followed by a 400°C annealing process that both oxidizes and crystallizes the material (see SM for process details). Regarding the electrolyte, the material must simultaneously show very high electronic resistivity and high proton conductivity. In our previous work, we identified nanoporous PSG as an outstanding SE (*36*–*38*) that displays both properties at room temperature without the need for hydration (i.e., humidity) (*33*). Finally, Pd was chosen as the gate metal and hydrogen reservoir (PdH_x), owing to its capability to take up and store hydrogen (*39*).

The three-terminal protonic programmable resistors studied in this work (Fig. 1A) were fabricated through a series of electron-beam lithography processes. A self-aligned gate structure was used to scale down device dimensions, avoiding mask alignment limitations. In a key design aspect of this process, the Pd layer was overlaid across a large region (Fig. 1B), which was then used as an etch mask for the PSG layer beneath. The height of the channel electrodes (Au) was calibrated such that the PSG layer could cover the sidewalls and the Pd layer that overlapped with the channel electrodes was disconnected from the gate electrode (Fig. 1C), avoiding the charging of those regions and unnecessary capacitance. This configuration was intended to maximize energy efficiency. The complete fabrication flow, additional metrology results, and yield details can be found in the SM.

Before testing, the devices were exposed to forming gas (3% H_2 in N_2) at room temperature for protonation of the Pd reservoir and then pumped down to vacuum. Future iterations of these protonic devices will be encapsulated to avoid environmental preconditioning steps. Figure 2A shows the channel conductance modulation of a 50-nm–by–150-nm device with 10-nm-thick PSG for 1000 protonating voltage pulses (V^+ = 10 V) followed by 1000 deprotonating pulses (V^- = −8.5 V). Between successive pulses, the channel conductance was read under drain-source voltage (V_{DS}) = 0.1 V and gate current (I_G) = 0 conditions and averaged for ~1 s. The devices displayed nearly ideal characteristics in terms of (i) high modulation speed, responding to 5-ns voltage pulses; (ii) nearly linear and symmetric behavior for incremental and decremental changes; (iii) conductance retention characteristics over durations longer than ~10^{10} times the unit pulse time (Fig. 2B); (iv) dynamic conductance range of 20×; (v) optimal base resistance of 88 megohms for readout (*19*, *32*); and (vi) preservation of these ideal properties without any degradation over extended time and use (Fig. 2C).

Moreover, the devices showed excellent energy efficiency under this ultrafast operation, the gate current supplied during each pulse being too small to be precisely measured for the small devices. The energy consumption during the transients was estimated to be ~2.5 fJ per pulse (see SM), which is a technology-agnostic overhead related to charging and discharging the gate capacitance. On the other hand, the energy consumed in proton transfer while the 5-ns voltage pulse was at its peak value was estimated to be ~15 aJ per pulse for the device whose performance is shown in Fig. 2. This value was based on a dc gate current measurement of ~30 nA at 10 V for a 750-nm–by–1000-nm device. This latter value is associated with the efficient shuttling of ions within the

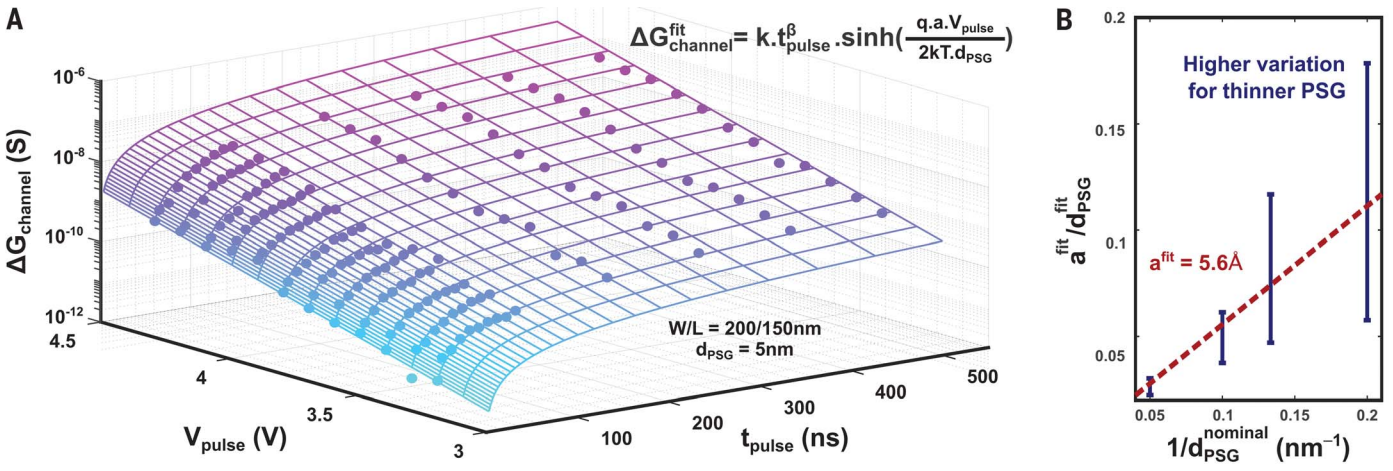

Fig. 3. Voltage dependence of conductance modulation. (**A**) Channel conductance change per pulse ($\Delta G_{channel}$) for different pairs of pulse amplitude (V_{pulse}) and duration (t_{pulse}) for a device with 5-nm PSG thickness. Points represent averaged experimental data over 1000 identical pulses for a given V_{pulse}-t_{pulse} pair, whereas the meshed surface represents the fitted result. The apparent hopping distance (a) and power factor (β) values used to fit the data are 5.6 Å and 1.2, respectively. The pulse time dependence of the protonation-deprotonation dynamics was empirically approximated as a power law ($\propto t_{pulse}^{1.2}$). (**B**) Estimation of the apparent hopping distance from repeating the same experiment detailed in Fig. 3A for 25 devices selected over four different chips with different PSG thicknesses. Error bars indicate the SD of values acquired from different devices with the same PSG thickness.

gate stack under the high electric field. Additional data and calculations related to the energy consumption can be found in the SM.

As a result, the combined material, processing, and performance benefits of the all–solid-state protonic devices demonstrated here exceed those of nonvolatile memory technologies.

To explain these desirable modulation characteristics, we have developed a model for device operation that consists of two key parts: (i) proton transport in the PSG and (ii) proton-coupled electron-transfer reaction rates at the PSG/electrode interfaces. Both processes have qualitatively similar formalisms and dependencies—the former is governed by hopping conduction in a disordered solid with random site energies (*4*–*7*, *10*), whereas the latter is determined by the Butler-Volmer charge-transfer kinetics (*8*, *9*). Conventionally, a high activation energy for protons to be freed from their sites (e.g., Si–O–H or P–O–H) results in low ionic conductivity (*40*) and ultimately limits the operation speed at room temperature. For amorphous SiO_2, this activation energy was reported to be 0.38 eV, which is much larger than the thermal energy in ambient conditions (25.9 meV) (*41*).

In the presence of a high electric field, the energy barrier to ion conduction is lowered in the field direction, yielding an enhanced proton hopping current

$$I_G(V_{pulse}) \propto \sinh\left(\frac{qaV_{pulse}}{2k_BTd_{PSG}}\right) \quad (1)$$

where q is the electron charge, k_B is the Boltzmann constant, T is the temperature, a is the apparent hopping distance, and d_{PSG} is the thickness of the electrolyte (*5*). Owing to the high resistance of the PSG layer, all of the pulse voltage is assumed to drop across the electrolyte ($V_{PSG} \sim V_{pulse}$; see SM for details). Figure 3A shows the experimentally observed conductance change per pulse ($\Delta G_{channel}$) as a function of pulse amplitude (V_{pulse}) and pulse time (t_{pulse}) for a device with d_{PSG} = 5 nm. Over a range of electric fields similar to that detailed in Fig. 2, the results shown in Fig. 3 closely followed Eq. 1, as indicated by the lines. Furthermore, experiments for devices with different d_{PSG} values allowed us to extract a hopping distance for the PSG of 5.6 Å (Fig. 3B). This result is in good agreement with previously reported values for amorphous silica glasses (*41*–*43*).

In our case, the electric field (≈1 V/nm) across the hopping distance ($a \approx 0.5$ nm) was so high that it might completely remove the activation barrier within the PSG (≈0.4 eV). This effect resolves the bottleneck of low proton conduction at room temperature, thus enabling high-speed operation. Such an effect was predicted in simulations (*44*) but has not been previously observed experimentally, as the required conditions are beyond the breakdown field or the electrochemical stability window of traditional electrolyte materials (*45*). Instead, PSG allows a high critical field (8 to 15 MV/cm) (*46*, *47*) in addition to a moderate base proton conductivity (*36*–*38*), making it an ideal electrolyte choice for this application. Moreover, under the classical statistical mechanics scenario, reducing the migration energy below a few k_BT boosts the likelihood of classical or quantum ballistic motion of protons. Indeed, in liquid water, proton motion is known to have a pronounced quantum character, with activationless quantum nuclear dynamics in some exchange events between water molecules (*48*). Thus, under such high electric fields, "quantum ballistic" transport of protons in solids may be realized.

Unlike the conductive filament devices controlled by angstrom-scale short-range ion motion at the tip of the filament (*49*), our protonic devices rely on long-range (~10 nm) and uniform partitioning of protons across the electrolyte with the protons then behaving as dynamic dopants in the entire H_yWO_3 channel (*32*, *35*). Furthermore, conducting filament formation is assisted by Joule heating, with local temperatures rising as high as ~500 K (*50*), whereas in our case ion transport is facilitated only by the electric field with negligible temperature increase ($\Delta T \sim 2$ K; see SM). As a result, although very-short-range ion transport at the tip of the filament can readily occur in subnanosecond time scales, the dependence of filament properties on the local microstructure and microchemistry makes those devices unpredictable and stochastic (*2*, *9*, *49*). On the contrary, devices studied in this work are modulated by macroscopic chemical and electronic modifications, which result in distinctly controllable and deterministic performance characteristics. These benefits make their acceleration to the nanosecond regime highly relevant.

