APPLIED SCIENCES AND ENGINEERING

Palimpsest memories stored in memristive synapses

Christos Giotis[1]*, Alexander Serb[1,2], Vasileios Manouras[1], Spyros Stathopoulos[1], Themis Prodromakis[1,2]

Biological synapses store multiple memories on top of each other in a palimpsest fashion and at different time scales. Palimpsest consolidation is facilitated by the interaction of hidden biochemical processes governing synaptic efficacy during varying lifetimes. This arrangement allows idle memories to be temporarily overwritten without being forgotten, while previously unseen memories are used in the short term. While embedded artificial intelligence can greatly benefit from this functionality, a practical demonstration in hardware is missing. Here, we show how the intrinsic properties of metal-oxide volatile memristors emulate the processes supporting biological palimpsest consolidation. Our memristive synapses exhibit an expanded doubled capacity and protect a consolidated memory while up to hundreds of uncorrelated short-term memories temporarily overwrite it, without requiring specialized instructions. We further demonstrate this technology in the context of visual working memory. This showcases how emerging memory technologies can efficiently expand the capabilities of artificial intelligence hardware toward more generalized learning memories.

INTRODUCTION

While neural networks in the cerebral cortex use an estimated 10^{13} to 10^{14} synapses to facilitate a plethora of cognitive abilities (*1*, *2*), their engineered counterparts require equivalent numbers of trainable parameters for a far narrower application spectrum (*3*, *4*). One candidate for explaining this discrepancy in learning capacity between biological and artificial intelligence (AI) suggests that synapses are able to consolidate multiple memories that can be revealed at different time scales—much like a palimpsest (*5*). Synapses can remember long-term plasticity events, namely, long-term potentiation (LTP) and long-term depression (LTD), while expressing altered states in the short term (*6*). This temporal partition enables the brain to use the same resources for multiple computation processes. The adoption of this flexibility by neuromorphic hardware is therefore a critical milestone toward the integration of AI in a wider range of on-the-edge, continuously-on learning systems.

Palimpsest storage is realized biologically via the bidirectional interaction of hidden biochemical processes affecting the manifestation of synaptic efficacy at different time scales (*5*) after each memory modification. These processes are characterized by their own degrees of plasticity (i.e., learning rates) and lifetimes (i.e., "forgetting time constants"). Sparsely presented memories induce fast changes in synaptic efficacy, but these quickly decay to reveal older but more persistent memories that have successfully affected less plastic but more long-lasting processes. The coexistence of these processes allows synapses to be both plastic in the short term, enabling incoming memories to be written easily, and rigid in the long term, thus preserving old memories of validated significance.

The flexibility promised by dynamic memory consolidation has naturally attracted the attention of AI hardware design and, particularly, that of memristive technologies, which have already showcased their potential in numerous neuromorphic applications (*7*–*12*). Memristor-based artificial synapses have demonstrated core plasticity functionality in the form of LTP/LTD. These implementations show how plasticity changes can become more pronounced in an analog regime when stimulation events are applied successively. These synaptic designs are largely based on phase-change memory (PCM) materials, which experience conductance changes when stimulation pulses are applied on them to emulate potentiation and depression. These designs achieve synaptic emulation, both by using standalone memristors (*12*–*14*) or by integrating them in more complex circuitry (*15*, *16*). While these studies have demonstrated the abilities of memristors to facilitate learning in artificial neural networks (ANNs), they have not considered how learned memories can be protected from continuous synaptic modifications—a crucial requirement for efficient online learning.

Both PCM- and resistive random-access memory (RRAM)–based memristors have been used to implement metaplasticity, i.e., tuning of the learning rate (*17*, *18*), on complementary metal-oxide semiconductor–based artificial synapses in spiking neural networks (*16*, *19*). Metaplasticity has been studied extensively because of its potential for protecting consolidated memories via tunable learning rates. In a similar vein, nonvolatile RRAM synapses use explicitly modulated bias voltage to tune their switching (i.e., learning) rate (*20*–*24*). However, these studies have not been evaluated in the context of dynamic memory consolidation for two reasons. First, their implementation of variable learning rates occurs from appropriately tuned stimulation variation, implying that the need for plasticity rate changes is known a priori. Thus, they cannot operate in an online learning environment where the need for consolidation is usually unknown in real time. In addition, although these synaptic models showcase both LTP and LTD, they focus only on manipulating learning rates unidirectionally. This means that plasticity rates vary only within the context of stronger or lesser potentiation/depression independently, for instance, an already potentiated synapse experiencing lower plasticity rates toward further potentiation. Nevertheless, for metaplasticity to function properly, it is also imperative for a synapse to mitigate for catastrophic forgetting and protect its learned state against modifications in either direction concurrently (*18*).

The protection of memory states has also been studied in the context of passive memory lifetime. Volatile RRAM has demonstrated short-term memory (STM) to long-term memory (LTM) transition where repeated presentations of the same memory induce

[1]Department of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK. [2]Centre for Electronics Frontiers, School of Engineering, University of Edinburgh, Edinburgh EH9 3FB, UK.
*Corresponding author. Email: c.giotis@soton.ac.uk

longer changes in synaptic states, albeit being irreversible and unidirectional (*10*, *25*–*27*). This means that they have only worked in the context of LTP, where successive potentiation leads to memory states that are available for longer time windows. While this serves as a strong foundation toward using the time dynamics of volatile memristors, these implementations also ignore the protection of consolidated states when opposing synaptic modifications occur, and hence, the issue of catastrophic forgetting remains unresolved. Last, simulated ANNs based on metaplasticity principles have demonstrated an increase in specialized learning capacity (*28*). While the authors comment that palimpsest capabilities can expand capacity toward uncorrelated memories, now, neither commercially available nor emerging memristive technologies have exhibited the necessary properties required for dynamic memory consolidation.

In this work, we built upon the studies that are mentioned above to bridge synaptic plasticity with automatic consolidation and memory protection, irrespective of plasticity direction. The characteristics of RRAM volatility (*29*, *30*) are exploited to emulate the function of the hidden biochemical processes that enable palimpsest consolidation. We harness the bidirectional volatile and nonvolatile responses of RRAM devices to practically realize two consolidation time scales in one device, effectively storing competing binary states in a single synapse with doubled STM and LTM capacity. Then, we upscale this principle to consolidate memory traces at variable consolidation intensity. Our technology can protect a strong memory in its long-term storage while allowing multiple short-term signals to take over the STM fleetingly and then quickly decay. Simple metaplasticity is also realized as a natural subset behavior when a given memory is consolidated consistently. Last, we show how this palimpsest memory can operate in a visual system where it also boasts unsupervised denoising abilities. Our memory system is unique in a number of key attributes. First, its expanded capacity is independent of the correlation between memories presented to it and even performs under fully destructive interference. Moreover, it can automatically sacrifice palimpsest capabilities for even stronger memory protection. These features unfold naturally as a result of single memristive device properties and do not require special biasing regimes or otherwise increased operational complexity.

