Ultralow power dual-gated subthreshold oxide neuristors : an enabler for higher order neuronal temporal correlations

John, Rohit Abharam; Tiwari, Nidhi; Chen, Yaoyi; Ankit; Tiwari, Naveen; Kulkarni, Mohit; Nirmal, Amoolya; Nguyen, Anh Chien; Basu, Arindam; Mathews, Nripan

2018

John, R. A., Tiwari, N., Chen, Y., Ankit, Tiwari, N., Kulkarni, M., . . . Mathews, N. (2018). Ultralow power dual-gated subthreshold oxide neuristors : an enabler for higher order neuronal temporal correlations. ACS nano, 12(11), 11263-11273. doi:10.1021/acsnano.8b05903

https://hdl.handle.net/10356/138428

https://doi.org/10.1021/acsnano.8b05903

Ultralow Power Dual Gated Sub-Threshold Oxide Neuristors: An Enabler For Higher Order Neuronal Temporal Correlations

Rohit Abraham John[1], *Nidhi Tiwari*[2], *Chen Yaoyi*[1], *Ankit*[1], *Naveen Tiwari*[1], *Mohit Kulkarni*[1], *Amoolya Nirmal*[1], *Anh Chien Nguyen*[1], *Arindam Basu*[3]*, *Nripan Mathews*[1,2]*

[1] School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

[2] Energy Research Institute @ NTU (ERI@N), Nanyang Technological University, Singapore 637553

[3] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

* Corresponding author

Prof. Arindam Basu (Email: arindam.basu@ntu.edu.sg)

Prof. Nripan Mathews (Email: Nripan@ntu.edu.sg)

Keywords: synaptic transistors, neuromorphic computing, heterosynaptic plasticity, homeostasis, classical conditioning, coincidence detection

ABSTRACT

Inspired by neural computing, the pursuit of ultralow power neuromorphic architectures with highly distributed memory and parallel processing capability has recently gained more traction. However, emulation of biological signal processing *via* artificial neuromorphic architectures does not exploit the immense interplay between local activities and global neuromodulations observed in biological neural networks, and hence are unable to mimic complex biologically plausible adaptive functions like heterosynaptic plasticity and homeostasis. Here, we demonstrate emulation of complex neuronal behaviours like heterosynaptic plasticity, homeostasis, association, correlation and coincidence in a single neuristor *via* a dual-gated architecture. This multiple gating approach allows one gate to capture the effect of local activity correlations and the second gate to represent global neuromodulations, allowing additional modulations which augment their plasticity and enabling higher order temporal correlations at a unitary level. Moreover, the dual-gate operation extends the available dynamic range of synaptic conductance while maintaining symmetry in the weight-update operation, expanding the number of accessible memory states. Finally, operating neuristors in the sub-threshold regime enables synaptic weight changes with high gain, while maintaining ultralow power consumption of the order of femto-Joules.

With massive parallelism, highly distributed memory, fault tolerance, self-learning, robustness and ultra-low power consumption, neural computing transcends serial processing in traditional von Neumann architectures made of fault-sensitive electronic transistors.[1,2] Learning in the human brain occurs *via* short and long-term strengthening and weakening of synaptic connections, interconnecting each neuron to up to 10^4 other neurons.[3] This modification of synaptic strength is referred to as functional plasticity (as opposed to structural plasticity) and exists at various time scales of operation. Short-term plasticity refers to temporary rapid changes in synaptic strength that act on a timescale of tens of milliseconds to a few minutes unlike long-term plasticity, which lasts from minutes to hours.[4] While short-term facilitation serves as a working memory, short-term depression removes auto-correlations.[5] On the other hand, long-term potentiation (LTP) contributes towards spatial memory storage while long-term depression (LTD) also encodes spatial features, weakens synapses selectively and clears old memory traces.[6] Combination of LTP and LTD, when triggered by spike time correlations is called spike time dependent plasticity (STDP).[7] In contrast to conventional memory storage and logic operations, neuromorphic applications set unique requirements on the hardware switching devices, namely- non-abrupt analog-like switching transitions and continuous distribution of conductance states with a wide dynamic range and linearity. The massively parallel signal processing of the human brain also demands emulation of the interconnectedness seen in biological neural networks, but is often overlooked. This calls for the need of device configurations that can not only emulate fundamental local synaptic features, but which also incorporates response to global neuromodulations at an elemental level.

While neuromorphic architectures based on conventional silicon circuitry are constrained by the requirement of additional clock cycles and peripheral circuitry for weight-updates and synchronisation, abrupt conductance transitions in drift-memristors limit the number of accessible states and hence, their plasticity.[8-10] Investigations on electrolyte-gated thin film transistors and ion-diffusive memristors have pushed the field forward with comprehensive emulation of short and

long-term plasticity rules, but fails to emulate complex neuronal behaviours like heterosynaptic plasticity and homeostasis, dependent on interplay between local activities and global neuromodulations. With a gated control of channel conductance exploiting time-dependent hysteresis, the thin-film transistor (TFT) configuration is promising in achieving dynamic and linear plasticity. Emulation of multi-factor or heterosynaptic plasticity that accounts for the immense interplay between local and global neuromodulations, necessitates synergistic gating strategies to control charge trapping mechanisms and address distinct memconductance states of semiconducting channels. This would allow development of highly plastic neuromorphic circuitry emulating biological adaptation mechanisms with drastically lesser neural elements, mitigating strict circuit density requirements. Maximizing and controlling the degree of such plasticity also calls for the utilization of high capacitance gate dielectrics intimately modulating the carrier concentration at the semiconductor-dielectric interface and satisfying the quest for ultralow power operation.

Very recently we demonstrated synergistic multiple gating of 2D chalcogenide neuristors which enabled metaplasticity and homeostasis in a single device.[11] However, these devices were limited by complex exfoliation and fabrication techniques which presently limit scalability, and the metaplasticity features were encoded only with bias deflections. With high intrinsic carrier mobility, optical transparency, and an established process flow developed for display applications, amorphous oxides of post-transition elements (In, Ga, Sn, Zn, W) serve as ideal semiconducting platforms to encode memconductances based on the semiconductor-dielectric capacitive coupling.[12,13] Low voltage operation of basic synaptic characteristics have been demonstrated by utilizing high-capacitance liquid electrolytes, but are disadvantaged by their chemically volatile nature.[14,15] This necessitates development of solid-state ionic dielectrics for ultra-low power operation of artificial synaptic devices.

Herein, artificial synapses with intimate electronic-ionic coupling were realized in a dual-gated electric-double-layer neuristor (EDLN) configuration with semiconducting indium-tungsten oxide (IWO) channels and solid-state ionic dielectric based on poly (vinylidene fluoride- co - hexafluoropropylene), P(VDF-HFP), and the ionic liquid 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl) imide, [EMI][TFSI]. High tensile strength of the structuring polymer allowed convenient fabrication of free-standing films and lamination on to the desired substrate *via* the 'cut and stick' processing strategy.[16] Ionic migration-relaxation kinetics in the ion gel mimicking Ca^{2+} influx in dendritic spines modulated the electronic conductance-state of the semiconducting channel, in addition to carrier trapping-detrapping mechanisms at the IWO-SiO_2 interface, while charge transport pathways emulated the synaptic cleft, and channel-conductance defined the synaptic weight. Temporally correlated pre- and post-synaptic action potentials initially created short-term volatile changes in the channel conductance/weights, which further consolidated to long-term non-volatile changes upon persistent training. With a multi-gated architecture, these neuromorphic transistors or neuristors could operate independently in a pristine ionotronic mode-capturing the effect of local activity correlations or electronic mode-capturing global neuromodulations. More importantly, they can operate synergistically in a dual-gated additive/subtractive mode addressing the immense interplay between local and global neuromodulations and accessing a multitude of memconductance states with dynamic linear plasticity. The dual gated approach created temporal cross-correlations among the various pre- and post-synaptic action potentials, introducing heterosynaptic or three-factor plasticity rules that emulate the effects of neuromodulators like dopamine or noradrenaline.[17,18] Persistent classical conditioning experiments spurred associative learning between the gating modes with features like extinction and recovery temporally encoded into the training algorithms. Sub-threshold operation of transistors resulted in conductance/weight changes with maximum gain and was utilized to create highly sensitive coincidence detectors operating at ultra-low power. These comprehensive results portray the advantages of a multi-gated configuration and sub-threshold operation of oxide

neuristors for ultralow-power, high-gain weight changes, facilitating development of metaplastic architectures with biologically plausible adaptation mechanisms dependent on local activity and global neuromodulations.