The ability to rapidly shuttle protons within bulk PSG at extreme speeds shifts the bottleneck of proton transport to the interfaces. Although the hydrogen uptake rate of Pd is high (*51*) and the PdH_x/PSG interfacial reaction is likely also efficient, the same cannot be said for proton insertion into polycrystalline WO_3 (*52*), which shows up as a PSG/H_yWO_3

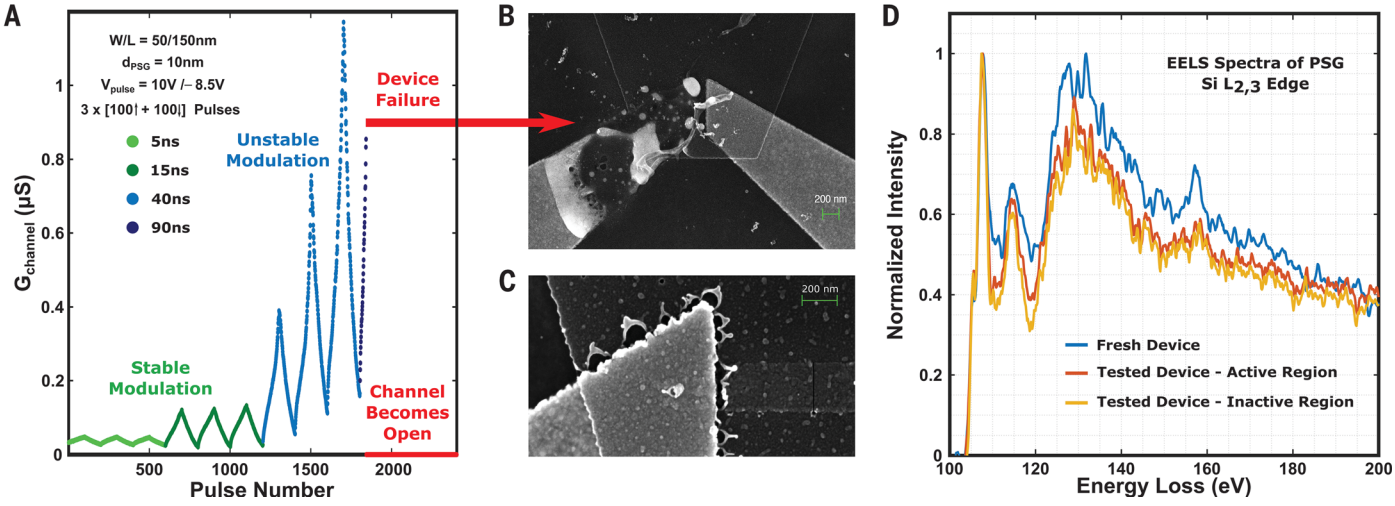

Fig. 4. Modulation dynamics for short and long pulse durations. (**A**) Channel conductance modulation of the device whose performance is presented in Fig. 2 for increasing pulse duration under the same programming voltages. Between pulses, the channel conductance is read and averaged for ≈1 s. (**B**) SEM image of the device after the experiment shown in (A). (**C**) SEM image of another device captured during early degradation. (**D**) Si-$L_{2,3}$ EELS spectra of the PSG layer for tested (red and orange) and fresh (blue) devices. The testing conditions of the device were identical to those given in Fig. 2, operating under the ultrafast regime only.

interfacial charge-transfer resistance. Because protons do not have high diffusivity in H_yWO_3 (*53*), it is likely that the inserted species cannot quickly vacate their sites near the interface, thus further reducing the insertion rate in a self-limiting fashion. These factors result in an excess of highly energetic protons at the WO_3/PSG interface and may lead to H_2 gas formation and buildup at the interface (*54*).

The electrical signature and morphological consequences of long–pulse-time stressing of the devices are captured in Fig. 4. In Fig. 4A, the device previously characterized in Fig. 2 was tested under the same V_{pulse} but for increasing t_{pulse}. Above 40 ns a cascading effect was apparent, in which the conductance change increased with each pulse, ultimately causing device failure for 90-ns pulses. A scanning electron microscopy (SEM) image of the device after this experiment is shown in Fig. 4B, whereas Fig. 4C shows an image of another device stopped at earlier stages (corresponding to the 40-ns pulse regime shown in Fig. 4A). These images show damage features that we suggest were consistent with H_2 gas evolution at the PSG/H_yWO_3 interface, nanobubble formation, and stress buildup.

Most importantly, as shown in Fig. 2C, no degradation was observed under high-speed (5 ns) operation. To provide further evidence of good endurance characteristics in the ultrafast regime, we performed electron energy-loss spectroscopy (EELS) for the SE in fresh and tested devices. Figure 4D shows that the Si $L_{2,3}$ energy-loss spectra were similar in active and inactive regions across the PSG layer, which indicates that there was no stoichiometry change or that the change was too subtle to be detected. Considering that the widest electrochemical stability window for room temperature protonic electrolytes was previously ~3.35 V (*55*), we attribute our devices' ability to operate stably and reversibly at 10 V/–8.5 V to the operation at nanosecond transients, where the PSG does not have to conform to the thermodynamic stability requirement for electrolytes (established after quasi-static wait times). Furthermore, these short time scales may simply be too fast for oxygen motion (slower than that of protons), which would otherwise cause degradation of material properties. We believe these are two key physical dynamics underlying the no-damage "protonic breakdown," a transient phenomenon that we took advantage of in this novel ultrafast long-range transport regime.

In summary, we explored the limits of all–solid-state ionics to answer the question of how much faster an electrochemical artificial synapse can operate relative to its biological counterpart. Under extreme electric field conditions, we demonstrated ultrafast modulation of nanoscale protonic programmable resistors with outstanding energy efficiency. Moreover, the devices exhibited many nonvolatile channel conductance states with good retention across a large and ideal dynamic range, which could be programmed reversibly, repeatedly, and symmetrically. Operation of the devices at extreme electric fields across the stack without any material degradation is the key breakthrough to enable such performance. Under these conditions, the activation energies that govern proton transport as well as charge-transfer reactions at the interface were substantially lowered, thus resulting in operation speed enhanced by six orders of magnitude at room temperature. Beyond artificial synapses, these findings open up possibilities in fields that require fast ion motion, such as microbatteries, artificial photosynthesis, and light-matter interactions.

REFERENCES AND NOTES

1. V. K. Sangwan, M. C. Hersam, *Nat. Nanotechnol.* **15**, 517–528 (2020).
2. Q. Xia, J. J. Yang, *Nat. Mater.* **18**, 309–323 (2019).
3. M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **1**, 22–29 (2018).
4. B. Roling, L. N. Patro, O. Burghaus, M. Gräf, *Eur. Phys. J. Spec. Top.* **226**, 3095–3112 (2017).
5. A. Röthel, S. Friedrich, R. Lühning, L. Heuer, *Z. Phys. Chem.* **224**, 1855–1889 (2010).
6. L. Onsager, *Science* **166**, 1359–1364 (1969).
7. M. J. Dignam, *J. Phys. Chem. Solids* **29**, 249–260 (1968).
8. T. Erdey-Grúz, M. Volmer, *Z. Phys. Chem.* **150A**, 203–213 (1930).
9. D. Ielmini, *IEEE Trans. Electron Dev.* **58**, 4309–4317 (2011).
10. S. Menzel, U. Böttger, M. Wimmer, M. Salinga, *Adv. Funct. Mater.* **25**, 6306–6325 (2015).
11. F. Palumbo *et al.*, *Adv. Funct. Mater.* **30**, 1900657 (2020).
12. W. Fedus, B. Zoph, N. Shazeer, arXiv:2101.03961 [cs.LG] (2021).
13. T. B. Brown *et al.*, in *Advances in Neural Information Processing Systems 33 (NeurIPS2020)*, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, 2020), pp. 1877–1901.
14. E. Strubell, A. Ganesh, A. McCallum, in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, L. Màrquez, Eds. (Association for Computational Linguistics, 2019), pp. 3645–3650.
15. I. L. Markov, *Nature* **512**, 147–154 (2014).
16. J. Choi *et al.*, in *Proceedings of the 2nd SysML Conference* (2019); https://mlsys.org/Conferences/2019/doc/2019/168.pdf.
17. X. Sun *et al.*, in *Advances in Neural Information Processing Systems 32 (NeurIPS2019)*, H. Wallach *et al.*, Eds. (Curran Associates, 2019); https://proceedings.neurips.cc/paper/2019/file/65fc9fb4897a89789352e211ca2d398f-Paper.pdf.
18. K. Steinbuch, *Kybernetik* **1**, 36–45 (1961).
19. T. Gokmen, Y. Vlasov, *Front. Neurosci.* **10**, 333 (2016).
20. T. Gokmen, M. Onen, W. Haensch, *Front. Neurosci.* **11**, 538 (2017).
21. S. Agarwal *et al.*, in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)* (IEEE, 2016), pp. 929–938.
22. G. W. Burr *et al.*, *IEEE Trans. Electron Dev.* **62**, 3498–3507 (2015).
23. S. Ambrogio *et al.*, *Nature* **558**, 60–67 (2018).
24. H. Jiang *et al.*, *Sci. Rep.* **6**, 28525 (2016).
25. S. Kim *et al.*, in *2019 IEEE International Electron Devices Meeting* (IEEE, 2019), pp. 847–850.
26. Y. Li *et al.*, *Adv. Mater.* **32**, e2003984 (2020).
27. Y. van de Burgt *et al.*, *Nat. Mater.* **16**, 414–418 (2017).
28. E. J. Fuller *et al.*, *Adv. Mater.* **29**, 1604310 (2017).
29. A. Chanthbouala *et al.*, *Nat. Mater.* **11**, 860–864 (2012).
30. A. Krogh, J. A. Hertz, in *Advances in Neural Information Processing Systems 4 (NIPS 1991)*, J. Moody, S. Hanson, R. P. Lippmann, Eds. (Morgan-Kaufmann, 1991), pp. 950–957.