RESULTS

Volatile memristors as candidates for palimpsest consolidation

Palimpsest consolidation relies on the premise that hidden variables (in the case of biology, resulting from complex biochemical processes) induce changes in synaptic efficacy acting across different time scales (*31*). While the operation of these processes is yet to be fully mapped, their phenotypic response can be modeled by the characteristics of fluid diffusion to first approximation (*5*). We can visualize a single synapse as an interconnected chain of progressively larger beakers, where the first beaker alone determines the synaptic weight and every subsequent beaker represents a hidden variable. The discrepancies in liquid levels across beakers affect the evolution of liquid distribution over time throughout the entire chain. This is illustrated in Fig. 1D. While stimulation of the model synapse (adding or removing liquid) occurs exclusively on that first beaker, repeated homopolar stimulation does eventually propagate to the hidden and crucially larger connected beakers farther down the chain. This is the phenomenon of consolidation. However, in a palimpsest memory, competing signals may still successfully write an opposing synaptic state with relatively little effort, albeit in the absence of further reinforcement, the consolidated liquid in the hidden beaker

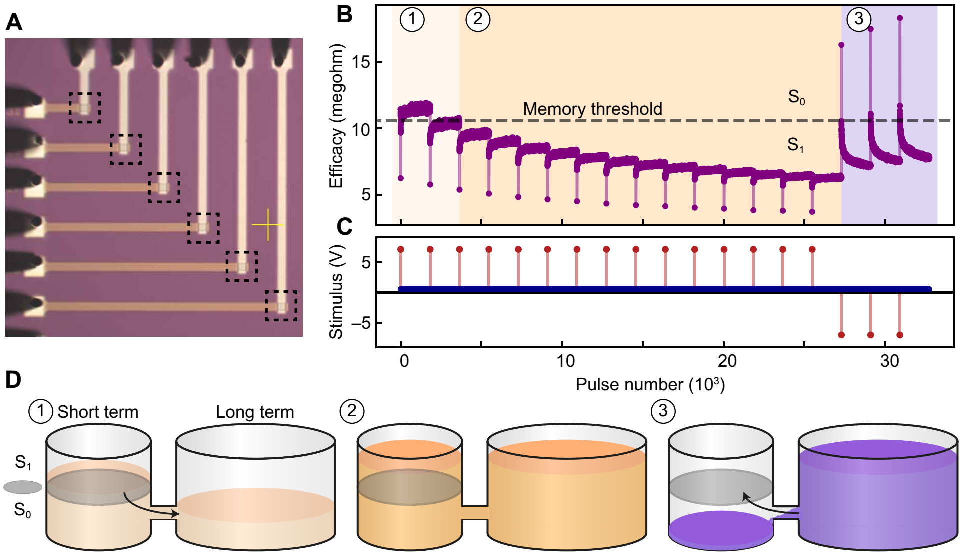

Fig. 1. Demonstration of a memristive synapse. (**A**) Microscope images of fabricated RRAM samples. Individual memristive synapses are highlighted (see black dashed squares) in the intersections of their corresponding conductive electrodes. (**B**) Schematic operation of a volatile memristor. Plasticity events cause a pronounced and rapid change in synaptic efficacy observed on a fast time scale (STM) and a smaller but more stable change on a slow time scale (LTM). The change in synaptic efficacy over time is shown in (B) following the memory pattern applied in (**C**). (**D**) Schematic equivalence of the three consolidation stages in (B) with the analogous beaker theory. The temporal evolution of the first beaker's liquid level is determined by the hidden state of the second beaker. The liquid levels of the first beaker at stages 1 and 3 are deliberately placed at disproportionate distances from the binary threshold to reflect the asymmetric response of the particular memristive synapse to potentiation and depression events.

will eventually revert the synapse to expressing the previously consolidated memory.

In this work, we have used TiO_2-based volatile memristive devices (see intersections in Fig. 1A and Materials and Methods) to examine the potential of palimpsest consolidation. Volatility is defined as the change in a device's state within a specified time window, and any change in its resistance R that outlasts this window is considered a nonvolatile residue (*29*). Volatile changes are easily induced but equally short lived (analogous to manipulating the visible synaptic state). Nonvolatile residues can accumulate in smaller steps and act as attractors for volatile decay $R(t)$ (*30*).

Our technology is a prime candidate for such conceptual demonstration due to the intrinsic characteristics that are observed within it. TiO_2 RRAM has already exhibited how it experiences polarity-dependent bidirectional volatility in a controllable manner (*29*). This characteristic is the essential driver behind our approach to palimpsest consolidation since it enables reversibility from STM to LTM states. Moreover, volatility in TiO_2 technologies can be induced reversibly for prolonged time periods without suffering from catastrophic state degradation. For more details, see fig. S1 where continuous device stimulation leads to stable volatility for up to 2500 short retention cycles and fig. S2 where volatility is continuously induced bidirectionally (for up to 8.5 hours) and under highly invasive stimulation (up to 1000 pulses per cycle) without hindering the reliability of the device. It should also be noted that while TiO_2 technology can support this work, it is by no means restrictive and standalone research on more technology candidates should be welcome independently, as argued later in Discussion.

Using these characteristics, we first demonstrate palimpsest memory on a single memristive synapse. Specifically, we aim to consolidate a binary state S_1 via LTP and then expressing the antipodal S_0 for a short duration. All plasticity events are uniformly distributed at 30 s intervals (see Materials and Methods for details). The results presented in Fig. 1 (B and C) illustrate three distinct consolidation stages. In stage 1, potentiation events push resistance R to below a predetermined binary threshold R_{thres}, corresponding to a short-lived expression of S_1, but the hidden nonvolatile state remains above R_{thres}, so in the absence of further reinforcement, the synapse reverts to the more consolidated S_0. In stage 2, additional potentiation events cause the synapse to undergo LTP. The hidden nonvolatile state is pushed below R_{thres} and S_1 is consolidated at the long-term time scale. Successive potentiation events produce diminishing nonvolatile residues, observed across stages 1 and 2, hinting toward the soft resistive state bounds observed in RRAM devices (*32*). These bounds are manifested via the diminishing changes in R after the corresponding observation time window has closed. Consequently, successive write events eventually fail to consolidate one state further, since they are not entrenching the analog synaptic state further from the binary state threshold. This is quantified in fig. S3, where the successive (%) nonvolatile residues are shown to diminish in magnitude for successive write events during long-term consolidation (also shown for all devices in hardware demonstration—see the next section). Bounded synaptic efficacy is to known aid the capacity of memory networks (*5*), and here, it results directly from device electrochemistry. Moreover, soft bounding naturally mitigates any asymmetry in device volatility since it allows for increased memory capacity even if potentiation and depression event strengths are not perfectly balanced (*8*, *33*). In stage 3, competing depression events cause antipodal plastic increases in R, manifesting as a temporary expression of S_0 in memory. However, because the hidden nonvolatile state has been consolidated below R_{thres}, the observed state overwrite is only realized short term, reverting to S_1 over time.

Overall, the artificial synapse is able to protect a hidden memory while, at the same time, reserving the flexibility to express another opposite memory atop it; memory capacity is thus doubled. Next, we note that frequently competing memory events inevitably drive the rigid state closer to R_{thres}. The synapse sacrifices stability at the slow time scale as a necessary trade-off for short-term plasticity. Last, despite our specific device family exhibiting asymmetric responses in potentiation and depression modifications, it is a good candidate for highlighting the consolidation applications due the high ratio between volatile and nonvolatile plasticity changes, a key functional parameter of the system.

Operation of memory system

Next, we construct a small memristive network composed of six synapses to consolidate two competing signals, M_1 ($\triangledown$ = 101100) and M_2 ($\blacktriangledown$ = 010011). These are antipodal binary vectors, and consequently, competition for storage consists a worst-case zero-sum game. The experimental setup is explained in Fig. 2A. Each bit b_n, $n \in \{0,1, ...,5\}$, of memories $M_{1,2}$ is written on a corresponding memristive synapse w_n (see Materials and Methods). Encoding palimpsest memories in this worst-case scenario implies that system performance would only increase in generalized applications, for instance, random uncorrelated binary vectors average a 50% similarity rate, leaving only the 50% subject to destructive interference effects.