RESULTS AND DISCUSSION

Learning and memory in the human brain happens through activity-dependent strengthening and weakening of synaptic connections, also known as **synaptic plasticity**.[3] Action potentials trigger opening of voltage-gated Ca^{2+} channels, transducing electrical potentials into chemical signals through the release of neurotransmitters. The neurotransmitter-receptor interaction activates ligand gated ion-channels, depolarizing or hyperpolarizing the resting membrane potential based on the type of channel (Figure 1 A).[19,20] Accumulated synaptic input may cause the post-synaptic cell in turn to fire an action potential and the temporal correlations between local pre-and post-synaptic firing activity is responsible for plasticity of the synaptic strengths. Similar activity triggered at 10^4 such neuronal junctions/synapses forms the basis of parallel computing and distributed memory in the human brain. Successful emulation of these neuronal dynamics require effective translation of modulation of local postsynaptic strength *via* changes in ion flux due to local presynaptic action-potentials. Additionally, since large groups of neurons receive common modulatory inputs in the form of neuromodulators such as dopamine, noradrenaline or acetylcholine signalling reward or surprise signals in biological networks, modulation of learning rules by such global factors need to be accounted for as well.[21] While most studies focus only on the emulation of plasticity due to local temporally correlated activity, global neuromodulations are often overlooked diluting the achievable plasticity. In the investigated EDLN configuration, pre- and post-synaptic potentials at the gate and source terminals modulated the post-synaptic channel conductance, read *via* the drain terminal (Figure 1 A). While pure electronic-mode transistors exhibited a field-effect mobility of 9 $cm^2V^{-1}s^{-1}$, large electrical-double-layer (EDL) capacitance of the ion gels resulted in efficient carrier accumulation and hence, high-performance ionotronic transistors (field-effect mobility ~ 79

$cm^2V^{-1}s^{-1}$) operating at ultralow voltage (V_{gs} ~ 1.5 V and V_{ds} ~ 1-100 mV) and power. (Supporting-Information Note-1 Figure S1). Each mode could be operated independently to program specific memconductance states into the semiconducting channel and exhibit plasticity. Concurrently, the dual-gating operation could be configured with one gate capturing the effect of local activity correlations and the second gate representing global neuromodulations, allowing the second gate terminal to modulate the weight changes caused by the first gate, augmenting the plasticity and enabling higher order temporal correlations at a unitary level.

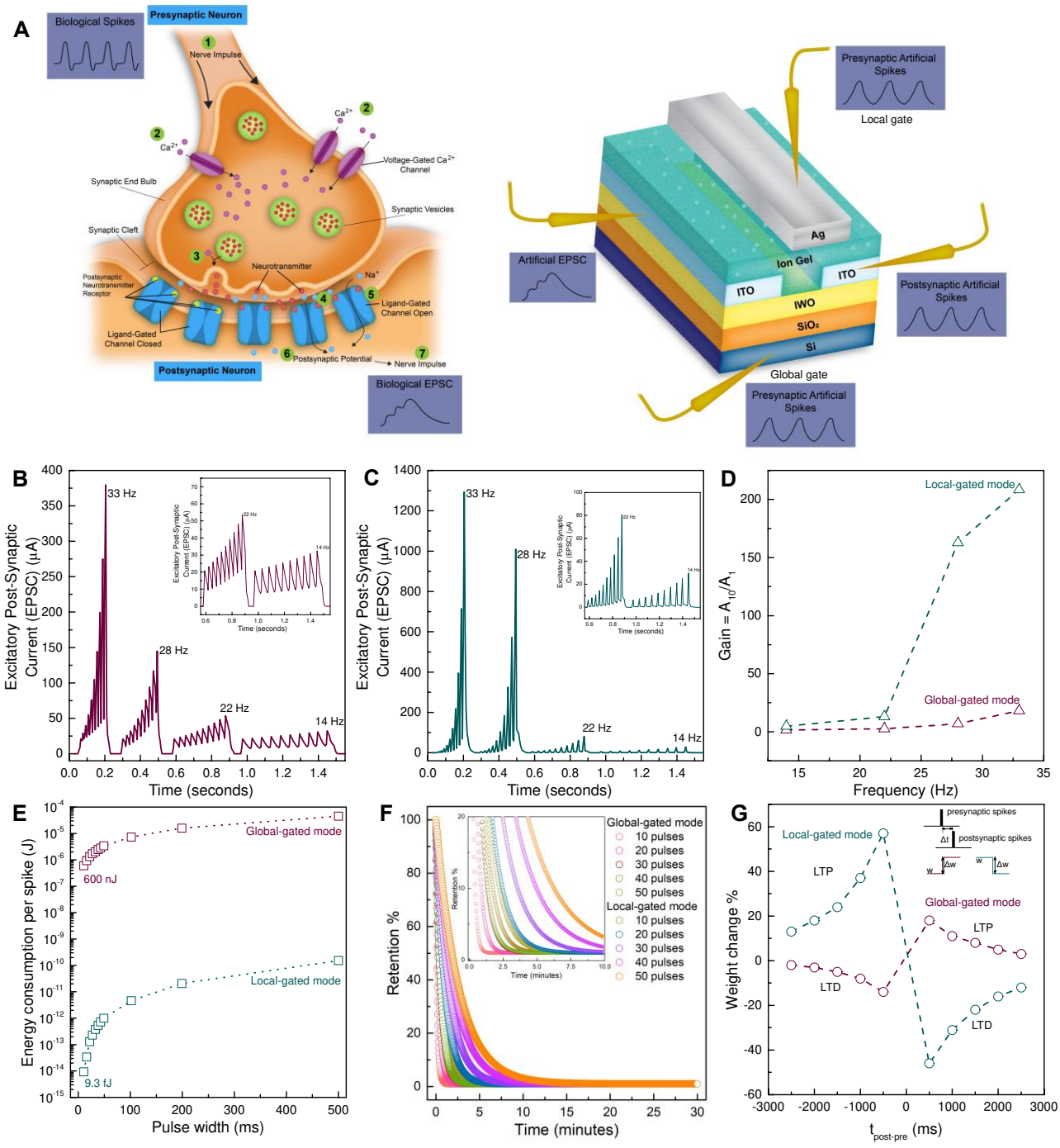

Figure 1. Interplay between short and long-term plasticity in oxide neuristors. (**A**) Synaptic transmission steps (left) and the proposed multi-gated architecture of analogous artificial oxide neuristors (right). Frequency dependent volatile changes in conductivity/weights create high-pass filtering of signals in the (**B**) global-gated and (**C**) local-gated modes of operation. (**D**) Frequency of presynaptic action potentials determine the gain of temporal filters. (**E**) Comparison of the ON-state energy consumption per spike (E_{on}) of oxide neuristors in the global and local-gated modes of operation. E_{on} was calculated from the equation $E_{on} = I_{peak,first} \times t \times V_{ds}$; where I_{peak} is the peak value of 1st generated EPSC, t is the spike duration, and V_{ds} is the applied drain voltage.[22] Scaling down of device dimensions and reducing spike durations to sub-millisecond levels could be also utilized as strategies to decrease the energy consumption per event even further. The values reported are for the best performing devices. (**F**) The forgetting curve depicts the decline of memory retention with time. Highest numerical value of the long-term weight changes in the two operational modes were normalized to compare the retention characteristics. Repetitive learning softened the forgetting behaviour in both modes as shown. (**G**) Relative timing between pre- and postsynaptic spikes created voltage differences across the neuristors, permanently strengthening and weakening the synaptic connections, also called Spike-timing-dependent plasticity (STDP). While the weight change pattern in the global-gated mode followed a STDP rule, a "reverse" STDP (rSTDP) rule was exhibited in the local-gated mode.