31. D. B. Strukov, R. S. Williams, . *Appl. Phys. A* **94**, 515–519 (2009).
32. X. Yao *et al*., *Nat. Commun.* **11**, 3134 (2020).
33. M. Onen, N. Emond, J. Li, B. Yildiz, J. A. del Alamo, *Nano Lett.* **21**, 6111–6116 (2021).
34. J. Tang *et al*., in *2018 IEEE International Electron Devices Meeting* (IEEE, 2018).
35. A. Hjelm, C. G. Granqvist, J. M. Wills, *Phys. Rev. B* **54**, 2436–2445 (1996).
36. S. Prakash, W. E. Mustain, S. H. Park, P. A. Kohl, *J. Power Sources* **175**, 91–97 (2008).
37. M. Nogami *et al*., *J. Electrochem. Soc.* **151**, A2095 (2004).
38. D. Bhusari, J. Li, P. J. Jayachandran, C. Moore, P. A. Kohl, *Electrochem. Solid-State Lett.* **8**, A588 (2005).
39. Y. Li, Y. T. Cheng, *Int. J. Hydrogen Energy* **21**, 281–291 (1996).
40. R. Oesten, R. A. Huggins, *Ionics* **1**, 427–437 (1995).
41. R. A. B. Devine, G. V. Herrera, *Phys. Rev. B* **63**, 233406 (2001).
42. J. Godet, A. Pasquarello, *Phys. Rev. Lett.* **97**, 155901 (2006).
43. M. E. S. Beatty, E. I. Gillette, A. T. Haley, D. V. Esposito, *ACS Appl. Energy Mater.* **3**, 12338–12350 (2020).
44. M. Kunow, A. Heuer, *J. Chem. Phys.* **124**, 214703 (2006).
45. W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, G. Ceder, *Chem. Mater.* **28**, 266–273 (2016).
46. E. H. Snow, B. E. Deal, *J. Electrochem. Soc.* **113**, 263 (1966).
47. D. Serghi, C. Pavelescu, *Thin Solid Films* **186**, L25–L28 (1990).
48. R. Vuilleumier, D. Borgis, *J. Phys. Chem. B* **102**, 4261–4264 (1998).
49. D. Ielmini, *Semicond. Sci. Technol.* **31**, 063002 (2016).
50. M. von Witzleben *et al*., *Adv. Electron. Mater.* **3**, 1700294 (2017).
51. F. Lewis, *Platin. Met. Rev.* **4**, 132–137 (1960).
52. P. G. Dickens, S. Crouch-Baker, M. T. Weller, *Solid State Ion.* **18–19**, 89–97 (1986).
53. S. Burkhardt, M. T. Elm, B. Lani-Wayda, P. J. Klar, *Adv. Mater. Interfaces* **5**, 1701587 (2018).
54. S. Nakabayashi, R. Shinozaki, Y. Senda, H. Y. Yoshikawa, *J. Phys. Condens. Matter* **25**, 184008 (2013).
55. S. Wang *et al*., *Adv. Mater.* **34**, 2202063 (2022).
56. M. Onen *et al*., Data for Nanosecond Protonic Programmable Resistors for Analog Deep Learning, Zenodo (2022); doi.org/10.5281/zenodo.6506857.

ACKNOWLEDGMENTS

We thank J. Rozen, T. Todorov, P. Solomon, T. Gokmen, and M. Rasch from the IBM T. J. Watson Research Center for program support and fruitful discussions. Device fabrication took place at the facilities of MIT.nano and the MIT Electron-Beam Lithography Facility. Materials characterization took place at the facilities of MIT's Materials Research Laboratory. **Funding:** This project is funded by the MIT-IBM Watson AI Lab. N.E. would like to thank the Fonds de recherche du Québec et Nature et technologies (FRQNT) for financial support. **Author contributions:** M.O. conceived the original idea, fabricated devices, and performed experimental characterization. N.E. optimized the channel material and performed metrology studies. M.O. and N.E. designed experiments and analyzed data. B.W. designed and conducted TEM characterizations. All authors contributed to developing the theory and writing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials and can be found in Zenodo (*56*).

SUPPLEMENTARY MATERIALS

Submitted 26 February 2022; accepted 28 June 2022
10.1126/science.abp8064

Science
AAAS

Supplementary Materials for

Nanosecond protonic programmable resistors for analog deep learning

Murat Onen *et al*.

Corresponding authors: Murat Onen, monen@mit.edu; Ju Li, liju@mit.edu; Bilge Yildiz, byildiz@mit.edu; Jesús A. del Alamo, alamo@mit.edu

Science **377**, 539 (2022)
DOI: 10.1126/science.abp8064

The PDF file includes:

1. Fabrication Flow

- Atomic Layer Deposition (ALD) of 10/40 nm HfO_2/Al_2O_3 on 1x1 cm^2 SiO_2/Si pieces.
- Patterning of poly(methyl methacrylate) (PMMA, e-beam resist) with Elionix FLS-125 for channel layer lift-off.
- Reactive sputtering of WO_3 layer from metallic target at room temperature in O_2/Ar RF plasma using AJA sputtering system.
- Annealing of the WO_3 layer in 8:2 N_2:O_2 environment at 400 °C for 1 hour following a liftoff step.
- Patterning of poly(methyl methacrylate) (PMMA, e-beam resist) with Elionix FLS-125 for source/drain contact layer lift-off.
- Electron-beam evaporation of 35/5 nm of Au/Cr layer using AJA evaporation system, followed by lift-off step.
- Plasma-Enhanced Chemical Vapor Deposition (PECVD) of PSG layer using 1420 sccm N_2O, 12 sccm SiH_4, and 12 sccm PH_3 (2% in H_2) at 100°C, with a RF plasma power of 60 W at 380 kHz.
- Patterning of poly(methyl methacrylate) (PMMA, e-beam resist) with Elionix FLS-125 for gate contact layer lift-off.
- Electron-beam evaporation of 10 nm of Pd layer using AJA evaporation system, followed by lift-off step.
- Reactive Ion Etching (RIE) of the PSG layer using Pd layer as the hard mask, under CF_4 plasma at 100W.
- Patterning the bilayer of poly (methylglutarimide) and Microposit S1813 positive tone photoresist, using Heidelberg-MLA 150 for pad layer lift-off.
- Electron-beam evaporation of 150/15 nm of Au/Cr layer using AJA evaporation system, followed by lift-off step.

Regarding the yield rate of this process, a total of 4056 devices were fabricated on 6 1x1 cm^2 chips with 4 different PSG and 2 different WO_3 thicknesses. Out of the 246 devices we measured, 17 were shorted (the most aggressively scaled devices), 114 were too resistive to operate and 115 functioned successfully giving a functional yield of 50%. Note that the yield is significantly higher, 75%, for the 55 devices with thicker WO_3 layer.

Note that in those experiments, the criterion for being functional was the ability to reversibly modulate channel conductance, yielding modulation characteristics similar to those shown in

Figures 2 and **3**. Since most device geometries as well as the operation conditions were different, we do not have enough statistical data to draw conclusions regarding device-to-device variation.

Over 9 months of experimentation, we have observed minor degradation effects, where the base conductance of devices increased. We attribute this change to self-protonation of the inactive devices due to proton leakage across PSG (as all Pd layers are protonated together in the gas environment, even for untested devices). Importantly, these changes could be easily reversed by application of negative pulses, deprotonating the channel such that high resistivity states can be restored. Moreover, similar to self-protonation, a self-deprotonation was also observed following the exposure of devices to oxygen (air) which quickly sinks all hydrogen out of the PdH_x reservoir. Encapsulation of these devices and controlled protonation of the Pd layer during fabrication steps should ultimately mitigate this issue. Devices showed no appreciable change when they were stored in a N_2 box.

2. Metrology

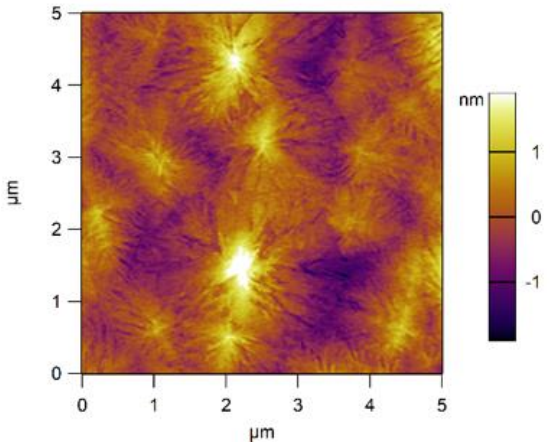

Figure S1. Atomic force microscopy (AFM) image of the reactive sputtered WO_3 layer after the annealing step showing surface topography deviations below +/-1nm.

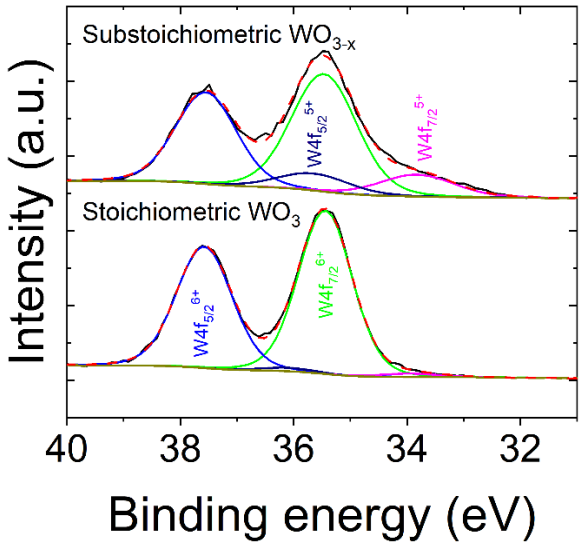

Figure S2. X-ray photoelectron spectroscopy (XPS) of the atomic layer deposited WO_3 layer (top) used in our previous work (*33*) and the reactive sputtered + annealed WO_3 used in this work. The annealing step oxidizes the WO_{3-x} layer to fully-stoichiometric WO_3 which is has less base conductance and provides higher sensitivity for protonation.

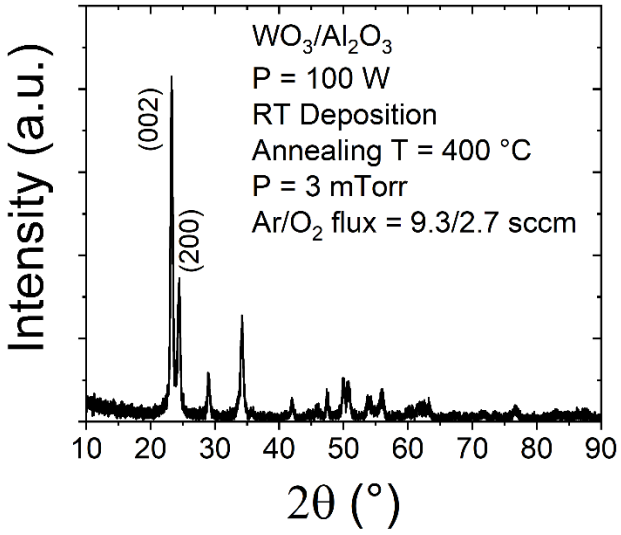

Figure S3. X-ray diffraction (XRD) spectra of the reactive sputtered WO_3 after annealing step showing characteristics of polycrystalline material in monoclinic phase.