Examining the analog resistance values gives further insight on how the two memory signals interact across the synapses. These synapses cycle through the same consolidation stages which have been outlined in Fig. 1, B to D and have been discussed in the previous section. Progressive modifications push the hidden state further away from R_{thres} as reflected from the "deeper" resistance values, at which point, M_1 is strongly consolidated at the long term. When M_2 is presented in memory, individual synapses are resistant to encoding the requested bit states. Specifically, synapses $w_{1,4,5}$ that have undergone LTD fail to fully encode the respective M_2 states in the two first write events. This stems from the mentioned asymmetry between the devices' volatile responses in opposite directions. However, all synapses are pushed closer to R_{thres} suggesting a retreat from strong M_1 consolidation. In the final M_2 event, the memory is fully written at the short term before M_1 is reinstated.

Memory performance is macroscopically examined in Fig. 2B. The overlap between M_1 and the system's state is shown against time. Because of the antipodal relationship between M_1 and M_2, an x% overlap between the system and the former implies a $(100 - x)$% overlap with the latter. The time axis has been truncated since everywhere between the second writing of M_1 at $t = 50$ s and the first presentation of M_2 at $t = 450$ s the overlap with M_1 is solidly 100%. This is also evident by examining Fig. 2A. The signal overlap is visibly quantized because of the small size of the synaptic circuit. Ultimately, each presentation of M_2 is progressively more successful at overwriting the consolidated M_1, whose recovery becomes progressively slower yet still achievable. Reinstation of M_1 is nonmonotonic. This is attributable to noise, which becomes a deciding factor when R is close to R_{thres}.

To further quantify how noise is expected to affect memory performance, we run the following experiment: First, using existing

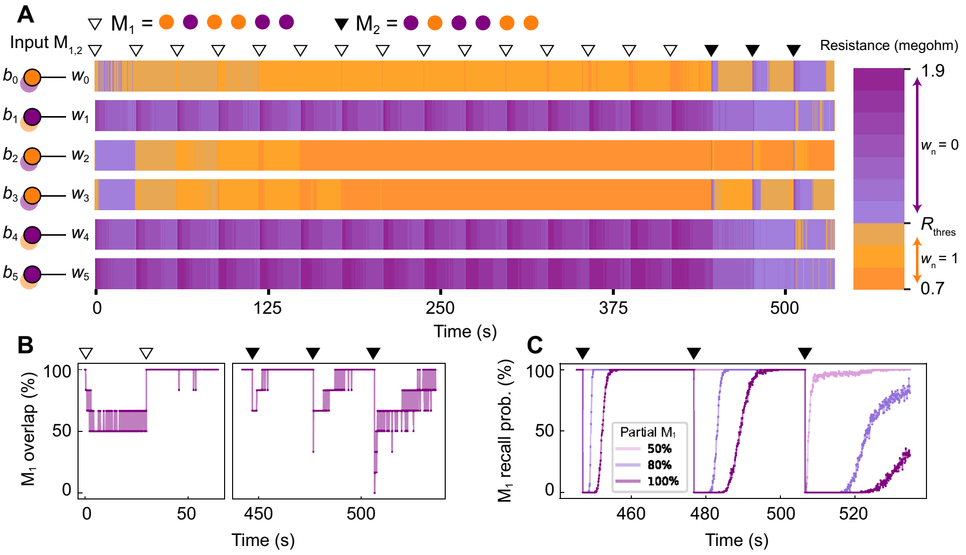

Fig. 2. Coexistence of STM and LTM in memristive synapses. (**A**) A set of six memristive synapses representing weights w_n, $n \in \{0,1, \ldots,5\}$ is fed plasticity instructions b_n as per patterns M1 and M2. Orange and purple stand for potentiation and depression, respectively. (**B**) Evolution of the memory trace overlap with M1 under the course of the 18 write events. One hundred percent overlap translates to perfect storage of M1, and since M1 and M2 are binary and mutually orthogonal, 0% overlap translates to perfect storage of M2. (**C**) Probability of recalling at least x% of M1, where $x \in \{50\%, 80\%, 100\%\}$.

volatility modeling methods (*30*), we obtain an expected ideal estimate of behavior under the stimulation protocol of Fig. 2A. Next, we calculate the noise distribution observed in real device data. Last, we run 200 simulations, each contaminated with different random noise, and evaluate the recall performance of M_1 after the presentations of M_2 at the end of each run (see Materials and Methods for details). Figure 2C shows the probability of recalling M_1 fully (100% signal overlap) or partially (at least 80 or 50% overlap) after applying the M_2 write events. As expected, the ability for recall of M_1 drops after each presentation of M_2, translating into longer recovery times and possibly lower achievable maximum recall values. Nevertheless, by the third M_2 presentation, the system continues to achieve at least 80% recall more than 80% of the time. This illustrates that the degradation of recall performance as competing memories progressively overwrite each other is smooth, which implies that effective performance will be affected by the required recall accuracy. Most AI systems already function on the basis of average estimations of learned signals, which implies a good quality partial recall suffices in practice (*3*).

The performance of the memory can be generalized even further by considering the palimpsest behavior within the easier context of random and uncorrelated memory traces. In these runs, emphasis has been shifted from using very large plastic capabilities that magnify the contrast between LTM and LTM time scales (see Figures 1 and 2) to ensure symmetrical performance under potentiation and depression modifications (see Materials and Methods). In practice, we have sacrificed the high ratio between volatile jump and nonvolatile residues for symmetrical, bidirectional volatile responses. Here, we assume a fully uniform memory network to isolate the system's dynamics from devices' operation variability. To that extent, the network is constructed in simulation using existing modeling methods for memristive volatility (see Materials and Methods).

A memory network composed of 100 synapses is subject to an ongoing stream of 500 input memories that are chosen randomly. Synaptic modifications occur evenly spaced every 10 s. In Fig. 3 (A to C), 10 random memories are written in the system before an LTM is consolidated with an intensity of $s = 2$ repetitions. A relatively low value for s has been chosen on purpose to prevent a very deep entrenchment of LTM in the rigid time scale. As observed in Fig. 3A, this prevents LTM from being fully written in the system. The memory overlap with three randomly chosen input memories (STMs) is also shown against the 50% noise floor. Spikes mark the presentation of these memories to the system. The general LTM overlap is surpassed by STM_{1-3} immediately after these are written in the system and before LTM is reinstated as the strongest memory. Some notable observations deserve mention here. First, LTM's failure to achieve a perfect signal overlap can be explained by the preceding plasticity events. Stochastic stimulation can cause spontaneous entrenchment of random synapses, which then challenges the consolidation of LTM. Second, the overall overlap of both LTM and successive STM signals falls over time, highlighting the slow but continuous degradation of memories as more recent inputs are received. Last, all STM signals retain an above-chance representation strength for a long time even if the interval during which they are the dominant memory is short. Hence, even at one-shot scenarios, the network exhibits high capacity in familiarity recalls, meaning it can distinguish whether multiple memories have been presented before or not.