Neuronal plasticity occurs at various levels of brain organization and involves changes in the efficiency of transmission-functional plasticity and/or changes in the number of connections between neurons- structural plasticity. Classified based on the timescales of operation, short and long-term plasticity rules defining learning and memory were studied by recording excitatory/inhibitory post synaptic currents (E/IPSCs) in response to pre- and post-synaptic training sequences.[10,23-24] These standard plasticity rules were implemented independently in both global and local-gated modes of operation to benchmark the nature of plasticity achieved. A training sequence consisting of 10 presynaptic action potentials of pulse width 20 ms (presynaptic amplitude = + 20 V (global-gated mode) and + 1.5 V (local-gated mode), drain bias = + 1 V (global-gated mode) and + 0.1 V (local-gated mode)) induced dynamic high-pass temporal filtering of signals *via* volatile changes in channel-conductance with a frequency-dependent gain as shown in Figures 1 B-D.[25] An extension of the paired-pulse facilitation/depression phenomena (Supporting-Information Note-2 Figure S2), polarity of the presynaptic action potentials determined the nature of post-synaptic signals (excitatory/ inhibitory) and the frequency of activation decided the gain. Slow nature of ion migration-relaxation dynamics of the ionotronic-mode (local-gated) resulted in superior facilitation (181 %) and depression indices (55 %) and thus, larger filter gains (209 at 33 Hz) when compared to the electronic-mode (global-gated) (PPF = 138 %, PPD = 70 %, Gain = 18 at 33 Hz, Figures 1 B-D, Supporting-Information Note-2 Figure S2, Table T1).[11,26] Short and temporary nature of the training sequences resulted in the decay of post-synaptic currents (PSCs) back to the initial state due to carrier trapping (global-gated) and ion relaxation (local-gated) mechanisms, analogous to **short-term plasticity**.[5] With neuristors pushed to operate at 1 mV (V_{ds}), the local-gated operation augmented this high-gain behaviour with an ultralow power consumption ~ 9.3 fJ per synaptic event (calculated from peak value of the 1st EPSC), comparable to biological synapses (~ 10 fJ per event) and one of the lowest reported till date (Figure 1 E, Supporting-Information Note-3 Figure S3, Table T2).[11]

Persistent training sequences with longer pulse width consolidated the weight changes, leading to non-volatile changes in channel-conductance and hence **long-term plasticity**.[10,27] Being an experience-dependent phenomena, the magnitude of weight changes could be tuned by the number and amplitude of presynaptic training spikes and the modulatory drain bias, resulting in controlled facilitation/depression in accordance with the quantal and probabilistic neurotransmitter release model.[28] Slow relaxation nature of the ions in the local-gated mode once again resulted in larger weight changes and higher retention of the memory states as depicted by the **Ebbinghaus forgetting curves** (Figure 1 F, Supporting-Information Note-4 Figure S4). Spaced repetition resulted in softening of the downward slope of the forgetting curve, indicating modulation of the strength of memory and process of forgetting that occurs with the passage of time.

Memory traces of past experiences were then encoded as temporal correlations between pre- and post-synaptic neurons, resulting in a temporally asymmetric form of Hebbian learning or **Spike Timing Dependent Plasticity (STDP)**.[29] Effective translation of temporal correlations into voltage amplitude differences created long-term facilitation/depression in global/local-gated synapses on the repeated arrival of presynaptic spikes a few hundred milliseconds before the postsynaptic action potentials. Positive effective writing voltages above the threshold resulted in permanent pinning of the ions in the ion gel at the semiconductor-dielectric interface saturating the Helmholtz layer, thereby causing facilitation in the local-gated mode. On the other hand, positive effective writing voltages resulted in depression in the global-gated mode indicating electron detrapping during the pulse application and subsequent slow trapping of carriers on removal of positive pulses/bias (Supporting-Information Note-4 Figure S4).[11,30] Changes to the order of arrival reversed the polarity of effective writing voltages and direction of weight-change and followed an asymmetric Hebbian rule in the global-gated mode and an asymmetric anti-Hebbian in the local-gated mode of operation (Figure 1 G, Supporting-Information Note-5 Figure S5) similar to learning rules observed in apical dendritic synapses.[31] Such reverse STDP rules have been shown to be computationally useful for

associative learning in concert with post-spike hyperpolarization.[32,33] While the exact roles and interplay between Hebbian and anti-Hebbian rules are still not clear, it is reassuring that our neuristors can be configured to display either rule providing the flexibility desirable in neuromorphic hardware without the need for additional circuits as would be needed in standard CMOS implementations.

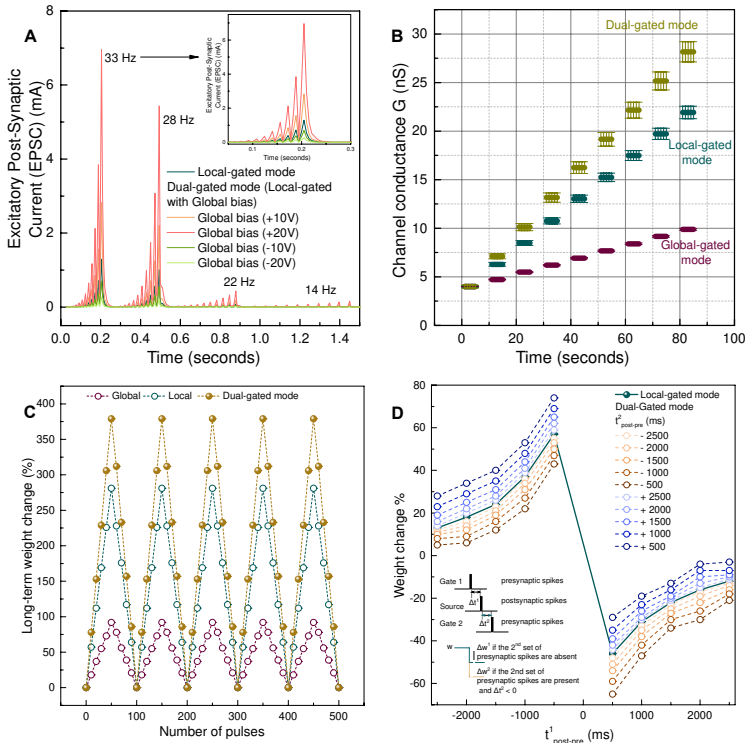

Figure 2. Dual-gated configuration enables multi-level memory and higher order temporal correlations in neuristors. (**A**) Modulation of temporal filter characteristics with additional gating. Level shifting was achieved with the second control gate and without any additional circuitry. (**B**) Linear variation of channel conductance with write operations featuring 8 (3-bit storage) distinct conductance levels for all 3 operational modes but spanning different conductance ranges. Symmetric presynaptic voltage write-pulses of constant magnitude + 1.5 V (local-gated), +20 V (global-gated) and number (10) were applied to generate the conductance linearity curve. The memconductance states were addressed using a drain bias of + 0.1 V. (**C**) Controlled facilitation and depression were achieved in our devices by applying a series of potentiating and depressing presynaptic spikes. Potentiating spikes for the global-gated mode involved spikes of amplitude - 20 V and pulse width 500 ms, and + 1.5 V for the local-gated mode. The dual-gated mode allowed weight modulations with improved strength. For the dual-gated mode, both global and local presynaptic spikes were applied simultaneously. (**D**) The dual gating approach allowed the second gate terminal to modulate the weight changes caused by the first gate, augmenting the plasticity and enabling higher order temporal correlations. The nature and degree of the effective conductance/weight change depended on temporal correlations between the postsynaptic sequence and two different input sequences (pre-synaptic and modulatory), resulting in heterosynaptic plasticity.