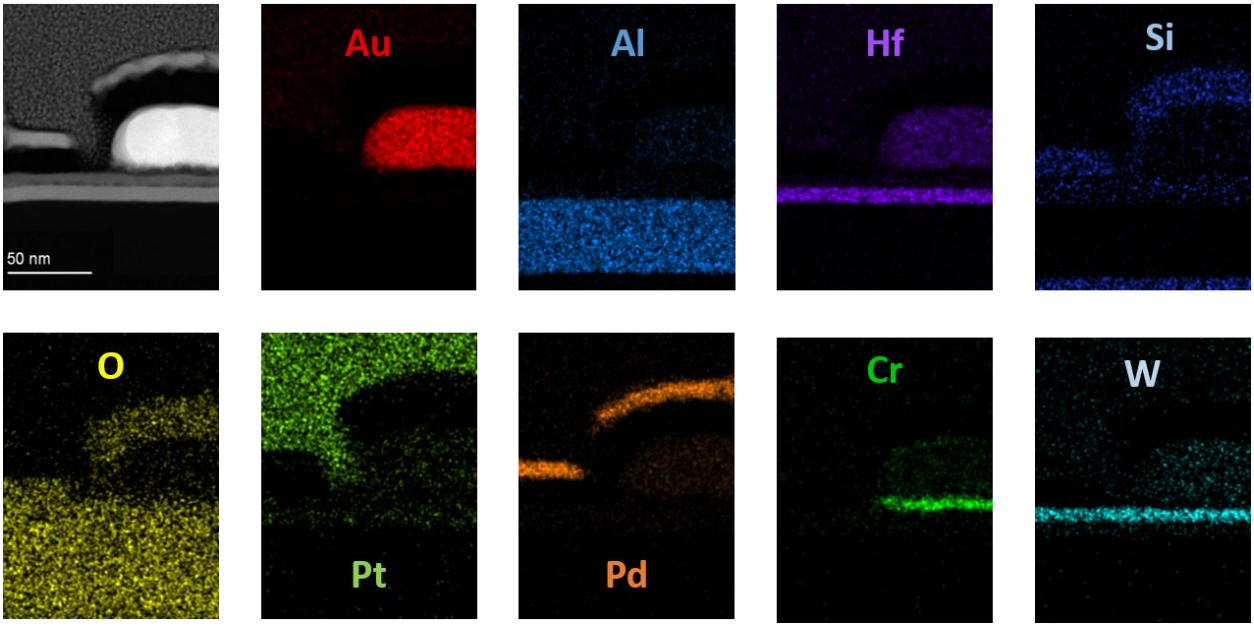

Figure S4. Energy Dispersive Spectroscopy (EDS) mapping in the transmission electron microscope (TEM) of the elements in a protonic memristor studied in this work.

3. Experimental Setup

All electrical measurements were performed using a micro-probe station enclosed chamber (MPS-PT) manufactured by NEXTRON, Korea. In experiments performed in forming gas (FG, 3% H_2 in N_2) all 4 DC probes were first connected to the pads of the device under test (1 to Source, 1 to Drain and 2 to Gate terminals). The chamber was then purged by flowing in FG for 60 s with both gas inlet and outlet open, followed by another 60 s with outlet shut to create positive pressure inside the chamber, ultimately pumped down to vacuum.

Three of the probes (1 Source, 1 Drain and 1st Gate) were then connected to the Source Measurement Units (SMUs) of a Keysight B1500 Semiconductor Analyzer, while the fourth probe (2nd Gate) was connected to the Pulse Generation Unit (PGU) of the same instrument. The experiment sequence and data acquisition were controlled via an in-house developed MATLAB suite.

During the application of the pulses, $V_{DS} = 0$ was maintained using the SMUs. On the other hand, for reading of the channel conductance, V_{DS} was swept between -0.1 and 0.1V while gate terminal was floated by the SPGU. Asymmetric pulse voltages during protonation/deprotonation was chosen to ensure better modulation symmetry. The 5 ns pulses used in **Figure 2** were at the limits of the Pulse Generation Unit of the Keysight B1500 Semiconductor Analyzer we used (see **Figure S8** for waveform).

4. Additional Results

To estimate the pulse energy, we first attempted to read the I_G and V_G during the application of the pulse, using a Keysight CX3324 waveform analyzer. However, due to the very low gate

currents for devices with small areas (such as the ones in **Figures 1-4**), we could not resolve those values at high pulse speeds. Therefore, as an alternative, we experimented with a larger device from the same chip used for **Figure 2** (10 nm thick PSG) with W = 1000 nm and L = 750 nm. At the same gate voltage of 10V used in **Fig. 2**, an average gate current of 30.5 nA was recorded. Assuming area scaling, an estimate gate current of 300 pA should be passing through the device in **Fig.2**. Using this value, the static energy dissipation (when the pulse voltage is constant at its peak value) is calculated as E_{static} = 10V × 300 pA × 5 ns = 15 aJ, which is the energy dissipated during proton shuttling.

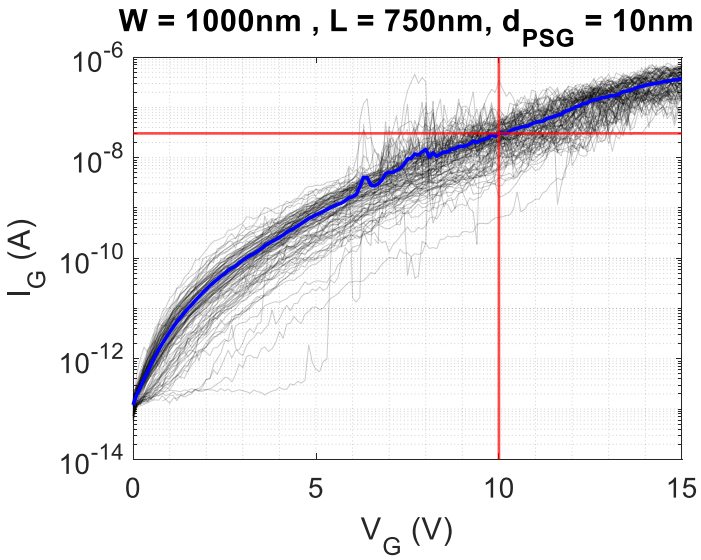

Figure S5. I_G-V_G curve for a large (1000nm × 750nm) protonic device used to estimate the gate current of the smaller device from the same chip shown in **Fig. 2**. At 10V, average recorded gate current is 30.5nA (marked with red lines).

On the other hand, during the charging and discharging of the gate capacitance (i.e. rise and fall times of the pulses), there exists a dynamic energy dissipation, which is computed as $E_{dynamic} = C_{gate}V_{gate}^2$, where $C_{gate} = \frac{\epsilon.A_{gate}}{d_{PSG}} = \frac{4\epsilon_o \times 150\ nm \times 50\ nm}{10\ nm} = 26.55$ aF yielding a $E_{dynamic}$ = 100V² × 26.55 aF = 2.655 fJ.

To estimate the maximum possible temperature increase in these devices, we first calculate the power under 50% pulse-width-modulation P = 2.66 fJ / 10 ns = 0.27 µW. Then using a simple 1-D model and taking the thermal conductivity of both PSG and WO_3 at k ~ 1.4 W/m.K, we find the thermal resistance R = 20 nm / [(30 × 60 nm²).(1,4 W/m.K)] = 8.10⁶ K/W, yielding a maximum temperature increase T = 0.25 µW × 8.10⁶ = 2K, which we therefore neglected.

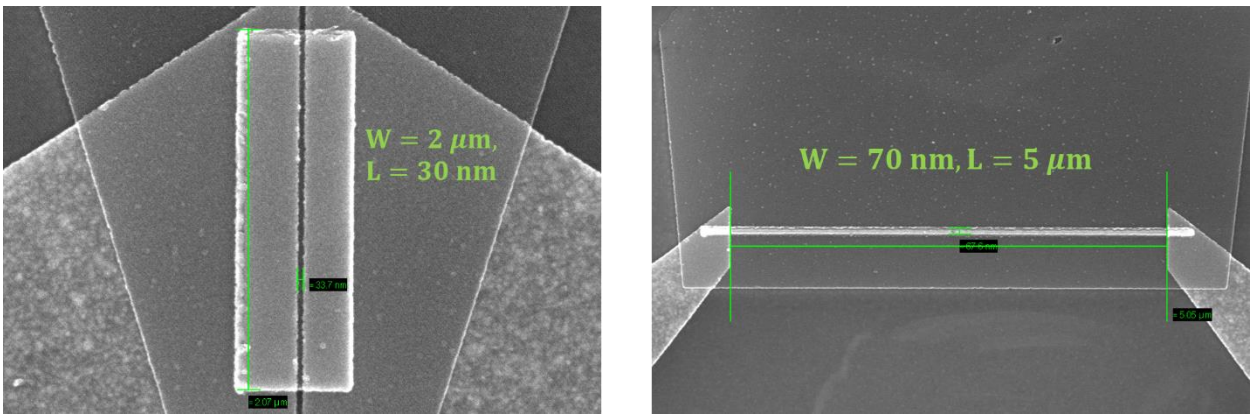

Figure S6. Scanning electron microscopy (SEM) images of protonic devices with higher aspect ratios fabricated and tested in this study.

We also note that in Eq.1 we wrote V_{pulse} to denote the voltage across the PSG electrolyte. The reason we assume $V_{PSG} \sim V_{pulse}$ is that when we quadruple the thickness of the PSG (which should quadruple R_{PSG}), the voltage required to get the same operation increases by 4×.If, say R_{PSG} were to be equal to R_{others}, where R_{others} cover all electronic and ionic resistances at the bulk and interfaces of the materials that are not bulk PSG, then quadrupling the thickness would have required only 2.5× the voltage. Therefore, we conclude $R_{PSG} >> R_{others}$ and thus $V_{pulse} \sim V_{PSG}$. Results from devices with four different PSG thicknesses is given below in **Fig. S7** showing overlapping modulation curves for V_{pulse}/d_{PSG}.