Because acceptance thresholds for absolute signal overlaps are directly related to specific application needs, we focus on the relative strength difference between the consolidated LTM and incoming STMs. Figure 3B shows which memory is dominant (has the highest degree of correlation with the actual state of the memory network) in 10 s observation windows following each presentation of an STM. The LTM has been consolidated just before the commencement of the first STM trial. Blue regions indicate LTM dominance and red regions indicate STM dominance. For the first 100 random STMs, the 10 s observation window is dominated by

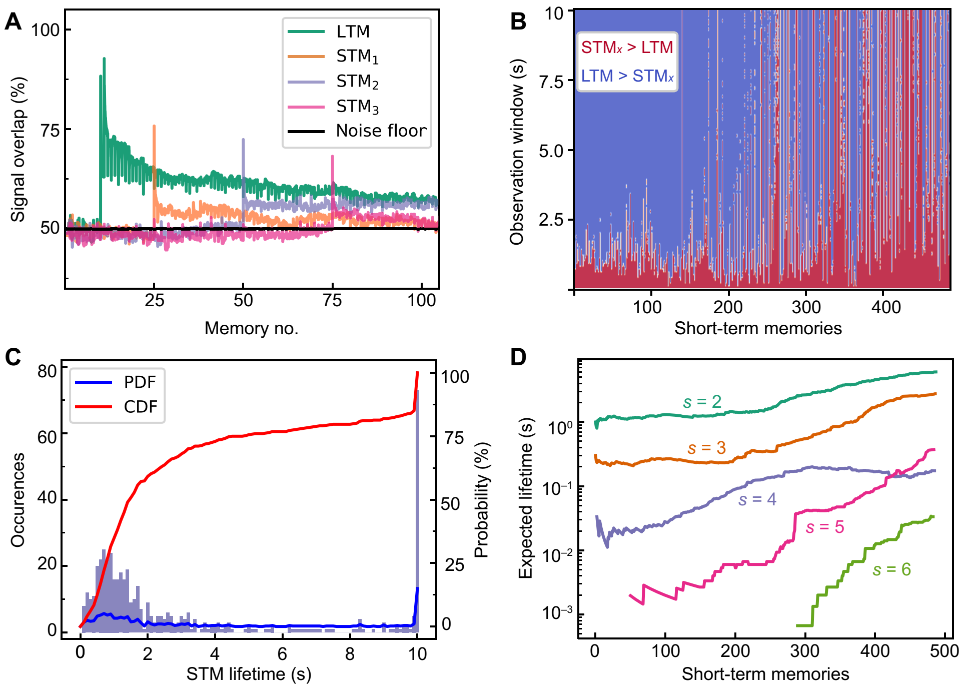

Fig. 3. Memory performance under continuous stimulation. (**A**) Evolution of a LTM and following STM signals. Each randomly chosen STM briefly surpasses LTM before the latter reinstates its dominance. Decaying signals retain higher than chance overlaps (they remain above the 50% noise floor). (**B**) Observation of all random STMs after LTM consolidation. Red regions indicate that the total time STMx signal is stronger than LTM; blue regions suggest the opposite. (**C**) Histogram of distribution of all STM total lifetimes on the left *y* axis. The corresponding probability density function (PDF) and cumulative distribution function (CDF) curves are illustrated via the right *y* axis. (**D**) Weighted expected STM lifetime for different LTM consolidation strengths "*s*." The green *s* = 2 line corresponds to the memory data presented in (A) to (C).

(quickly restored) LTM signals. However, as more patterns are presented to the memory, the consolidated LTM pattern degrades, and more recent signals prevail; red sections become longer and denser as LTM restoration collapses.

STM lifetime is defined as the total time period where some STM$_x$ signal dominates over the LTM (see Fig. 3C). The histogram of lifetime occurrences is shown on the left *y* axis, while the corresponding probability and cumulative distributions are shown on the right *y* axis. The data can be split into three main segments: First, a large bulk of STMs surviving between 0 and 2 s, mainly populated by STMs presented toward the beginning of the test and representing about 50% of the signals. Second, a relatively sparsely populated trough between ~2 and 10 s reflects the fact that, after ~2 s, the volatile component of the synaptic dynamics has for the most part relaxed. It should be noted that there is some preliminary evidence to suggest that volatility might function on phases, characterized by an initial phase of rapid decay, followed by a slower decay of the residue at a much smaller time constant (*29*). This subtle phenomenon, however, requires further study. Last, the peak at 10 s bins together any cases where the LTM would either be restored at more than 10 s or fail to be restored and thus appears as a prominent peak. This relationship is also reflected by the distribution's probability density function (PDF), as depicted by the blue line. Overall, while LTM remains consolidated in the rigid time scale, the network tends to rebound to it within that 2 s time frame. This is highlighted by the lifetime cumulative distribution function (CDF) shown in red. The probability that some STM$_x$ survives for up to 2 s is approximately 60%, while the probability that it never gets written is about 20%.

We also note that the system is able to retain LTMs more rigidly if they are more intensely entrenched/consolidated. The same experiment has been repeated for a range of LTM consolidation intensities $s = \{2,3,4,5,6\}$. In Fig. 3D, we show averaged STM lifetimes over a sliding window of 150 STM presentations for each consolidation intensity (data points before STM no. 150 only average over all previous STMs). As consolidation intensity increases, the ability to write any STM on top of LTM reduces significantly. Early lifetimes decrease by about one order of magnitude per *s* level until the first few tens (*s* = 5) or hundreds (*s* = 6) of STMs fail to surpass the LTM completely (expected lifetime = 0 s). This is because increased consolidation of LTM causes the rigid nonvolatile residues to shift further away from the (binary) efficacy threshold, reducing the efficiency of potentiation/depression events in changing the weight. Metaplastic properties are thus realized implicitly by our synapses. This shows how the trade-off between capacity and recollection accuracy can be controlled: High levels of consolidation appear to safeguard the LTM against at least hundreds of incoming STMs.

Visual working memory

Short-term attention

After evaluating performance with random memory streams, we examine the network's operation with statistically correlated signals. In particular, we draw inspiration from existing theory on working memory (*34*) to construct a vision network with short-term attention that is complementary to its memory capacity. To address this, we have set up a synaptic network composed of 100 × 100 simulated artificial synapses that observes incoming binary images (see Materials and Methods). We consolidate one image in the LTM and then sequentially store two images $ST_{1,2}$ in short-term capacity. In addition, we test the network's ability to consolidate only the relevant information within LTM. To do this, we present the network with three independently noise-contaminated LTM (NLTM) variations before writing ST_1 and ST_2. Thus, signal correlation is observed at two levels: All different images are at least 70% similar to each other, while the noisy NLTM variations are at least 80% similar. The timeline of synaptic modifications is shown in Fig. 4A. Time is set to $T = 0$ s at the occurrence of ST_1.

The image memories are shown in Fig. 4B. Each pixel corresponds to a memristive synapse. A random NLTM is illustrated. Snapshots of the network's evolution are shown in Fig. 4C. In this scenario, NLTMs are contaminated at a 10% noise level. The first snapshot (top left corner) shows the LTM trace just before ST_1 is written in the network. At $T = 0$ s, ST_1 successfully overtakes the network for a short time frame. Later snapshots show the network's transition to its long-term state. By time $T = 9$ s, the network has recovered LTM. The system's state at that point constructs a visibly cleaner representation of LTM than the final NLTM trace shown before ST_1 is presented. At $T = 10$ s, the third image, ST_2 is written in the memory network. Noticeably, the decay of ST_2 is first caused by the pixels that are common in LTM and ST_1. These have been consolidated by both memories and are thus reinstated faster. It should be noted that while part of the LTM signal is reinstated faster after the observation of ST_2, (namely, synapses that have previously been consolidated both by LTM and ST_1), this does not favor the overall recovery of LTM. As seen at Fig. 4D, the recovery rate of LTM is slower after ST_2, stemming from the signal's progressive contamination. This is yet another depiction of the gradual LTM degradation that is observed at Fig. 3 (A and B). Eventually, LTM becomes dominant again at time $T = 17$ s.