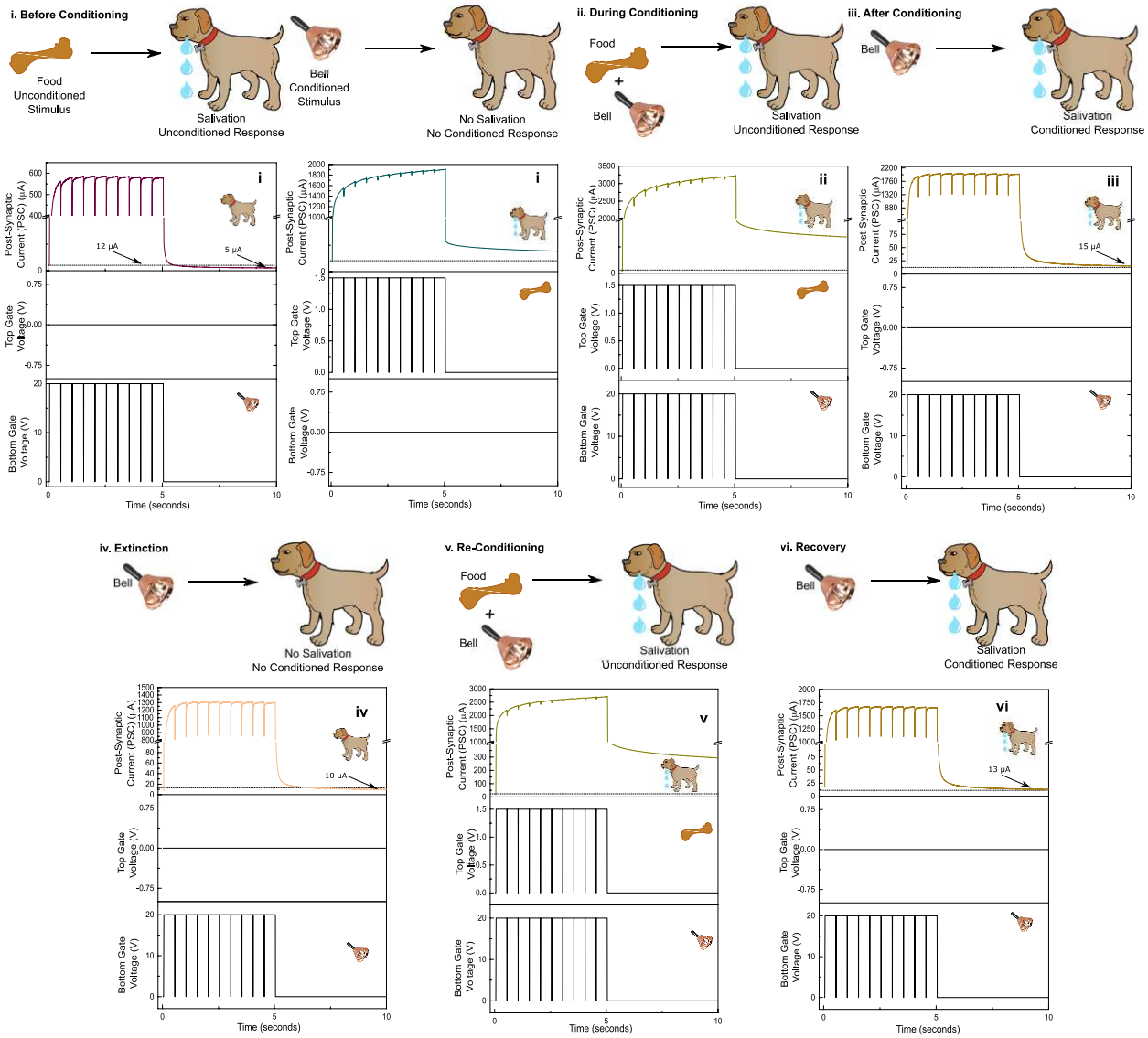

Figure 3. Schematic of the concept of Pavlov's dog experiment for associative memory. Prior to conditioning, the Pavlov's dog salivated to unconditioned stimuli alone (stage: i). Here, 12 μA was set as the threshold for salivation. Persistent training with simultaneous application of unconditioned and conditioned stimuli created association between them, after which the Pavlov's dog salivated to independent application of conditioned stimuli (stages: ii-iii). The association was weakened and became extinct with time on training with conditioned stimuli alone (stage: iv). However, retraining helped recover this association within a shorter time scale (stage: v-vi).

The proposed dual-gated architecture of the neuristors offer unique advantages in implementing traditional machine learning update rules as well as more bio-realistic plasticity mechanisms. Like all electronic hardware, neuromorphic circuitry also suffers from restricted tunability and unit-level fluctuations causing parameter inaccuracies to serious malfunctioning. In network implementations like the Hopfield model, this results in multiple solutions for each input due to the dynamically changing nature of threshold voltages of individual neurons which create localized minima states for each input.[34,35] In competitive learning networks, high intrinsic firing rates of some neurons may

prohibit other neurons from learning unless homeostasis is introduced.[36] However, inclusion of homeostasis in silicon synapses require additional circuits requiring more power and area.[37] In the proposed multi-gated configuration, the additional gate could be biased using homeostatic feedback inputs, applying correction currents to the input of neurons as shown in Figure 2 A. Carrier accumulation in the local-gated mode could be modulated *via* necessary biasing at the global gate and vice versa, level-shifting the temporal filter characteristics. Positive biasing at the global gate shifted the entire temporal filter characteristics in the positive direction, whereas negative biasing resulted in a negative shift without affecting the frequency-gain characteristics. In contrast to the STDP-based homeostasis previously reported which focussed on only long-term weight changes, this implementation extends the homeostatic control to short-term plasticity rules, regulating the stability of neural networks at both time scales of operation.[11]

While scaling strategies to increase the data storage capabilities of non-volatile memories are hampered by stringent photolithography and other fabrication complexities, development of multilevel memories seems to be a more viable alternative for neuromorphic machine learning. Creating such multilevel memories would require a device with dynamic wide linear conductance range and distinguishable read-out states.[38] In terms of computing, cognitive tasks like pattern recognition and classification also demand synaptic devices with excellent conductance linearity (CL) and multi-level cell (MLC) characteristics.[39] Most drift memristors depict restricted memory windows with non-linear weight updates and asymmetric behaviour between the processes of conductance increase (potentiation) and decrease (depression) due to the differences in switching kinetics of filament formation and dissolution.[40] With a wide linear conductance range that can be modulated by pre and post-synaptic voltages, neuristors become superior substitutes for computationally intensive applications like dot product engines (DPEs), which require such 'blind update' schemes.[41,42] As a proof of concept, linear symmetric postsynaptic conductance changes featuring 8 distinct levels (equivalent to 3-bit storage) were recorded for both global and local-gated

operational modes with symmetric presynaptic write-pulses of constant magnitude (Figure 2 B). The individual steps of linear increment depended on the net capacitive coupling in the three modes of operation. While conductance/ weight changes in the global-gated mode resulted from carrier trapping-detrapping mechanisms at the semiconductor-dielectric interface, ion migration-relaxation effects in the ion gel induced similar modulations of synaptic weights in the local-gated mode. With electrical-double-layer formation across nanometer thick Helmholtz layers, the local gate utilized stronger capacitive coupling to widen the linear conductance range. Configuring the neuristor in a dual-gated mode enhanced charge carrier accumulation due to a combination of carrier trapping and ion migration effects, resulting in a larger memory window. In system level operation with provisions for pulsing either gate or both, this leads to an increase in number of conductance states with an increased storage capacity of ~ 4.5 bits/synapse from the original 3 bits. Thus, adopting such multi-gated architectures to create multi-level memories becomes a superior alternative with reduced quantization errors in mapping continuous weights to discrete conductance levels. With more optimized writing schemes, the linear conductance range and the number of accessible states can be further widened and is limited only by the extent of triode region of the neuristor.

Implementation of online learning requires symmetric weight increment and decrement operations for outer product weight update rules. This symmetry along with low “write” noise and low switching voltages and currents determine the efficiency of crossbar neuromorphic computational kernels like vector–matrix multiplication and parallel rank 1 weight updates, and significantly improves classification accuracy of backpropagation schemes.[41] In contrast to structural transformations which decides the switching behaviour in memristors and phase-change memories (PCM), the gating mechanisms employed here create softer permanent changes in channel conductance in these neuristors, resulting in low “write” noises. The large electrical-double-layer capacitances also result in ultralow switching voltages with the local-gated mode of operation. Pulsing the neuristors with optimized writing schemes resulted in near-linear weight changes with

symmetry in facilitation and depression. With extremely large electrical-double-layer capacitances, the rate of modulation with local-gated mode always dominated the global-gated mode at low biases, exhibiting a much larger memory window. Additive operation of the two modes strengthened the degree of facilitation/depression with higher slopes compared to the independent operation (Figure 2 C). Analogous to **homeostasis**, subtractive operation of the second gate could be utilized to create a negative feedback to offset excessive excitation or inhibition in the other mode.[43,44] For example, potentiation with the global-gate could be depressed by local-gated mode at a rate different from a pure global-gated operation (Supporting-Information Note-6 Figure S6). Thus operation in the dual-gated mode allowed programming of weight changes with fine precision, with the net weight change defined by the overall capacitive coupling across the semiconducting channel.

Co-existence of multiple forms of synaptic plasticity would increase the processing capability and memory storage capacity of neuristors. In this configuration, the additional gate control terminal was utilized as a switch controlled *via* presynaptic rate-coded schemes or neuromodulation schemes like dopaminergic, noradrenergic, muscarinic, and nicotinic receptors to dynamically regulate weight changes, resulting in a higher order of plasticity, also called **heterosynaptic plasticity or three-factor learning**.[17,45–50] Dual-gating resulted in superposition of carrier trapping-detrapping and ion migration-relaxation effects with tuneable additive/subtractive operation, resulting in heterosynaptic plasticity. Compared to metaplastic memristive implementations based on activity priming, this incarnation supports weight updates based on time differences between three signals, facilitating complex learning algorithms involving interactions between local temporal correlations and presence or absence of global neuromodulatory signals, modulating the STDP windows on demand (Figure 2 D, Supporting-Information Note-7 Figure S7 A).[51,52] Implementing such higher order correlations in CMOS requires added circuits in the synapse and periphery.[53,54] However, in the proposed dual-gated neuristor configuration, the second presynaptic terminal created temporally

correlated voltage amplitude differences across the channel in addition to the ones created by the first gate, resulting in multiple correlated terms without aid of any additional peripheral circuitry. The nature and degree of the effective conductance/weight change here depended on temporal correlations between the postsynaptic sequence and two different input sequences (pre-synaptic and modulatory), resulting in higher order correlations that can be used as a supplementary selection node. For example, in standard STDP measurements, operation in the local-gated mode resulted in long-term depression on repeated arrival of the presynaptic spikes a few hundred milliseconds before the postsynaptic action potentials. On activation of the dual-gated mode, this net long-term depression was further modulated by temporal correlations of the additional global modulatory terminal, which added to the depression if repeated arrival of the modulatory spikes occurred a few hundred milliseconds after the postsynaptic action potentials. Repeated arrival of the global modulatory input a few hundred milliseconds before the postsynaptic action potentials created facilitation effects that counter acted the depression due to the first set of pre-post spikes. It should be noted that homeostasis can be obtained by connecting the pre-synaptic spikes to both the gates -- it is thus a subset of the range of possible dynamics achievable by this neuristor.