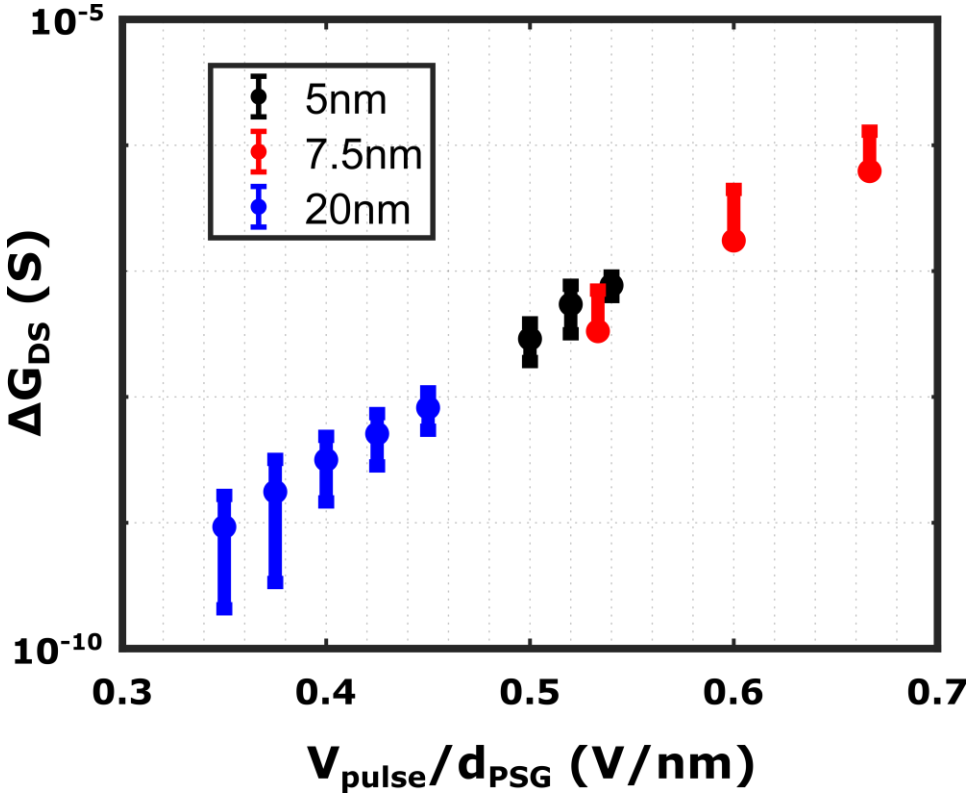

Figure S7. Scaling of voltage linearly with PSG thickness, indicating that the PSG resistance is higher than the other resistances in the gate stack. Error bars refer to the standard deviation of the channel conductance change at a given point for incremental and decremental directions.

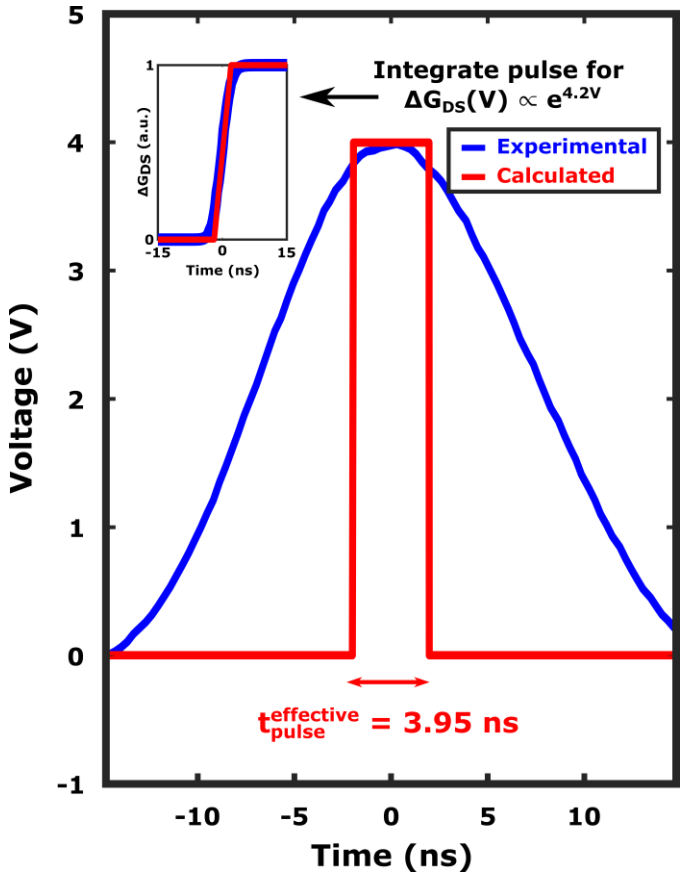

Figure S8. The voltage waveform (blue) of the fastest pulses (referred to as 5 ns in the manuscript) captured by Tektronix MSO44 Waveform Analyzer (1 GHz) and its calculated equivalent square pulse (red). To calculate the effective pulse time, the experimental waveform was first integrated using the function $\Delta G\_DS\ (V) \propto e^{4.2V}$ acquired by the fitting displayed in Figure 3. Then, the same was applied for a perfect square wave with same maximum amplitude, yielding an estimated effective pulse time of ~ 4 ns.

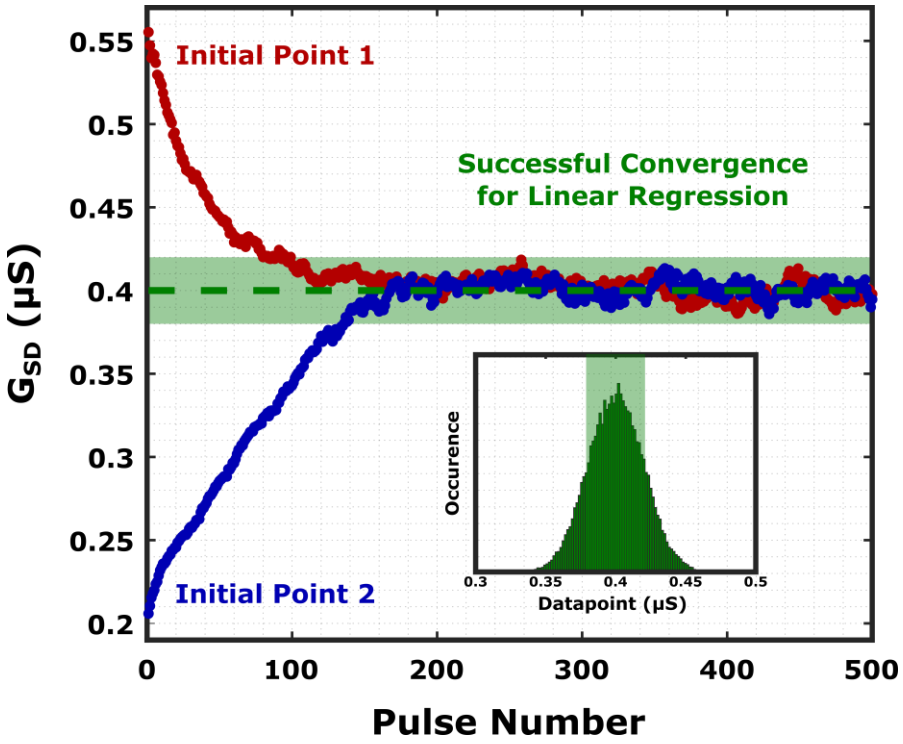

Figure S9. Single-parameter linear regression experiment conducted on a nanoscale protonic programmable resistor (W = 150 nm, L = 75 nm, d_{PSG} = 10 nm) programmed with voltage pulses of ± 10 V, 5 ns. The dataset was artificially generated to be around 0.4 μS with a standard deviation of ± 0.2 μS. For both initial points above and below this optimum value, device conductance converged to the target region and remained in the vicinity of this optimum, evidencing successful optimization.

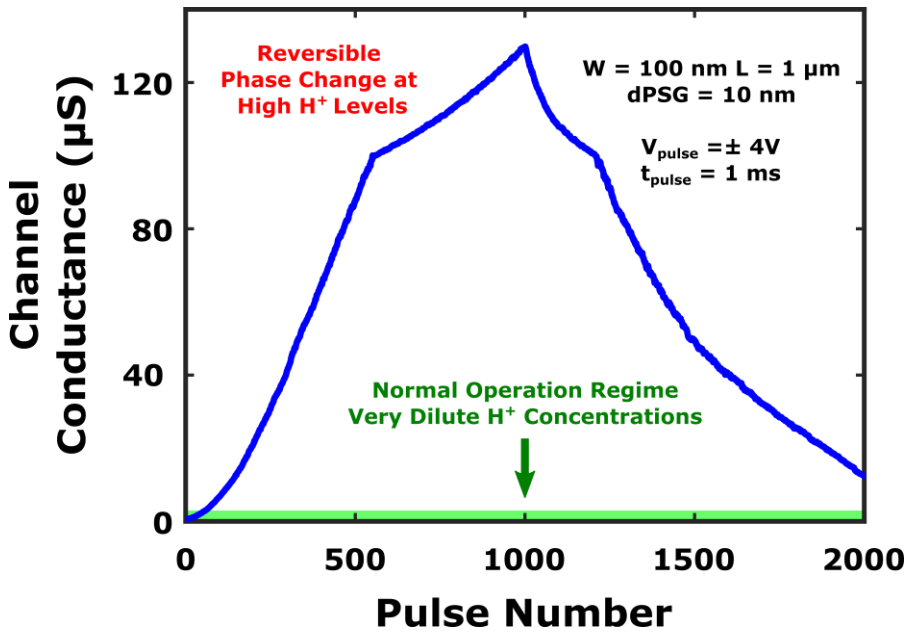

Figure S10. Conductance modulation behavior of a protonic programmable resistor, covering a wide range of proton concentrations. A reversible phase transition is evidenced, which is in good agreement with the suggestions of previous literature for high concentrations of protons in WO_3 (*57*).