Unsupervised memory reconstruction

We are also interested in whether the system can identify statistical significance without supervision. By comparing the LTM trace in the top left corner and the last snapshot, we observe that the network is able to automatically denoise consolidated signals. This property

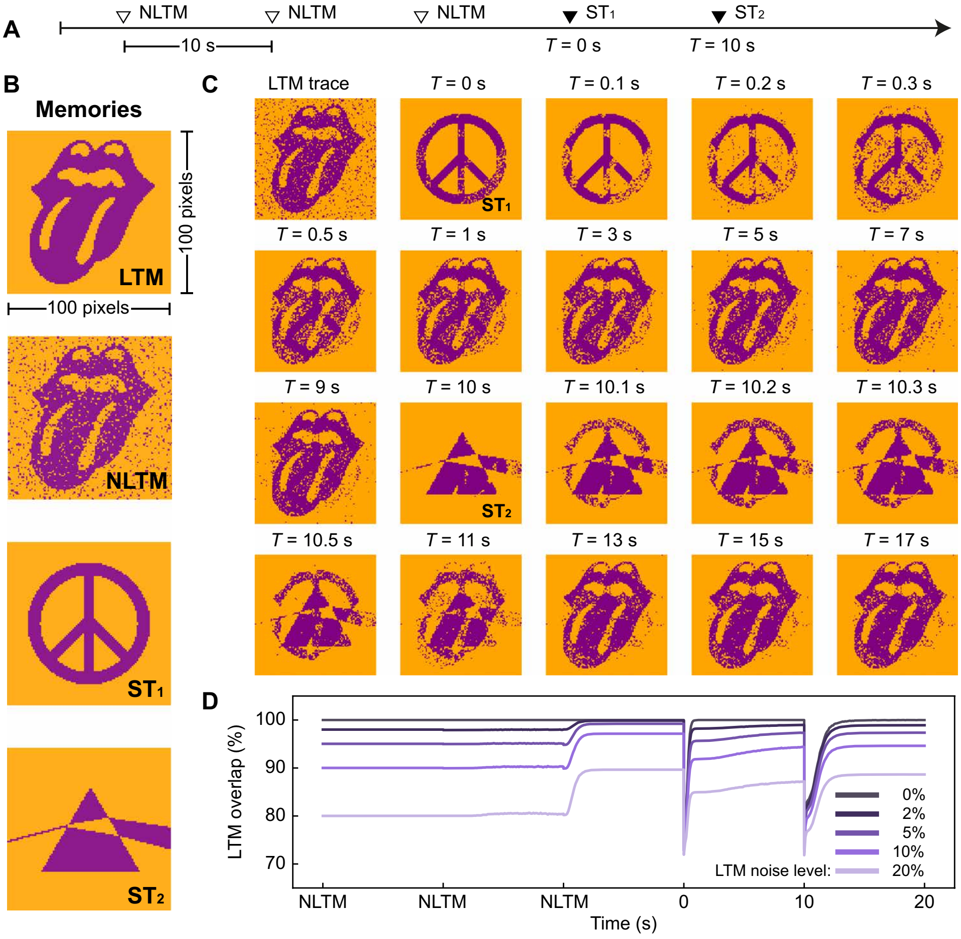

Fig. 4. Consolidation of binary images. (**A**) Operation timeline: Images are written in memory every 10 s. Random noisy representations of a LTM (NLTM and LTM, respectively) are written three times before two images are written, once each, in the short term ($ST_{1,2}$). (**B**) Original signals of image memories. An NLTM depiction is chosen at random. (**C**) Temporal snapshots of memory states. The initial snapshot corresponds to last observation before ST_1 is seen. Time is referenced after that instant. The moments when the two short-term signals are written were noted. In both cases, a noisy trace of each short-term signal can be accessed in the memory immediately after observation. Then, the short-term signal gradually degrades, while LTM is reinstated in memory. (**D**) Outline of the LTM signal overlap, as it evolves following the timeline in (A). The initial overlap reflects the corresponding noise level. The experiment is repeated for different NLTM noise levels. Even if a noiseless version of LTM is never presented to the system per se, the memory that achieves above the baseline overlaps vis-à-vis LTM after consolidation has finished. This high overlap is retained even after the ST_1 and ST_2 are seen by the network.

arises naturally from the fact that random information, presented only sparsely, is less susceptible to consolidation than bits written with higher intensity, as shown in Fig. 3. Hence, the network is able to average over many noisy representations and converge closer to the actual, but never explicitly presented, LTM signal. These denoising properties are clearly visible in Fig. 4D, where the experiment has been repeated for several noise levels on LTM and average signal overlaps are shown versus time. The overlap at the time of the first NLTM occurrence reflects the corresponding noise levels. However, after writing all NLTM signals, the overlap between the network and the actual LTM signal is increased significantly entirely spontaneously by almost 10% at the limit scenario of 20% noise level. The representation of LTM remains at above initial noise levels even after the modifications induced by $ST_{1,2}$.

DISCUSSION

In this work, we focused on binary synapses, which are known to support adequate learning in mathematical models (*35*) and deep-learning algorithms (*28*). The weight of the synapse is a binarized version of its resistive state, and the interplay between intense bidirectional volatility and small nonvolatile residues underlies its palimpsest capability. This concept can be naturally extended to synaptic weights of higher resolution. Also, we would like to point out that, beyond the possible extension to higher efficacy resolution, further improvement toward the wider adoption of this technology can be realized in two key areas. First, systematic studies for improving the fabrication uniformity of memristive synapses would be of high interest, particularly in the scope of large-scale hardware demonstration. Second, while we have prioritized conceptual clarity of results over writing speed and energy efficiency, these parameters are crucial milestones before the integration of this technology in real-world online learning applications. For these reasons, it is clear that future implementations of this work are by no means limited to our selected TiO_2 technology. As long as a candidate technology exhibits bidirectional volatility, it could then be evaluated according to application-specific needs, e.g., write speed, retention time, energy efficiency, etc. Nevertheless, the scope of this study has been strictly focused to the conceptual derisking of palimpsest synapses, which boast very interesting properties in several areas.

We note that unidirectional volatility is already sufficient to support the transition from STM to LTM (*25*–*27*). However, this work differentiates that consolidated memories are also protected from competing memory signals, something that was overlooked by previous studies. Moreover, our synapses' absolute capacity is effectively doubled, and palimpsest functionality has thoroughly been evaluated both in hardware and simulation demonstrations (see table S3 for a detailed comparison with memristive synaptic implementations). These features can be directly attributed to the bidirectional nature of our RRAM technology. It should be noted that bidirectional volatility in these devices has already been characterized for observation windows of up to 2 min (*29*), which could practically extend the memory lifetime of our synapses.

Another remark about this palimpsest memory is that the contents of the memory are in general imperfect reflections of the desired memory. This is not unusual per se since neuro-inspired systems work on the basis of imperfect information typically by default (classifiers sort noisy inputs into neat classes), but, in palimpsest memories, we have the additional factor of LTM-STM relations to consider.

The capabilities of this technology can be interpreted in several distinct ways. First, the palimpsest network can be evaluated in its capacity to recall multiple memories concurrently. While acceptable recall accuracy levels are relative with respect to application-specific needs, the absolute capacity of the network is tied to the number of available time scales in the memristive synapses. Here, nonconsolidated memories can only access the short-term network slot, and thus consecutive STMs interfere destructively with each other. Nevertheless, the correlation statistics of incoming memory streams play a decisive role in the degradation of old signals. To that extent, applications that can afford more noisy recollections are also able to recall a consolidated LTM and multiple random STMs with a mean 50% correlation simultaneously, as illustrated in Fig. 3A. While simple metaplasticity can also suffice for generalizing over multiple highly correlated memories [see (*28*)], the advantage of our technology arises from the ability to remember consolidated states even when the attention of working memory falls on uncorrelated signals. An intuitive representation of at least two palimpsest memories coexisting in the system can be seen in Fig. 4C at snapshots $T \in [0.3\text{–}3]$ s and $T \in [10.1\text{–}11]$ s. In that scenario, memory degradation after recovery is much weaker. Further expanding the consolidation capacity and the initial signal overlap of the network will require manipulating the switching and relaxation dynamics of the memristive synapses such that they operate more flexibly in a proportionally larger number of time scales. This investigation in material science and a resulting more complex electrochemical device structure are certainly of great interest.