Classical conditioning experiments also created **association** between the presynaptic inputs as shown in Figure 3.[55,56] Initially, while unconditioned ionically gated stimuli (equivalent to food) activated salivation / unconditioned post-synaptic response (PSC > 12 μA), electronically gated voltage pulses / conditioned stimuli (equivalent to bell) failed to trigger salivary response (PSC < 12 μA, stage: i) in the artificial Pavlov's dog. Persistent simultaneous training resulted in effective association between the two stimuli marked by post-synaptic responses higher than the salivation response threshold with large retention (stages: ii-iii). However, repeated training sequences with the unconditioned stimuli alone resulted in extinction of the association (stage: iv), akin to forgetting in the human brain. But the association could be recovered back by lesser number of retraining sequences, analogous to the biological learning curve (stages: v-vi). Compared to the

associative learning demonstrated between the excitatory photonic and ionic gating modes, this implementation represents a more robust form of association because of the counteracting nature of conductance change observed with the ionic (facilitation) and electronic (depression) gating modes for the same polarity of training sequences.[11] Persistent conditioning of the semiconducting channel with the dominating capacitive coupling of the ionic gating helped overcome this counteracting nature and established robust association between the two gating modes.

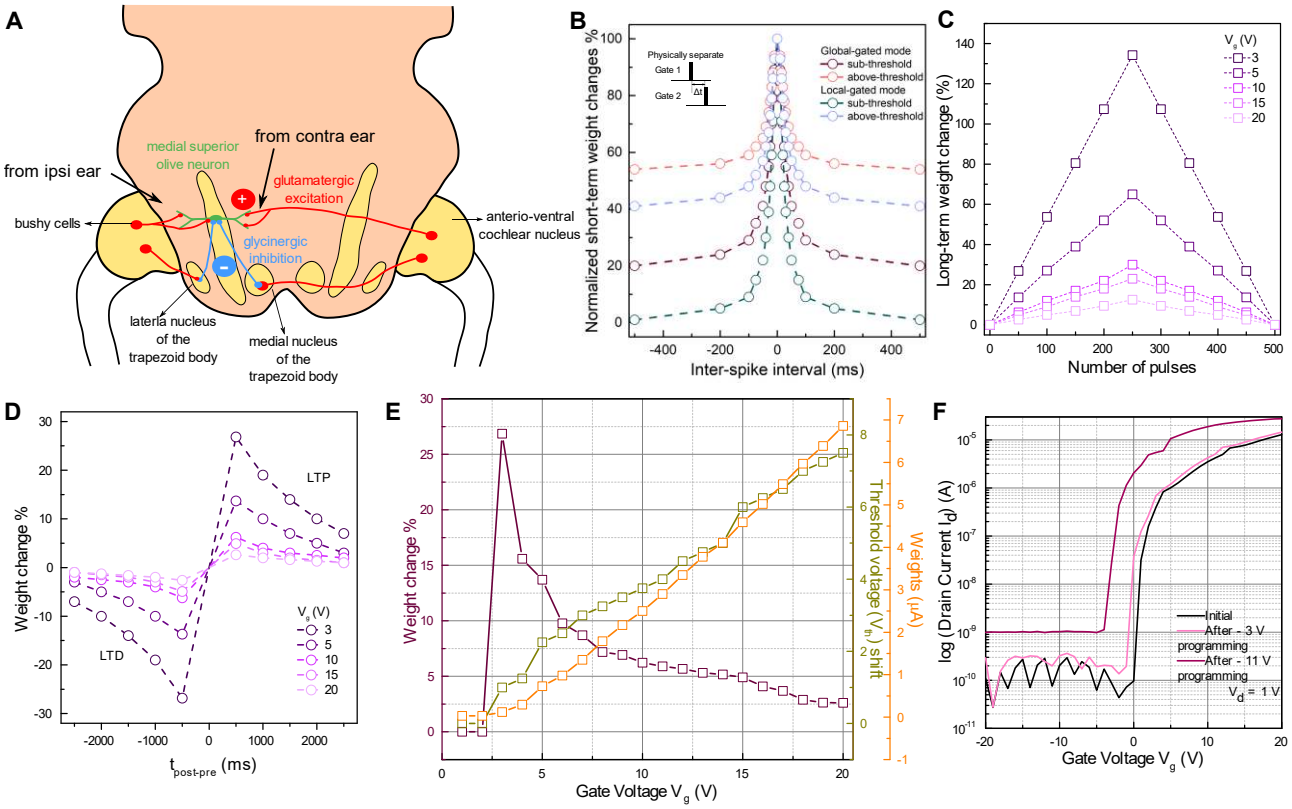

Figure 4. Sub-threshold operation of neuristors enable weight changes with high gain while maintaining ultralow power operation. **(A)** Schematic of the neuronal basis underlying interaural time difference (ITD) based audio signal-processing**, (B**) Input spikes applied with a delay at two physically separate gate terminals, akin to ITD as proposed by Jeffress model, created symmetric weight changes or coincidence. Coincidence detectors operating in the sub-threshold region depicted a wider weight change window, reflecting higher sensitivity. **(C-D)** Long-term weight changes and STDP learning rules programmed in the sub-threshold region yielded higher conductance changes when compared to above-threshold region. **(E-F)** Programming of neuristors create long-term weight changes in channel conductance (weight), reflected as threshold voltage (V_{th}) shifts of neuristors as a function of the programming gate voltage.

Finally, we provide insight into advantages of operating neuristors in the sub-threshold region, where the gate voltages are less than the threshold voltage and the current changes exponentially with change in gate voltage.[57] The versatility of the neuristor is demonstrated by using it to mimic **coincidence detection** properties of auditory brainstem neurons (Figure 4 A).[58,59] Spatial location

of a sound source is estimated by an organism using the interaural time differences (ITD) as proposed by Jeffress model. Though details of the architecture differ between models, the general architecture of tapped delay lines from each ear going to temporal correlators is a common feature. Spike trains with different delays coming from the two ears were presented to two physically separated gate terminals of the neuristor, operating independently in the global and local-gated modes (Figure 4 B, Supporting-Information Note-7 Figure S7 B). Akin to interaural sound localization schemes, symmetric weight-changes were obtained as a function of the delay between the two gates, irrespective of the order of their arrival. The sensitivity of the weight change to changes in delay was 2 to 4 times more when the neuristor was biased in sub-threshold regime as opposed to above-threshold regime due to exponential change in currents. Thus, sub-threshold operation resulted in detectors more sensitive to spatiotemporal correlations with a wider memory window.

The sub-threshold operation also provided higher sensitivity for long-term weight change experiments. Temporally correlated STDP measurements were performed at each step of gate voltage corresponding to the transfer characteristics. Input waveforms were designed in such a way that the reading voltages were equivalent to the minimum voltage level (V_{min}) and the effective writing voltages corresponded to the difference between the maximum and minimum voltage levels (V_{max}-V_{min}) as shown in Figures 4 C-F, Supporting-Information Note-7 Figure S7 C. These temporally correlated waveforms created voltage amplitude differences, resulting in weight changes corresponding to V_{max}-V_{min} while the order of arrival of pre and post-synaptic sequences decided their facilitative/depressive nature. Normalizing the weight changes to the weights resulted in a gate dependent behaviour related to the threshold voltage shift of the neuristors. Clearly, smaller gate voltages corresponding to sub-threshold operation resulted in higher percentage weight changes for the same temporal differences pointing to increased sensitivity. Operation in the global-gated mode was chosen as the model here due to ease of understanding due to hysteresis-free behaviour; however, the conclusions are valid for local-gated mode as well. The cross-over point of the weight-

changes and absolute weights indicated the region at which the neuristor could be biased for a good trade-off between high-sensitivity in weight change as well as high value of absolute weights.