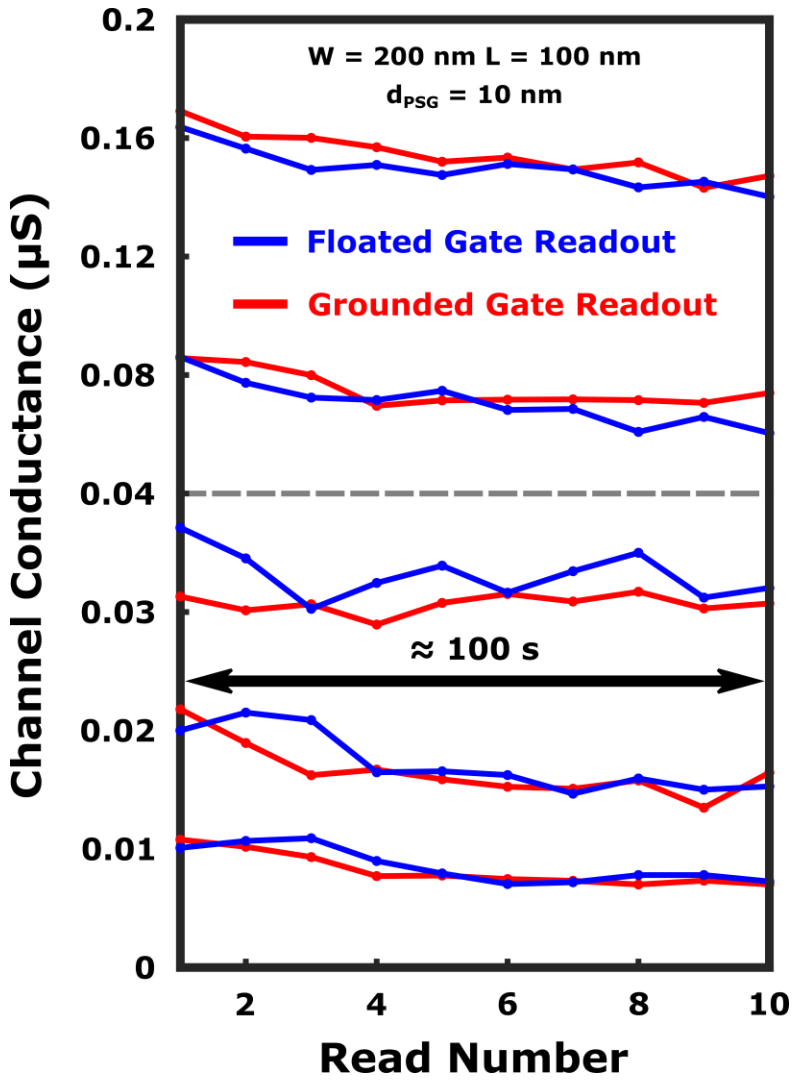

Figure S11. Retention characteristics for ≈ 100 s for floated and grounded gate conditions (W = 200 nm, L = 100 nm, d_{PSG} = 10 nm) programmed at different conductance states. No noticeable difference was observed for two different gate biasing conditions.

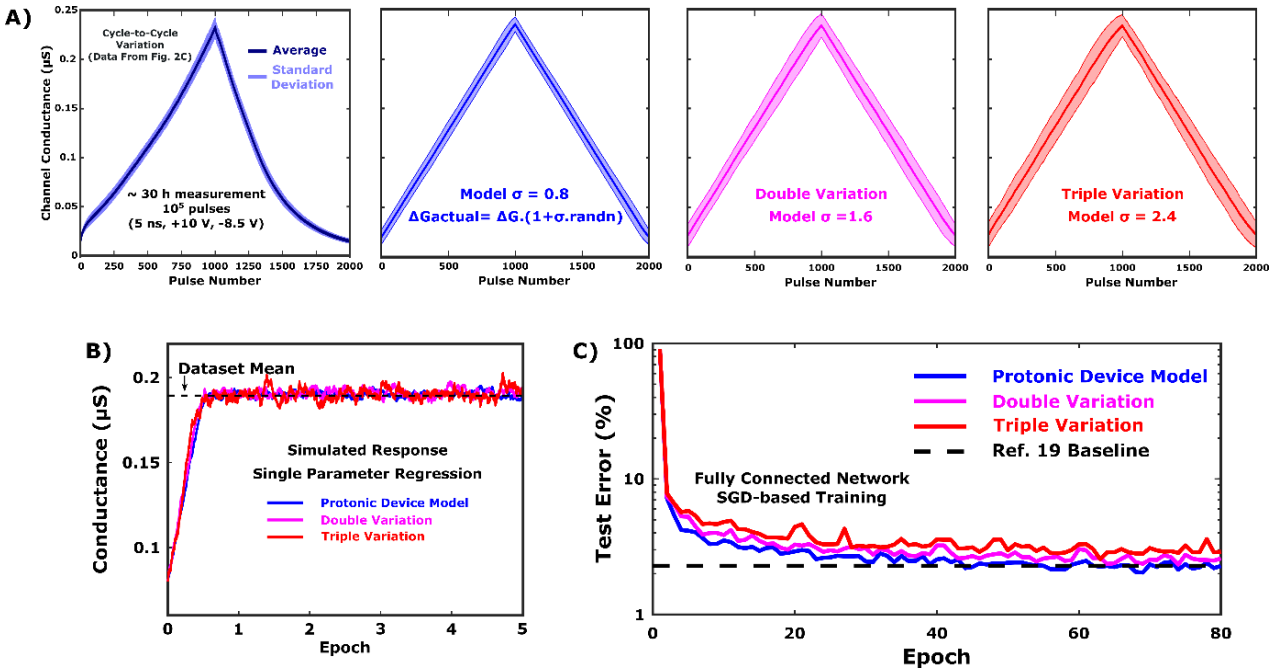

Figure S12. (A) Modulation variation analysis for the results presented in **Figure 2**C and modeling of variation where the experimental data is fit with the equation $\Delta G^{actual} = \Delta G \times (1 + \sigma * randn)$, where $randn$ is a random variable with normal distribution. A σ value of 0.8 was used to reproduce the experimental behavior. The center line shows the average conductance modulation at each pulse whereas the shaded are signifies their standard deviation. **(B)** Simulated single-parameter linear regression for device models in (A), replicating the task experimentally shown in **Fig.S9**. **(C)** The learning curve for a fully connected neural network trained on MNIST dataset following the same methodology presented in Ref. (19). Results show similar classification error results for different amounts of modulation variation displaying high variation tolerance of analog deep learning.

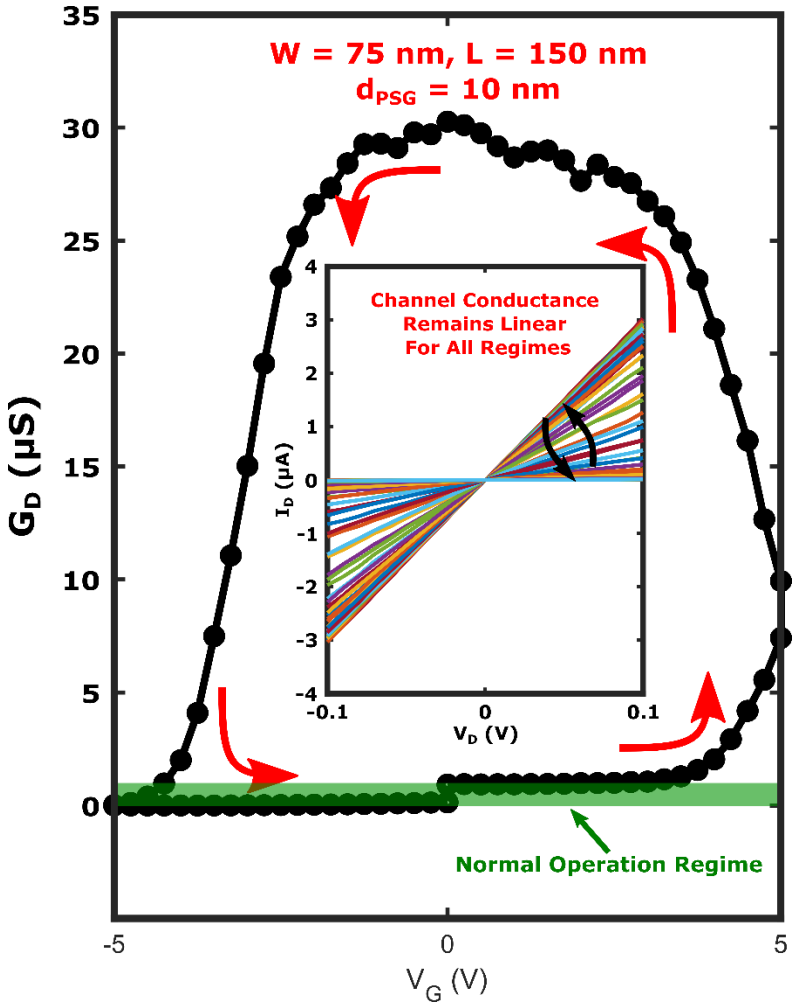

Figure S13. ID-VG sweep for a protonic programmable resistor (W = 75 nm, L = 150 nm, dPSG = 10 nm). The linearity of the channel conductance remains intact at all gate voltages. The ID-VD sweep at each VG takes ~5 s, depending on the channel current levels due to auto-ranging (i.e. lower currents, longer integration times). As a result of applying positive voltage for prolonged times, the channel becomes protonated during the sweeps, which was evidenced as an increase in conductance at positive gate voltages followed by a decrease for negative ones. Note that the regular operation regime studied in this manuscript was the low-conductance with dilute proton concentrations, marked by the green region.

5. Assessing the Retention Requirement for Analog Deep Learning Applications

As discussed in the main body, devices for analog deep learning (i.e. training), need to be primarily optimized for state transition (i.e. modulation) rather than state preservation (i.e. nonvolatility). The retention of the weight parameters is obviously important such that the information learned from previous inputs are not lost faster than the still-ongoing optimization. However, given the purpose of the network is not to memorize correct labels for the training dataset but to find generalized features instead, retaining exact values for extended durations of time loses relevance.

The question then becomes, how long retention is enough for accurate training? First of all, the answer would obviously be smaller than the total training time, and longer than the cycle-time (i.e. the time it takes between an input is presented to a given weight and it receives the corresponding update). We suggest that the time-constant of the (exponential-assumed) retention-loss should be comparable to the time it takes to train a full-epoch for a given network (which is $\approx 1/100^{th}$ of training time). Indeed, one can find that this suggestion is empirically observed in many previous studies (*23, 58, 59*)

For example, consider training of the largest and most complicated neural network ever, GPT-3 (*13*). The 175B parameter version of the GPT-3 takes 3.10^{23} FLOPs to train. Solving for number of passes (number of operations = number of parameters × weight reuse × 3 × number of passes), one needs 3×10^{23} / ($175.10^9 \times 4 \times 3$) = 1.4×10^{11} passes. Note that although the GPT-3 is immensely complex, its low weight reuse is certainly advantageous for analog deep learning applications. For an architecture intending to train GPT-3, the network sequencing would surely be pipelined, so we can assume that at any given time, at least 60% of the arrays are active (i.e. utilization factor = 60%), while the signal integration time is set at 100 ns (*19*) to have fast operation while maintaining a decent signal to noise ratio. So effective time for an input to pass the network would be 100 x 3 / 0.6 = 500 ns. Therefore, the total training time would be $1.4\times10^{11} \times 500$ ns $=$ 19.4 hours. Assuming training takes a relatively standard 100 epochs, a single epoch time should be 700 s. In conclusion, we suggest retention time-constants (not to be confused with single-state worth change) around 100-1000 s should be sufficient for many critical applications.