Moreover, this technology bears some interesting similarities to how real estate is used for multiple storage in visual working memory systems (*34*, *36*–*38*). Our results within the context of a visual working memory encapsulate best its relevant capabilities. As it is evident from Fig. 4, palimpsest operation may not necessarily need to expand absolute memory capacity to provide computational advantages. Contrarily, it can be enabled using a neural network flexibly without suffering the cost of forgetting older but consolidated signals. This flexibility and LTM reconstructive ability can enhance the performance of in-memory computing (*12*, *39*, *40*) where systems are required to adapt quickly to incoming stimuli and is thus of direct relevance to neuro-inspired applications. In these scenarios, systems benefiting from palimpsest functionality shift their resources on spontaneous online tasks while retaining a core consolidated functionality. As shown in Fig. 3 (B and D), this can occur for hundreds of uncorrelated short-term signals without explicit needs for reinforcing the consolidated counterpart.

Last but not the least, the short-lived span of overlapping memories resembles short-term attention mechanisms, which have recently shown promise toward more complex AI algorithms (*41*). Attention mechanisms can also be implemented using the high-capacity STM familiarity filters that are exhibited here (a familiarity filter is a memory that recognizes when a memory input is present inside the memory even if it no longer has enough information to reconstruct the memory). Illustrated in Fig. 3A, at least 50 uncorrelated memories can pass the familiarity filter simultaneously.

Our memory also implements unsupervised (LTM) memory reconstruction in hardware, supporting previously linked theories of consolidation (*5*, *18*) and optimal recall in the CA3 area of the hippocampus (*42*). This partition of memory storage is an advantageous adaptation since only information that is relevant to a specific cognitive task is needed for undergoing the said task. The dual temporal capacity that is exhibited by our devices resembles the bistable switching that is known to govern synaptic plasticity (*31*). Specifically, the accumulation of nonvolatile residues after LTP/LTD can be thought of as an equivalent mechanism to calcium/calmodulin-dependent protein kinase II, which is considered to be a primary molecular memory mechanism (*43*, *44*).

MATERIALS AND METHODS

Device fabrication

Devices used in this work are vertical metal-insulator-metal (MIM) structures with electrode dimensions of 20 μm by 20 μm as depicted in Fig. 1A. The initial fabrication step was to thermally grow 200 nm of SiO_2 on top of 6-inch silicon wafers, which were used as substrates for the process. This thermal oxide serves as an insulator, separating all devices from the silicon substrate. Each of the three layers in the MIM structure was deposited by following a four-step process, namely, lithography, short reactive ion etching, deposition, and lift-off. Lithography was completed with an EVG 620 TB mask aligner to expose each mask pattern on a negative tone AZ 2070 resist. After each lithography step, a short O_2 plasma cleaning step ensures cleanliness of the area, which has been prepared for material deposition, by removing resist residuals. The deposition step was completed by electron beam evaporation for metal materials and by magnetron sputtering for the active layer material. Bottom electrodes were deposited using the Leybold LAB 700 E-beam equipment. Initially, 5 nm of titanium (Ti) adhesion layer was deposited on top of the thermal oxide and, in continuation, 20 nm of gold (Au) was deposited. After bottom layer deposition, the wafer was soaked in *N*-methyl-2-pyrrolidone (NMP) overnight for lift-off. The middle layer consists of TiO_2 deposited with an EvoVac angstrom engineering dc sputtering equipment. The active layer consists of 25 nm of TiO_2 sputtered at room temperature from a metal Ti target in a 4% O_2/Ar atmosphere and 3-mTorr pressure at 200 W. Following the active layer deposition step, lift-off was carried out with the OPTIwet ST30 tool, which ensures a clean lift-off by spraying hot NMP (60°C) with 3-mbar pressure on the wafer for 30 min. Last, top electrodes were deposited by following the same process as described for the bottom electrodes, with the deposited material in this case being 12 nm of platinum (Pt).

Memristive synapse setup

For all of our experiments, single volatile devices were operated in a binary fashion, dictated by resistance R compared to a chosen threshold value R_{thres}. Plasticity changes were induced following the rule in Eq. 1. Binary signals equal to "1" induce to "0" are causing the synapses to become depressed.

$$\text{plasticity event} = \begin{cases} \text{potentiation, if } V > 0 \\ \text{depression, otherwise} \end{cases} \tag{1}$$

Accordingly, the binary weight w was computed using Eq. 2

$$w = \begin{cases} 1, \text{if } R < R_{\text{thres}} \\ 0, \text{otherwise} \end{cases} \tag{2}$$

RRAM volatility modeling and noise extraction

Memristive volatility was quantified using existing modeling methods (*30*). Specifically, $R(t)$ was expressed via Eq. 3

$$R(t) = \alpha\, e^{-\left(\frac{t}{\tau}\right)^{\beta}} + \gamma \tag{3}$$

Volatility noise was calculated as the percentage (%) difference between the ideal model and real device as shown in Eq. 4. RRAM noise was seen to follow a characteristic Gaussian (normal) distribution. This distribution is in line with existing literature on RRAM reading specific noise that attributes the phenomenon to the activation and deactivation of electron traps in conductive filaments affecting memristive states over time (*45*). More analytical noise data are shown in figs. S4 and S5. By extracting the distribution's mean value μ and SD σ, noise could then be added stochastically on ideal data using Eq. 5

$$(\%)\,\Delta R = \text{noise} = \frac{R_{\text{ideal}} - R}{R} \times 100\% \tag{4}$$

$$p(\text{noise}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(\text{noise}-\mu)^2}{2\sigma^2}} \tag{5}$$

Consolidation of fully destructive memories

To begin our study, we wished to examine our networks performance on the worst-case scenario of two antipodal and fully competing binary memories. We used six memristive synapses that were independently stimulated such that M_1 = [101100] and M_2 = [010011] were written in the long- and short-term time scales, respectively. Specifically, M_1 and M_2 were presented in memory for 15 and 3 consecutive times, respectively. For this conceptual demonstration of our technology, we biased our devices using single ±7-V stimuli for a total duration of 100 μs, while analog states were read at 0.5 V. This profile yielded a significant contrast ratio between plastic and rigid efficacy changes, which has allowed a clearer depiction of palimpsest state overwrites.

It has already been shown that TiO_2-based memristors exhibit more pronounced volatile phenomena when stimulations invasiveness increases, either in the form of larger pulsing amplitudes or numbers of pulses per stimulation (*30*). Here, our choice of ±7 V ensures that the memristive synapses experience large enough volatile changes in efficacy to showcase reversible transitions from LTM to STM. This occurs both from potentiated to depressed states and vice versa. Moreover, this relatively high amplitude has been chosen to ensure that volatility is present even after a single programming pulse per write event, thereby minimizing write speeds. A more detailed depiction of TiO_2 volatility in the context of single pulse write events is shown in fig. S6. These relationships are discussed in greater detail in (*30*). Furthermore, our choice of 0.5-V reading voltage has been made to guarantee good signal-to-noise ratio while ensuring that any read operations are performed at a noninvasive interfacing regime. Thus, synaptic weights can be read at any time with no interference to the consolidated memories. Current-voltage relationships for TiO_2 devices are illustrated at fig. S7 and verify this noninvasive reading regime.