CONCLUSIONS

Previous reports on artificial neuromorphic devices explored the role of liquid electrolytes in demonstrating basic synaptic characteristics like short and long-term plasticity. However, most studies ignore the interplay between local activities and global neuromodulations and hence, fail to emulate complex biologically plausible adaptive functions like heterosynaptic plasticity and homeostasis. In this work, the dual-gated architecture enables emulation of complex neuronal behaviours like heterosynaptic plasticity, homeostasis, association, correlation and coincidence in a single neuristor . The ionotronic mode captures the effect of local activity correlations while the electronic mode represents global neuromodulations, allowing additional modulations, augmented plasticity and higher order temporal correlations at a unitary level. The additional modulatory gate could be biased using homeostatic feedback inputs, applying correction currents to the input of neurons, regulating neural network activity at short-time scales. Extension of this concept to longer time scales results in a homeostatically controlled STDP behaviour. With optimized writing schemes, the dual-gate operation extends the dynamic range of synaptic conductance, increasing the data storage ability of the neuristors. Limited only by the triode region of operation, these neuristors exhibit symmetry in the weight-update operation, significant for cognitive tasks like pattern recognition and classification and online machine learning algorithms. The dual gating configuration also creates temporal cross-correlations between the charge coupling modes, resulting in higher order temporal relations beyond the standard doublet STDP rules- akin to hetersosynaptic plasticity. Compared to other metaplastic implementations where specific bias voltages at the second and third gates modulated the STDP window, this implementation establishes temporal cross correlations between the gating modes emulating metaplasticity in a more biorealistic fashion and resembling complex neuronal models like triplet STDP rules.[11] Classical conditioning

experiments creates associative learning between the gating modes with features like extinction and recovery temporally encoded into the training algorithms. Sub-threshold operation of transistors results in conductance/weight changes with maximum gain and is utilized to create highly sensitive coincidence detectors operating at ultra-low power of the order of femto-Joules. While thermal noise in sub-threshold FETs is higher than its above-threshold counterparts, the signal to noise ratio (SNR) can be improved by summing outputs of many synapses together and scales as square root of N where outputs of N synapses are summed. This is exactly the mode of operation in neural networks and hence sub-threshold operation is a good fit for these kinds of applications where the fundamental operation is a vector-vector multiply.[60] The second issue of capacitive crosstalk is largely mitigated by the three terminal structure of the memristive device and hence sneak paths can be cut-off effectively using floating gates.[61] Compared to previous reports on artificial synapses which focus on only local pre- and postsynaptic activities, such dual-gated neuristors facilitate emulation of more complex biological adaptation mechanisms encompassing local activities and global neuromodulations, paving way for more intelligent computing systems.

EXPERIMENTAL

Solid-state ionic dielectric preparation and characterization: The ionic liquid [EMI][TFSI] was initially dried in vacuum for 24 hours at a temperature of 70 °C. Next, P(VDF-HFP) and [EMI][TFSI] were codissolved in acetone with a weight ratio of 1:4:7. The ion gels were further dried in vacuum at 70 °C for 24 hours to remove the residual solvent, after which it was cut with a razor blade, and then laminated onto the substrate of choice.

Neuristor fabrication and characterization: IWO thin films (thickness ~7 nm) were deposited on SiO_2/Si wafers at room temperature using a RF magnetron sputtering technique with a In_2O_3:WO_3 (a-IWO) (98 : 2 wt %) target at a gas mixing ratio of Ar : O_2 (20 : 1), total chamber pressure of 5 mtorr and RF power of 50W. ITO source and drain contacts (thickness ~100 nm) were then sputter deposited through a shadow mask using a In_2O_3: SnO_2 (90 : 10 wt %) target. The devices were then

annealed at 200°C for 30 minutes in ambient environment for optimized transistor performance. Ion gels were laminated on to these to create dual-gated neuristors. Electrical measurements were carried out in a Desert Cryogenics (Lakeshore) probe station using Keithley 4200-SCS semiconductor characterization system. Capacitance measurements were carried out using an Alpha A Analyzer, Novocontrol analyser over a frequency range of 1 Hz to 10 kHz.

Associated Content

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

Transfer and output characteristics of hysteretic transistors, current-time curves indicating short-term plasticity, comparison of energy consumption with state of the art, scaling of energy consumption with device area, current-time curves indicating long-term plasticity, input waveforms for STDP measurements, LTP-LTD conductance changes depicting advantages of the dual-gated approach, input waveforms for heterosynaptic plasticity, coincidence detection and long-term weight changes as a function of gate voltage.

Author Information

Corresponding Author

Prof. Nripan Mathews (Email: Nripan@ntu.edu.sg)

Prof. Arindam Basu (Email: arindam.basu@ntu.edu.sg)

Notes: The authors declare no competing financial interest.

Acknowledgements. The authors would like to acknowledge the funding from MOE Tier 1 grants RG87/16, RG 166/16 and MOE Tier2 grant MOE2016-T2-1-100. We thank Prof. Koh Soo Jin Adrian (Department of Mechanical Engineering, NUS) for access to dielectric spectroscopy measurements.

References.

(1) Gkoupidenis, P.; Koutsouras, D. A.; Malliaras, G. G. Neuromorphic Device Architectures with Global Connectivity through Electrolyte Gating. *Nat. Commun.* 2017, 8.

(2) Wright, C. D.; Hosseini, P.; Diosdado, J. A. V. Beyond Von-Neumann Computing with Nanoscale Phase-Change Memory Devices. *Adv. Funct. Mater.* 2013, 23, 2248–2254.

(3) Abbott, L. F.; Nelson, S. B. Synaptic Plasticity: Taming the Beast. *Nat. Neurosci.* 2000, 3, 1178–1183.

(4) Zucker, R. S. Short-Term Synaptic Plasticity. *Annu. Rev. Neurosci.* 1989, 12, 13–31.

(5) Blitz, D. M.; Foster, K. A.; Regehr, W. G. Short-Term Synaptic Plasticity: A Comparison of Two Synapses. *Nat. Rev. Neurosci.* 2004, 5, 630–640.

(6) Bliss, T. V; Collingridge, G. L. A Synaptic Model of Memory: Long-Term Potentiation in the Hippocampus. *Nature* 1993, 361, 31.

(7) Serrano-Gotarredona, T.; Masquelier, T.; Prodromakis, T.; Indiveri, G.; Linares-Barranco, B. STDP and STDP Variations with Memristors for Spiking Neuromorphic Learning Systems. *Front. Neurosci.* 2013, 7, 2.

(8) Indiveri, G.; Liu, S.-C. Memory and Information Processing in Neuromorphic Systems. *Proc. IEEE* 2015, 103, 1379–1397.

(9) Wang, Z.; Joshi, S.; Savel'ev, S. E.; Jiang, H.; Midya, R.; Lin, P.; Hu, M.; Ge, N.; Strachan, J. P.; Li, Z. Memristors with Diffusive Dynamics as Synaptic Emulators for Neuromorphic Computing. *Nat. Mater.* 2017, 16, 101–108.

(10) Ohno, T.; Hasegawa, T.; Tsuruoka, T.; Terabe, K.; Gimzewski, J. K.; Aono, M. Short-Term Plasticity and Long-Term Potentiation Mimicked in Single Inorganic Synapses. *Nat. Mater.* 2011, 10, 591–595.

(11) John, R. A.; Liu, F.; Chien, N. A.; Kulkarni, M. R.; Zhu, C.; Fu, Q.; Basu, A.; Liu, Z.; Mathews, N. Synergistic Gating of Electro-Iono-Photoactive 2D Chalcogenide Neuristors: Coexistence of Hebbian and Homeostatic Synaptic Metaplasticity. *Adv. Mater.* 2018, 1800220.

(12) Fortunato, E.; Barquinha, P.; Martins, R. Oxide Semiconductor Thin-film Transistors: A Review of Recent Advances. *Adv. Mater.* 2012, 24, 2945–2986.

(13) Nomura, K.; Ohta, H.; Takagi, A.; Kamiya, T.; Hirano, M.; Hosono, H. Room-Temperature Fabrication of Transparent Flexible Thin-Film Transistors Using Amorphous Oxide Semiconductors. *Nature* 2004, 432, 488–492.

(14) Shi, J.; Ha, S. D.; Zhou, Y.; Schoofs, F.; Ramanathan, S. A Correlated Nickelate Synaptic Transistor. *Nat. Commun.* 2013, 4.