6. Additional Discussions for Endurance Under High Electric Fields

As explained in the main body, the electrical voltages and fields we apply to the device stack are high, beyond the breakdown field or the electrochemical stability window of traditional electrolyte materials (*45*), even the enlarged stability window of protonic electrolytes of ~3V (*55, 60*). Therefore, the fact the device operates stably and reversibly is unexpected and shows how our device requires a different understanding compared to solid-state battery systems. Our explanation is that because the PSG operates at 10V/-8V only for nanosecond transients, it does not have to conform to the thermodynamic stability requirement for electrolytes, which requires quasi-static (or nearly infinite time) waits. Indeed, we have seen our devices do "exsplode" if we stay at 10V/-8V for longer than a critical time. Before such critical time, however, our PSG does not sustain permanent damage. Therefore, these dynamics are fundamentally different from thermally activated hopping transport in batteries. Furthermore, the proton is the only ion among all cations and anions in solid-state chemistry to have a zero (actually, slightly negative) Shannon ionic radius of −0.04Å, in contrast to the 0.90Å of Li^+, thus generating very little strain/stress as it enters or exits solids. We believe these are two key aspects of the physics underlying the low-damage "protonic breakdown", a transient phenomenon that we took advantage of in the 10V/-8V operations.

7. Assessing the Modulation Variation Effects for Analog Deep Learning Applications

The experimental data shown in **Fig. 1**C was replotted above in **Fig.S12**A to better display the modulation variation of the devices. Moreover, this behavior can be simply modeled by assuming a conductance modulation $\boldsymbol{\Delta G^{actual} = \Delta G(1 + \sigma * randn)}$, where $\boldsymbol{\Delta G}$ is the average change in conductance per step (linear assumption), $\boldsymbol{randn}$ is a random variable with normal distribution,

and $\boldsymbol{\sigma}$ is the magnitude of variability. In our case, the experimental data is best represented with a $\boldsymbol{\sigma}$ value of 0.8 (**Fig.S12**A).

Importantly, this level of modulation variability is not an issue for analog deep learning applications. In a training scenario, the programming of these devices is performed in small steps and in a highly iterative manner under the feedback control of an optimization algorithm. Therefore, even if a certain update is not accurate, the following inputs would generate a correction which would ultimately allow successful convergence as long as the modulation variations are random. Moreover, even if the device modulation had $\boldsymbol{\sigma} = \mathbf{0}$, the updates would still have variability due to the dataset. In other words, none of the individual updates would show the direction of the true gradient but their average across the full dataset would. For example, assume a single-parameter regression example aiming to find a dataset mean $\boldsymbol{\mu}$ using an analog device. At a given time when $\boldsymbol{G} > \boldsymbol{\mu}$, an input $\boldsymbol{x}$ can still be larger than $\boldsymbol{G}$ and cause an incremental update, causing $\boldsymbol{G}$ to change away further away from the optimum ($\boldsymbol{\mu}$). Nonetheless, training over the complete dataset would change $\boldsymbol{G}$ towards $\boldsymbol{\mu}$ and enable successful convergence (which is observed experimentally in **Fig.S9**).

To further support this understanding, training simulations were performed using device models extracted from experimental data. First, **Fig.S12**B shows a simulated version of the single parameter regression (similar to the experimental demonstration in **Fig.S9**). As expected, the learning curves show no significant difference from one another, even for a device model with three times the experimental variation (red curves). Moreover, **Fig.S12**C shows a similar experiment on a more complex setting, where a fully-connected neural network was trained on MNIST dataset. The details of the simulator and the network can be found in Ref. 19. These simulations once again evidenced that the modulation variation at these levels is not an issue for analog deep learning.

References and Notes

1. V. K. Sangwan, M. C. Hersam, Neuromorphic nanoelectronic materials. *Nat. Nanotechnol.* **15**, 517–528 (2020). doi:10.1038/s41565-020-0647-z Medline

2. Q. Xia, J. J. Yang, Memristive crossbar arrays for brain-inspired computing. *Nat. Mater.* **18**, 309–323 (2019). doi:10.1038/s41563-019-0291-x Medline

3. M. A. Zidan, J. P. Strachan, W. D. Lu, The future of electronics based on memristive systems. *Nat. Electron.* **1**, 22–29 (2018). doi:10.1038/s41928-017-0006-8

4. B. Roling, L. N. Patro, O. Burghaus, M. Gräf, Nonlinear ion transport in liquid and solid electrolytes. *Eur. Phys. J. Spec. Top.* **226**, 3095–3112 (2017). doi:10.1140/epjst/e2017-70073-9

5. A. Röthel, S. Friedrich, R. Lühning, L. Heuer, Theoretical description of ion conduction in disordered systems: From linear to nonlinear response. *Z. Phys. Chem.* **224**, 1855–1889 (2010). doi:10.1524/zpch.2010.0020

6. L. Onsager, The motion of ions: Principles and concepts. *Science* **166**, 1359–1364 (1969). doi:10.1126/science.166.3911.1359 Medline

7. M. J. Dignam, Ion transport in solids under conditions which include large electric fields. *J. Phys. Chem. Solids* **29**, 249–260 (1968). doi:10.1016/0022-3697(68)90069-3

8. T. Erdey-Grúz, M. Volmer, Zur theorie der wasserstoff überspannung. *Z. Phys. Chem.* **150A**, 203–213 (1930). doi:10.1515/zpch-1930-15020

9. D. Ielmini, Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Dev.* **58**, 4309–4317 (2011). doi:10.1109/TED.2011.2167513

10. S. Menzel, U. Böttger, M. Wimmer, M. Salinga, Physics of the switching kinetics in resistive memories. *Adv. Funct. Mater.* **25**, 6306–6325 (2015). doi:10.1002/adfm.201500825

11. F. Palumbo, C. Wen, S. Lombardo, S. Pazos, F. Aguirre, M. Eizenberg, F. Hui, M. Lanza, A review on dielectric breakdown in thin dielectrics: Silicon dioxide, high-k, and layered dielectrics. *Adv. Funct. Mater.* **30**, 1900657 (2020). doi:10.1002/adfm.201900657

12. W. Fedus, B. Zoph, N. Shazeer, Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. arXiv:2101.03961 [cs.LG] (2021).

13. T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, "Language models are few-shot learners" in *Advances in Neural Information Processing Systems 33* (*NeurIPS2020*), H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, H. Lin, Eds. (Curran Associates, 2020), pp. 1877–1901.

14. E. Strubell, A. Ganesh, A. McCallum, "Energy and policy considerations for deep learning in NLP" in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, L. Màrquez, Eds. (Association for Computational Linguistics, 2019), pp. 3645–3650.

15. I. L. Markov, Limits on fundamental limits to computation. *Nature* **512**, 147–154 (2014). doi:10.1038/nature13570 Medline

16. J. Choi, S. Venkataramani, V. Srinivasan, K. Gopalakrishnan, Z. Wang, P. Chuang, "Accurate and efficient 2-bit quantized neural networks" in *Proceedings of the 2nd SysML Conference* (2019); https://mlsys.org/Conferences/2019/doc/2019/168.pdf.

17. X. Sun, J. Choi, C.-Y. Chen, N. Wang, S. Venkataramani, V. Srinivasan, X. Cui, W. Zhang, K. Gopalakrishnan, "Hybrid 8-bit floating point (HFP8) training and inference for deep neural networks" in *Advances in Neural Information Processing Systems 32* (*NeurIPS2019*), H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, R. Garnett, Eds. (Curran Associates, 2019); https://proceedings.neurips.cc/paper/2019/file/65fc9fb4897a89789352e211ca2d398f-Paper.pdf.

18. K. Steinbuch, Die Lernmatrix. *Kybernetik* **1**, 36–45 (1961). doi:10.1007/BF00293853

19. T. Gokmen, Y. Vlasov, Acceleration of deep neural network training with resistive cross-point devices: Design considerations. *Front. Neurosci.* **10**, 333 (2016). doi:10.3389/fnins.2016.00333 Medline

20. T. Gokmen, M. Onen, W. Haensch, Training deep convolutional neural networks with resistive cross-point devices. *Front. Neurosci.* **11**, 538 (2017). doi:10.3389/fnins.2017.00538 Medline

21. S. Agarwal, S. J. Plimpton, D. R. Hughart, A. H. Hsia, I. Richter, J. A. Cox, C. D. James, M. J. Marinella, "Resistive memory device requirements for a neural algorithm accelerator" in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*(IEEE, 2016), pp. 929–938;doi:10.1109/IJCNN.2016.7727298.

22. G. W. Burr, R. M. Shelby, S. Sidler, C. Di Nolfo, J. Jang, I. Boybat, R. S. Shenoy, P. Narayanan, K. Virwani, E. U. Giacometti, B. N. Kurdi, H. Hwang, Experimental demonstration and tolerancing of a large-dcale neural network (165 000 synapses) using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Dev.* **62**, 3498–3507 (2015). doi:10.1109/TED.2015.2439635

23. S. Ambrogio, P. Narayanan, H. Tsai, R. M. Shelby, I. Boybat, C. di Nolfo, S. Sidler, M. Giordano, M. Bodini, N. C. P. Farinha, B. Killeen, C. Cheng, Y. Jaoudi, G. W. Burr, Equivalent-accuracy accelerated neural-network training using analogue memory. *Nature* **558**, 60–67 (2018). doi:10.1038/s41586-018-0180-5 Medline

24. H. Jiang, L. Han, P. Lin, Z. Wang, M. H. Jang, Q. Wu, M. Barnell, J. J. Yang, H. L. Xin, Q. Xia, Sub-10 nm Ta channel responsible for superior performance of a HfO_2 memristor. *Sci. Rep.* **6**, 28525 (2016). doi:10.1038/srep28525 Medline

25. S. Kim, T. Todorov, M. Onen, T. Gokmen, D. Bishop, P. Solomon, K. Lee, M. Copel, D. B. Farmer, J. A. Ott, T. Ando, H. Miyazoe, V. Narayanan, J. Rozen, "Metal-oxide based, CMOS-compatible ECRAM for Deep Learning Accelerator" in *2019 IEEE International Electron Devices Meeting*, (IEEE, 2019), pp. 847–850.