Plasticity changes where induced following the rule in Eq. 1, and the corresponding synaptic states or "weights" were calculated using Eq. 2. A binary threshold value R_{thres} = 10.6 megaohms was explicitly chosen such that individual device histories achieve the best overlap possible. A representation of our logic is shown in fig. S8. We also note that while the device samples occupy different positions in our testing wafer samples (see Fig. 1A), the differences in the induced line resistances are negligible compared to the baseline

R_{thres} value. Specifically, the line resistance in the worst-case scenario for b_2 is 0.003% of the total measured R. Detailed values for all line resistances are provided in table S1.

Last, to obtain the results shown in Fig. 2C, we have performed 200 simulations on the basis of the results shown in fig. S8. Specifically, the noise distribution for each memristive synapse has been extracted using the raw data and the ideal fittings in fig. S6 (obtained using Eqs. 3 and 4). Then, using Eq.5, we have added characteristic noise on all ideal device responses, independently, in each simulation run. The data shown in Fig. 2C are the statistical results obtained from these simulations.

Random memory stream

For this study, we aimed at evaluating our technology's ability to consolidate memories that were uncorrelated in nature. Moreover, we required a network sufficiently large to reflect the statistics of its performance in a smooth manner and avoid the quantization errors shown in Fig. 2. Hence, we devised a network that is composed of 100 identical memristive synapses in simulation by using Eqs. 3 to 5. The synapses' binary threshold was extracted via applying alternating plasticity events and observing the natural occurring equilibrium position (see figs. S5, S9, and S10). Here, R_{thres} is chosen at a value of 11 megaohms. In the worst-case scenario, the device line resistance is about 0.001% of total R. Detailed values for all line resistances are provided in table S1.

The operation parameters of these synapses are included in table S2. To evaluate the technology's performance at a memory level, our main priority has been the symmetrical response to LTP/LTD events such that no binary state is consolidated de facto over time. Devices have this time been stimulated using 500 train pulses (500-μs width each and 10-μs interpulse) at 1.4 and −2.6 V for potentiation and depression events, respectively. Resistance was read at 1.0 V. This profile uses asymmetric stimulation energy but ensures equal writing speeds. Our choices have been made while ignoring the energy efficiency and write speeds of our systems in favor of conceptual and operational clarity. However, volatility in the millisecond range has been reported in HfO_2-based memristors with programming voltages as low as 0.3 V (*46*), which is a promising pathway toward less energy consuming solutions. This smaller time scale may be in line with existing volatility modeling work that shows a clear decrease in volatile phenomena with decreasing stimulation amplitude (*30*). Consequently, dedicated study on ways of increasing RRAM volatility in weaker stimulation regimes is still of great importance, as mentioned in Discussion.

STM lifetime statistics

To compute the histogram of short-time life occurrences, the maximum short-term lifetime of 10 s has been quantized using bins of size 0.1 s. By transitioning from a continuous to a discrete time domain, the total number of occurrences for each bin, o_n, are calculated, resulting to an occurrence vector $\boldsymbol{O} = \{o_0, o_1, \ldots, o_n\}$. The lifetime PDF can then be obtained by dividing $\boldsymbol{O}$ with the total number of STM signals, as shown in Eq. 5.

$$\text{PDF} = \frac{\boldsymbol{O}}{\#\text{STM signals}} \times 100\% \tag{6}$$

Last, the CDF can be obtained by computing the cumulative sum of the PDF for each bin such that

$$\text{CDF}_k = \sum_{k=0}^{i} \text{PDF}_i,\ i \in [0, n] \tag{7}$$

Consolidation of binary images

In this section, we used a memory network of 100 × 100 identical synapses to consolidate binary images. These are the same synapses that were described in Materials and Methods, and the same operation profile was used. The first image was implicitly reconstructed in the LTM using noisy variation of the original signal. Noise was added to the signal by independently choosing to flip each bit with a probability $P = \{0\%, 2\%, 5\%, 10\%, 20\%\}$.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abn7920

REFERENCES AND NOTES

1. G. M. Shepherd, *The Synaptic Organization of the Brain* (Oxford Univ. Press, 2004).
2. C. Koch, *Biophysics of Computation: Information Processing in Single Neurons* (Oxford Univ. Press, 1998).
3. Y. Lecun, Y. Bengio, G. Hinton, *Deep Learning* (Nature Publishing Group, 2015), vol. 521.
4. X. Xu, Y. Ding, S. X. Hu, M. Niemier, J. Cong, Y. Hu, Y. Shi, Scaling for edge inference of deep neural networks. *Nat. Electron.* **1**, 216–222 (2018).
5. M. K. Benna, S. Fusi, Computational principles of synaptic memory consolidation. *Nat. Neurosci.* **19**, 1697–1706 (2016).
6. W. C. Abraham, Metaplasticity: Tuning synapses and networks for plasticity. *Nat. Rev. Neurosci.* **9**, 387–399 (2008).
7. S. Stathopoulos, A. Khiat, M. Trapatseli, S. Cortese, A. Serb, I. Valov, T. Prodromakis, Multibit memory operation of metal-oxide bi-layer memristors. *Sci. Rep.* **7**, 17532 (2017).
8. A. Serb, J. Bill, A. Khiat, R. Berdan, R. Legenstein, T. Prodromakis, Unsupervised learning in probabilistic neural networks with multi-state metal-oxide memristive synapses. *Nat. Commun.* **7**, 12611 (2016).
9. I. Gupta, A. Serb, A. Khiat, R. Zeitler, S. Vassanelli, T. Prodromakis, Real-time encoding and compression of neuronal spikes by metal-oxide memristors. *Nat. Commun.* **7**, 12805 (2016).
10. R. Berdan, E. Vasilaki, A. Khiat, G. Indiveri, A. Serb, T. Prodromakis, Emulating short-term synaptic dynamics with memristive devices. *Sci. Rep.* **6**, 18639 (2016).
11. J. H. Yoon, Z. Wang, K. M. Kim, H. Wu, V. Ravichandran, Q. Xia, C. S. Hwang, J. J. Yang, An artificial nociceptor based on a diffusive memristor. *Nat. Commun.* **9**, 417 (2018).
12. I. Boybat, M. Le Gallo, S. R. Nandakumar, T. Moraitis, T. Parnell, T. Tuma, B. Rajendran, Y. Leblebici, A. Sebastian, E. Eleftheriou, Neuromorphic computing with multi-memristive synapses. *Nat. Commun.* **9**, 2514 (2018).
13. S. La Barbera, D. R. B. Ly, G. Navarro, N. Castellani, O. Cueto, G. Bourgeois, B. De Salvo, E. Nowak, D. Querlioz, E. Vianello, Narrow heater bottom electrode-based phase change memory as a bidirectional artificial synapse. *Adv. Electron. Mater* **4**, 1800223 (2018).
14. 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-scale neural network (165 000 Synapses) using phase-change memory as the synaptic weight element. *IEEE Trans. Electron Devices.* **62**, 3498–3507 (2015).
15. S. Ambrogio, N. Ciocchini, M. Laudato, V. Milo, A. Pirovano, P. Fantini, D. Ielmini, Unsupervised learning by spike timing dependent plasticity in phase change memory (PCM) synapses. *Front. Neurosci.* **10**, 56 (2016).
16. Y. Demirağ, F. Moro, T. Dalgaty, G. Navarro, C. Frenkel, G. Indiveri, E. Vianello, M. Payvand, PCM-Trace: Scalable Synaptic Eligibility Traces with Resistivity Drift of Phase-Change Materials, in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, (2021), pp. 1–5.
17. W. C. Abraham, M. F. Bear, Metaplasticity: The plasticity of synaptic plasticity. *Trends Neurosci.* **19**, 126–130 (1996).
18. S. Fusi, P. J. Drew, L. F. Abbott, Cascade models of synaptically stored memories. *Neuron* **45**, 599–611 (2005).
19. S. Brivio, D. Conti, M. V. Nair, J. Frascaroli, E. Covi, C. Ricciardi, G. Indiveri, S. Spiga, Extended memory lifetime in spiking neural networks employing memristive synapses with nonlinear conductance dynamics. *Nanotechnology* **30**, 015102 (2019).
20. X. Zhu, C. Du, Y. Jeong, W. D. Lu, Emulation of synaptic metaplasticity in memristors. *Nanoscale* **9**, 45–51 (2017).
21. Q. Wu, H. Wang, Q. Luo, W. Banerjee, J. Cao, X. Zhang, F. Wu, Q. Liu, L. Li, M. Liu, Full imitation of synaptic metaplasticity based on memristor devices. *Nanoscale* **10**, 5875–5881 (2018).