(15) Yang, J.; Ge, C.; Du, J.; Huang, H.; He, M.; Wang, C.; Lu, H.; Yang, G.; Jin, K. Artificial Synapses Emulated by an Electrolyte-Gated Tungsten-Oxide Transistor. *Adv. Mater.* 2018, 1801548.

(16) Lee, K. H.; Kang, M. S.; Zhang, S.; Gu, Y.; Lodge, T. P.; Frisbie, C. D. "Cut and Stick" Rubbery Ion Gels as High Capacitance Gate Dielectrics. *Adv. Mater.* 2012, 24, 4457–4462.

(17) Frémaux, N.; Gerstner, W. Neuromodulated Spike-Timing-Dependent Plasticity, and Theory of Three-Factor Learning Rules. *Front. Neural Circuits* 2016, 9, 85.

(18) Gerstner, W.; Lehmann, M.; Liakoni, V.; Corneil, D.; Brea, J. Eligibility Traces and Plasticity on Behavioral Time Scales: Experimental Support of NeoHebbian Three-Factor Learning Rules. *arXiv Prepr.* arXiv1801.05219 2018.

(19) Rubin, J. E.; Hayes, J. A.; Mendenhall, J. L.; Del Negro, C. A. Calcium-Activated Nonspecific Cation Current and Synaptic Depression Promote Network-Dependent Burst Oscillations. *Proc. Natl. Acad. Sci.* 2009, 106, 2939–2944.

(20) Greengard, P. The Neurobiology of Slow Synaptic Transmission. *Science*. 2001, 294, 1024–1030.

(21) Loewenstein, Y.; Seung, H. S. Operant Matching Is a Generic Outcome of Synaptic Plasticity Based on the Covariance between Reward and Neural Activity. *Proc. Natl. Acad.*

Sci. 2006, 103, 15224–15229.
(22) Shen, A. M.; Chen, C.-L.; Kim, K.; Cho, B.; Tudor, A.; Chen, Y. Analog Neuromorphic Module Based on Carbon Nanotube Synapses. *ACS Nano* 2013, 7, 6117–6122.
(23) Castro-Alamancos, M. A.; Connors, B. W. Short-Term Synaptic Enhancement and Long-Term Potentiation in Neocortex. *Proc. Natl. Acad. Sci.* 1996, 93, 1335–1339.
(24) Perea, G.; Araque, A. Astrocytes Potentiate Transmitter Release at Single Hippocampal Synapses. *Science*. 2007, 317, 1083–1086.
(25) Fortune, E. S.; Rose, G. J. Short-Term Synaptic Plasticity as a Temporal Filter. *Trends Neurosci.* 2001, 24, 381–385.
(26) John, R. A.; Ko, J.; Kulkarni, M. R.; Tiwari, N.; Chien, N. A.; Geok, N.; Leong, W. L.; Mathews, N. Flexible Ionic-Electronic Hybrid Oxide Synaptic TFTs with Programmable Dynamic Plasticity for Brain-Inspired Neuromorphic Computing. *Small* 2017, 13.
(27) Daoudal, G.; Debanne, D. Long-Term Plasticity of Intrinsic Excitability: Learning Rules and Mechanisms. *Learn. Mem.* 2003, 10, 456–465.
(28) Arnold, A. J.; Razavieh, A.; Nasr, J. R.; Schulman, D. S.; Eichfeld, C. M.; Das, S. Mimicking Neurotransmitter Release in Chemical Synapses *via* Hysteresis Engineering in MoS2 Transistors. *ACS Nano* 2017, 11, 3110–3118.
(29) Song, S.; Miller, K. D.; Abbott, L. F. Competitive Hebbian Learning through Spike-Timing-Dependent Synaptic Plasticity. *Nat. Neurosci.* 2000, 3, 919–926.
(30) Tian, H.; Guo, Q.; Xie, Y.; Zhao, H.; Li, C.; Cha, J. J.; Xia, F.; Wang, H. Anisotropic Black Phosphorus Synaptic Device for Neuromorphic Applications. *Adv. Mater.* 2016, 28, 4991–4997.
(31) Letzkus, J. J.; Kampa, B. M.; Stuart, G. J. Learning Rules for Spike Timing-Dependent Plasticity Depend on Dendritic Synapse Location. *J. Neurosci.* 2006, 26, 10420–10429.
(32) Hussain, S.; Liu, S.-C.; Basu, A. Hardware-Amenable Structural Learning for Spike-Based Pattern Classification Using a Simple Model of Active Dendrites. *Neural Comput.* 2015, 27, 845–897.
(33) Albers, C.; Westkott, M.; Pawelzik, K. Perfect Associative Learning with Spike-Timing-Dependent Plasticity. *Advances in Neural Information Processing Systems*; 2013; pp 1709–1717.
(34) Indiveri, G.; Linares-Barranco, B.; Hamilton, T. J.; Van Schaik, A.; Etienne-Cummings, R.; Delbruck, T.; Liu, S.-C.; Dudek, P.; Häfliger, P.; Renaud, S. Neuromorphic Silicon Neuron Circuits. *Front. Neurosci.* 2011, 5, 73.
(35) Tankimanova, A.; Maan, A. K.; James, A. P. Level-Shifted Neural Encoded Analog-to-Digital Converter. *arXiv Prepr.* arXiv1801.00448 2018.
(36) Diehl, P. U.; Cook, M. Unsupervised Learning of Digit Recognition Using Spike-Timing-Dependent Plasticity. *Front. Comput. Neurosci.* 2015, 9.
(37) Bartolozzi, C.; Indiveri, G. Global Scaling of Synaptic Efficacy: Homeostasis in Silicon Synapses. *Neurocomputing* 2009, 72, 726–731.
(38) Leydecker, T.; Herder, M.; Pavlica, E.; Bratina, G.; Hecht, S.; Orgiu, E.; Samorì, P. Flexible Non-Volatile Optical Memory Thin-Film Transistor Device with over 256 Distinct Levels Based on an Organic Bicomponent Blend. *Nat. Nanotechnol.* 2016, 11, 769–775.
(39) Sung, C.; Lim, S.; Kim, H.; Kim, T.; Moon, K.; Song, J.; Kim, J.-J.; Hwang, H. Effect of Conductance Linearity and Multi-Level Cell Characteristics of TaO x-Based Synapse Device on Pattern Recognition Accuracy of Neuromorphic System. *Nanotechnology* 2018.
(40) Frascaroli, J.; Brivio, S.; Covi, E.; Spiga, S. Evidence of Soft Bound Behaviour in Analogue Memristive Devices for Neuromorphic Computing. *Sci. Rep.* 2018, 8, 7178.
(41) Fuller, E. J.; Gabaly, F. El; Léonard, F.; Agarwal, S.; Plimpton, S. J.; Jacobs-Gedrim, R. B.; James, C. D.; Marinella, M. J.; Talin, A. A. Li-ion Synaptic Transistor for Low Power Analog Computing. *Adv. Mater.* 2017, 29.
(42) Keene, S. T.; Melianas, A.; Fuller, E. J.; Van De Burgt, Y.; Talin, A. A.; Salleo, A. Optimized Pulsed Write Schemes Improve Linearity and Write Speed for Low-Power