26. Y. Li, E. J. Fuller, J. D. Sugar, S. Yoo, D. S. Ashby, C. H. Bennett, R. D. Horton, M. S. Bartsch, M. J. Marinella, W. D. Lu, A. A. Talin, Filament-free bulk resistive memory

enables deterministic analogue switching. *Adv. Mater.* **32**, e2003984 (2020). doi:10.1002/adma.202003984 Medline

27. Y. van de Burgt, E. Lubberman, E. J. Fuller, S. T. Keene, G. C. Faria, S. Agarwal, M. J. Marinella, A. Alec Talin, A. Salleo, A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nat. Mater.* **16**, 414–418 (2017). doi:10.1038/nmat4856 Medline

28. E. J. Fuller, F. El Gabaly, F. Léonard, S. Agarwal, S. J. Plimpton, R. B. Jacobs-Gedrim, C. D. James, M. J. Marinella, A. A. Talin, Li-ion synaptic transistor for low power analog computing. *Adv. Mater.* **29**, 1604310 (2017). doi:10.1002/adma.201604310 Medline

29. A. Chanthbouala, V. Garcia, R. O. Cherifi, K. Bouzehouane, S. Fusil, X. Moya, S. Xavier, H. Yamada, C. Deranlot, N. D. Mathur, M. Bibes, A. Barthélémy, J. Grollier, A ferroelectric memristor. *Nat. Mater.* **11**, 860–864 (2012). doi:10.1038/nmat3415 Medline

30. A. Krogh, J. A. Hertz, "A simple weight decay can improve generalizaition" in *Advances in Neural Information Processing Systems 4* (*NIPS 1991*), J. Moody, S. Hanson, R.P. Lippmann, Eds. (Morgan-Kaufmann, 1991), pp. 950–957.

31. D. B. Strukov, R. S. Williams, Exponential ionic drift: Fast switching and low volatility of thin-film memristors. *Appl. Phys. A* **94**, 515–519 (2009). doi:10.1007/s00339-008-4975-3

32. X. Yao, K. Klyukin, W. Lu, M. Onen, S. Ryu, D. Kim, N. Emond, I. Waluyo, A. Hunt, J. A. Del Alamo, J. Li, B. Yildiz, Protonic solid-state electrochemical synapse for physical neural networks. *Nat. Commun.* **11**, 3134 (2020). doi:10.1038/s41467-020-16866-6 Medline

33. M. Onen, N. Emond, J. Li, B. Yildiz, J. A. del Alamo, CMOS-compatible protonic programmable resistor based on phosphosilicate glass electrolyte for analog deep learning. *Nano Lett.* **21**, 6111–6116 (2021). doi:10.1021/acs.nanolett.1c01614 Medline

34. J. Tang, D. Bishop, S. Kim, M. Copel, T. Gokmen, T. Todorov, S. Shin, K. T. Lee, P. Solomon, K. Chan, W. Haensch, J. Rozen, "ECRAM as scalable synaptic cell for high-speed, low-power neuromorphic computing" in *2018 IEEE International Electron Devices Meeting* (IEEE, 2018); doi:10.1109/IEDM.2018.8614551.

35. A. Hjelm, C. G. Granqvist, J. M. Wills, Electronic structure and optical properties of WO_3, $LiWO_3$, $NaWO_3$, and HWO_3. *Phys. Rev. B* **54**, 2436–2445 (1996). doi:10.1103/PhysRevB.54.2436 Medline

36. S. Prakash, W. E. Mustain, S. H. Park, P. A. Kohl, Phosphorus-doped glass proton exchange membranes for low temperature direct methanol fuel cells. *J. Power Sources* **175**, 91–97 (2008). doi:10.1016/j.jpowsour.2007.09.060

37. M. Nogami, Y. Tarutani, Y. Daiko, S. Izuhara, T. Nakao, T. Kasuga, Preparation of P_2O_5-SiO_2 glasses with proton conductivity of ~100 mS/cm at room temperature. *J. Electrochem. Soc.* **151**, A2095 (2004). doi:10.1149/1.1817766

38. D. Bhusari, J. Li, P. J. Jayachandran, C. Moore, P. A. Kohl, Development of P-doped SiO_2 as proton exchange membrane for microfuel cells. *Electrochem. Solid-State Lett.* **8**, A588 (2005). doi:10.1149/1.2050587

39. Y. Li, Y. T. Cheng, Hydrogen diffusion and solubility in palladium thin films. *Int. J. Hydrogen Energy* **21**, 281–291 (1996). doi:10.1016/0360-3199(95)00094-1

40. R. Oesten, R. A. Huggins, Proton conduction in oxides: A review. *Ionics* **1**, 427–437 (1995). doi:10.1007/BF02375287

41. R. A. B. Devine, G. V. Herrera, Electric-field-induced transport of protons in amorphous SiO_2. *Phys. Rev. B* **63**, 233406 (2001). doi:10.1103/PhysRevB.63.233406

42. J. Godet, A. Pasquarello, Proton diffusion mechanism in amorphous SiO_2. *Phys. Rev. Lett.* **97**, 155901 (2006). doi:10.1103/PhysRevLett.97.155901 Medline

43. M. E. S. Beatty, E. I. Gillette, A. T. Haley, D. V. Esposito, Controlling the relative fluxes of protons and oxygen to electrocatalytic buried interfaces with tunable silicon oxide overlayers. *ACS Appl. Energy Mater.* **3**, 12338–12350 (2020). doi:10.1021/acsaem.0c02359

44. M. Kunow, A. Heuer, Nonlinear ionic conductivity of lithium silicate glass studied via molecular dynamics simulations. *J. Chem. Phys.* **124**, 214703 (2006). doi:10.1063/1.2198197 Medline

45. W. D. Richards, L. J. Miara, Y. Wang, J. C. Kim, G. Ceder, Interface stability in solid-state batteries. *Chem. Mater.* **28**, 266–273 (2016). doi:10.1021/acs.chemmater.5b04082

46. E. H. Snow, B. E. Deal, Polarization phenomena and other properties of phosphosilicate glass films on silicon. *J. Electrochem. Soc.* **113**, 263 (1966). doi:10.1149/1.2423929

47. D. Serghi, C. Pavelescu, D.c. dielectric breakdown in phosphosilicate glass films prepared by low temperature chemical vapour deposition. *Thin Solid Films* **186**, L25–L28 (1990). doi:10.1016/0040-6090(90)90518-I

48. R. Vuilleumier, D. Borgis, Quantum dynamics of an excess proton in water using an extended empirical valence-bond Hamiltonian. *J. Phys. Chem. B* **102**, 4261–4264 (1998). doi:10.1021/jp9807423

49. D. Ielmini, Resistive switching memories based on metal oxides: Mechanisms, reliability and scaling. *Semicond. Sci. Technol.* **31**, 063002 (2016). doi:10.1088/0268-1242/31/6/063002

50. M. von Witzleben, K. Fleck, C. Funck, B. Baumkötter, M. Zuric, A. Idt, T. Breuer, R. Waser, U. Böttger, S. Menzel, Investigation of the impact of high temperatures on the switching kinetics of redox-based resistive switching cells using a high-speed nanoheater. *Adv. Electron. Mater.* **3**, 1700294 (2017). doi:10.1002/aelm.201700294

51. F. Lewis, The hydrides of palladium and palladium alloys. *Platin. Met. Rev.* **4**, 132–137 (1960).

52. P. G. Dickens, S. Crouch-Baker, M. T. Weller, Hydrogen insertion in oxides. *Solid State Ion.* **18–19**, 89–97 (1986). doi:10.1016/0167-2738(86)90092-5

53. S. Burkhardt, M. T. Elm, B. Lani-Wayda, P. J. Klar, In situ monitoring of lateral hydrogen diffusion in amorphous and polycrystalline WO_3 thin films. *Adv. Mater. Interfaces* **5**, 1701587 (2018). doi:10.1002/admi.201701587

54. S. Nakabayashi, R. Shinozaki, Y. Senda, H. Y. Yoshikawa, Hydrogen nanobubble at normal hydrogen electrode. *J. Phys. Condens. Matter* **25**, 184008 (2013). doi:10.1088/0953-8984/25/18/184008 Medline

55. S. Wang, H. Jiang, Y. Dong, D. Clarkson, H. Zhu, C. M. Settens, Y. Ren, T. Nguyen, F. Han, W. Fan, S. Y. Kim, J. Zhang, W. Xue, S. K. Sandstrom, G. Xu, E. Tekoglu, M. Li, S. Deng, Q. Liu, S. G. Greenbaum, X. Ji, T. Gao, J. Li, Acid-in-clay electrolyte for wide-temperature-range and long-cycle proton batteries. *Adv. Mater.* **34**, 2202063 (2022). doi:10.1002/adma.202202063 Medline

56. M. Onen, N. Emond, B. Wang, D. Zhang, F. M. Ross, J. Li, B. Yildiz, J. A. del Alamo, Data for Nanosecond Protonic Programmable Resistors for Analog Deep Learning, Zenodo (2022); doi.org/10.5281/zenodo.6506857.

57. E. Cazzanelli, C. Vinegoni, G. Mariotto, A. Kuzmin, J. Purans, Raman study of the phase transitions sequence in pure WO3 at high temperature and in H_xWO_3 with variable hydrogen content. *Solid State Ion.* **123**, 67–74 (1999). doi:10.1016/S0167-2738(99)00101-0

58. Y. Li, S. Kim, X. Sun, P. Solomon, T. Gokmen, H. Tsai, S. Koswatta, Z. Ren, R. Mo, C. Yeh, W. Haensch, E. Leobandung, "Capacitor-based cross-point array for analog neural network with record symmetry and linearity" in *2018 IEEE Symposium on VLSI Technology* (IEEE, 2018); doi:10.1109/VLSIT.2018.8510648.

59. Y. Kohda, Y. Li, K. Hosokawa, S. Kim, R. Khaddam-Aljameh, Z. Ren, P. Solomon, T. Gokmen, S. Rajalingam, C. Baks, W. Haensch, E. Leobandung, "Unassisted true analog neural network training chip" in *2020 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2020); doi:10.1109/IEDM13553.2020.9371973.

60. L. Suo, O. Borodin, T. Gao, M. Olguin, J. Ho, X. Fan, C. Luo, C. Wang, K. Xu, "Water-in-salt" electrolyte enables high-voltage aqueous lithium-ion chemistries. *Science* **350**, 938–943 (2015). doi:10.1126/science.aab1595 Medline