22. C. Cheng, Y. Li, T. Zhang, Y. Fang, J. Zhu, K. Liu, L. Xu, Y. Cai, X. Yan, Y. Yang, R. Huang, Bipolar to unipolar mode transition and imitation of metaplasticity in oxide based memristors with enhanced ionic conductivity. *J. Appl. Phys.* **124**, 152103 (2018).
23. B. Liu, Z. Liu, I. S. Chiu, M. Di, Y. Wu, J. C. Wang, T. H. Hou, C. S. Lai, Programmable synaptic metaplasticity and below femtojoule spiking energy realized in graphene-based neuromorphic memristor. *ACS Appl. Mater. Interfaces* **10**, 20237–20243 (2018).
24. T. H. Lee, H. G. Hwang, J. U. Woo, D. H. Kim, T. W. Kim, S. Nahm, Synaptic plasticity and metaplasticity of biological synapse realized in a KNbO3Memristor for application to artificial synapse. *ACS Appl. Mater. Interfaces* **10**, 25673–25682 (2018).
25. T. Chang, S. H. Jo, W. Lu, Short-term memory to long-term memory transition in a nanoscale memristor. *ACS Nano* **5**, 7669–7676 (2011).
26. Z. H. Tan, R. Yang, K. Terabe, X. B. Yin, X. D. Zhang, X. Guo, Synaptic metaplasticity realized in oxide memristive devices. *Adv. Mater.* **28**, 377–384 (2016).
27. T. Ohno, T. Hasegawa, T. Tsuruoka, K. Terabe, J. K. Gimzewski, M. Aono, Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nat. Mater.* **10**, 591–595 (2011).
28. A. Laborieux, M. Ernoult, T. Hirtzlin, D. Querlioz, Synaptic metaplasticity in binarized neural networks. *Nat. Commun.* **12**, 2549 (2021).
29. C. Giotis, A. Serb, S. Stathopoulos, L. Michalas, A. Khiat, T. Prodromakis, Bidirectional volatile signatures of metal-oxide memristors-part I: Characterization. *IEEE Trans. Electron Devices.* **67**, 5158–5165 (2020).
30. C. Giotis, A. Serb, S. Stathopoulos, T. Prodromakis, Bidirectional volatile signatures of metal-oxide memristors-part II: Modeling. *IEEE Trans. Electron Devices.* **67**, 5166–5173 (2020).
31. U. S. Bhalla, Molecular computation in neurons: A modeling perspective. *Curr. Opin. Neurobiol.* **25**, 31–37 (2014).
32. I. Messaris, A. Serb, S. Stathopoulos, A. Khiat, S. Nikolaidis, T. Prodromakis, A data-driven verilog-A ReRAM Model. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **37**, 3151–3162 (2018).
33. S. Fusi, L. F. Abbott, Limits on the memory storage capacity of bounded synapses. *Nat. Neurosci.* **10**, 485–493 (2007).
34. G. Mongillo, O. Barak, M. Tsodyks, Synaptic theory of working memory. *Science* **319**, 1543–1546 (2008).
35. M. K. Benna, S. Fusi, Efficient online learning with low-precision synaptic variables, in *2017 51st Asilomar Conference on Signals, Systems, and* Computers (IEEE, 2017), pp. 1610–1614.
36. L. Matthey, P. M. Bays, P. Dayan, A probabilistic palimpsest model of visual short-term memory. *PLOS Comput. Biol.* **11**, 1004003 (2015).
37. S. J. Luck, E. K. Vogel, The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
38. W. Zhang, S. J. Luck, Discrete fixed-resolution representations in visual working memory. *Nature* **453**, 233–235 (2008).
39. M. Le Gallo, A. Sebastian, R. Mathis, M. Manica, H. Giefers, T. Tuma, C. Bekas, A. Curioni, E. Eleftheriou, Mixed-precision in-memory computing. *Nat. Electron.* **1**, 246–253 (2018).
40. S. G. Hu, Y. Liu, Z. Liu, T. P. Chen, J. J. Wang, Q. Yu, L. J. Deng, Y. Yin, S. Hosaka, Associative memory realized by a reconfigurable memristive Hopfield neural network. *Nat. Commun.* **6**, 7522 (2015).
41. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, in *Advances in Neural Information Processing Systems* (NIPS, 2017), vol. 2017-Decem, pp. 5999–6009.
42. C. Savin, P. Dayan, M. Lengyel, Optimal recall from bounded metaplastic synapses: Predicting functional adaptations in hippocampal area CA3. *PLoS Comput. Biol.* **10**, e1003489 (2014).
43. J. Lisman, H. Schulman, H. Cline, The molecular basis of CaMKII function in synaptic and behavioural memory. *Nat. Rev. Neurosci.* **3**, 175–190 (2002).
44. P. Miller, A. M. Zhabotinsky, J. E. Lisman, X.-J. Wang, The stability of a stochastic CaMKII switch: Dependence on the number of enzyme molecules and protein turnover. *PLOS Biol.* **3**, 0705–0717 (2005).
45. D. Veksler, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Muraviev, B. Chakrabarti, E. Vogel, D. C. Gilmer, P. D. Kirsch, in *IEEE International Reliability Physics Symposium Proceedings* (IEEE, 2013).
46. E. Covi, D. Ielmini, Y. H. Lin, W. Wang, T. Stecconi, V. Milo, A. Bricalli, E. Ambrosi, G. Pedretti, T. Y. Tseng, in *26th IEEE International Conference on Electronics, Circuits and Systems, ICECS 2019* (IEEE, 2019), pp. 903–906.

Acknowledgments
Funding: The authors acknowledge the support of the EPSRC FORTE Programme Grant (EP/R024642/1), the RAEng Chair in Emerging Technologies (CiET1819/2/93), and the EU project SYNCH (824162). **Author contributions:** Original idea: T.P. and A.S. Experimental design: A.S. and C.G. Investigation: C.G. Sample fabrication: S.S. and V.M. Supervision: T.P. and A.S. Writing—original draft: C.G. and A.S. Writing—review and editing: C.G., A.S., and T.P. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Simulations and analysis for this work have been done in Python. Existing data for hardware analysis and the necessary code for recreating the simulation data can be found in Zenodo (https://doi.org/10.5281/zenodo.6513918).

Submitted 20 December 2021
Accepted 9 May 2022
Published 22 June 2022
10.1126/sciadv.abn7920