Organic Neuromorphic Devices. *J. Phys. D. Appl. Phys.* 2018, 51, 224002.
(43) Pozo, K.; Goda, Y. Unraveling Mechanisms of Homeostatic Synaptic Plasticity. *Neuron* 2010, 66, 337–351.
(44) Turrigiano, G. G.; Nelson, S. B. Homeostatic Plasticity in the Developing Nervous System. *Nat. Rev. Neurosci.* 2004, 5, 97–107.
(45) Pawlak, V.; Wickens, J. R.; Kirkwood, A.; Kerr, J. N. D. Timing Is Not Everything: Neuromodulation Opens the STDP Gate. *Front. Synaptic Neurosci.* 2010, 2, 146.
(46) Abraham, W. C.; Bear, M. F. Metaplasticity: The Plasticity of Synaptic Plasticity. *Trends Neurosci.* 1996, 19, 126–130.
(47) Xie, X.; Seung, H. S. Learning in Neural Networks by Reinforcement of Irregular Spiking. *Phys. Rev. E* 2004, 69, 41909.
(48) Legenstein, R.; Pecevski, D.; Maass, W. A Learning Theory for Reward-Modulated Spike-Timing-Dependent Plasticity with Application to Biofeedback. *PLoS Comput. Biol.* 2008, 4, e1000180.
(49) Vasilaki, E.; Frémaux, N.; Urbanczik, R.; Senn, W.; Gerstner, W. Spike-Based Reinforcement Learning in Continuous State and Action Space: When Policy Gradient Methods Fail. *PLoS Comput. Biol.* 2009, 5, e1000586.
(50) Frémaux, N.; Sprekeler, H.; Gerstner, W. Reinforcement Learning Using a Continuous Time Actor-Critic Framework with Spiking Neurons. *PLoS Comput. Biol.* 2013, 9, e1003024.
(51) Zhu, X.; Du, C.; Jeong, Y.; Lu, W. D. Emulation of Synaptic Metaplasticity in Memristors. *Nanoscale* 2017, 9, 45–51.
(52) Tan, Z.; Yang, R.; Terabe, K.; Yin, X.; Zhang, X.; Guo, X. Synaptic Metaplasticity Realized in Oxide Memristive Devices. *Adv. Mater.* 2016, 28, 377–384.
(53) Huayaney, F. L. M.; Nease, S.; Chicca, E. Learning in Silicon beyond STDP: A Neuromorphic Implementation of Multi-Factor Synaptic Plasticity with Calcium-Based Dynamics. *IEEE Trans. Circuits Syst. I Regul. Pap.* 2016, 63, 2189–2199.
(54) Gopalakrishnan, R.; Basu, A. Triplet Spike Time-Dependent Plasticity in a Floating-Gate Synapse. *IEEE Trans. neural networks Learn. Syst.* 2017, 28, 778–790.
(55) Pavlov, I. P.; Anrep, G. V. Conditioned Reflexes; *Courier Corporation*, 2003.
(56) Tan, Z.-H.; Yin, X.-B.; Yang, R.; Mi, S.-B.; Jia, C.-L.; Guo, X. Pavlovian Conditioning Demonstrated with Neuromorphic Memristive Devices. *Sci. Rep.* 2017, 7, 713.
(57) Tsividis, Y.; McAndrew, C. Operation and Modeling of the MOS Transistor; *Oxford Univ. Press*, 2011.
(58) Agmon-Snir, H.; Carr, C. E.; Rinzel, J. The Role of Dendrites in Auditory Coincidence Detection. *Nature* 1998, 393, 268.
(59) Joris, P.; Yin, T. C. T. A Matter of Time: Internal Delays in Binaural Processing. *Trends Neurosci.* 2007, 30, 70–78.
(60) Agarwal, S.; Quach, T.-T.; Parekh, O.; Hsia, A. H.; DeBenedictis, E. P.; James, C. D.; Marinella, M. J.; Aimone, J. B. Energy Scaling Advantages of Resistive Memory Crossbar Based Computation and Its Application to Sparse Coding. *Front. Neurosci.* 2016, 9, 484.
(61) Brink, S.; Nease, S.; Hasler, P.; Ramakrishnan, S.; Wunderlich, R.; Basu, A.; Degnan, B. A Learning-Enabled Neuron Array IC Based upon Transistor Channel Models of Biological Phenomena. *IEEE Trans. Biomed. Circuits Syst.* 2013, 7, 71–81.

Ultralow Power Dual Gated Sub-Threshold Oxide Neuristors: An Enabler For Higher Order Neuronal Temporal Correlations

Rohit Abraham John[1], Nidhi Tiwari[2], Chen Yaoyi[1], Ankit[1], Naveen Tiwari[1], Mohit Kulkarni[1], Amoolya Nirmal[1], Anh Chien Nguyen[1], Arindam Basu[3], Nripan Mathews[1,2]**

[1] School of Materials Science and Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

[2] Energy Research Institute @ NTU (ERI@N), Nanyang Technological University, Singapore 637553

[3] School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

* Corresponding author

Prof. Arindam Basu (Email: arindam.basu@ntu.edu.sg)

Prof. Nripan Mathews (Email: Nripan@ntu.edu.sg)

Keywords: synaptic transistors, neuromorphic computing, heterosynaptic plasticity, homeostasis, classical conditioning, coincidence detection

Note 1: Transistor Characteristics

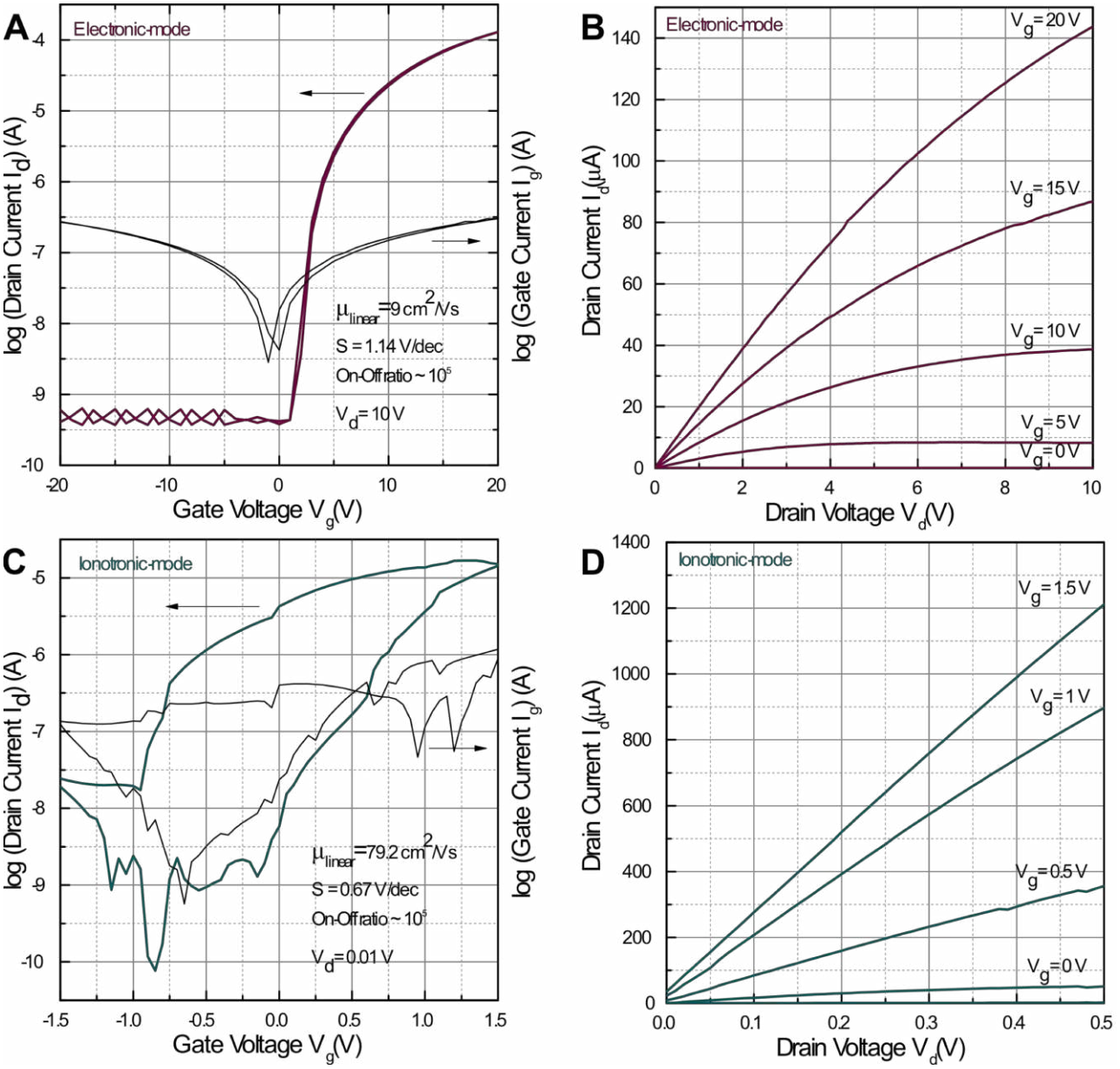

Figure S1. Transfer and Output characteristics of IWO TFTs in (**A-B**) electronic global-gated mode (**C-D**) ionotronic local-gated mode. While SiO_2 serves as the back-gate dielectric in the global-gated mode, ion gel serves as the top-gate dielectric in the local-gated mode.

The linear field-effect mobilities were extracted from the transfer characteristics using the equation $\mu = \frac{dI_d}{dV_g} \times \frac{L}{WC_iV_d}$, where L is the channel length (300 μm), W is the channel width (1000 μm) and C_i is the gate dielectric capacitance. The mobility was calculated to be around 9 $cm^2V^{-1}s^{-1}$ for the global-gated mode, and 79 $cm^2V^{-1}s^{-1}$ for the local-gated mode. The large electrical-double-layer

(EDL) capacitance resulted in efficient carrier accumulation within the small electrochemical window of the ion-gels, resulting in ultralow voltage and power operation.

Devices in the global-gated mode clearly depicted a negligible clockwise hysteresis, while local-gated mode exhibited a larger anticlockwise hysteresis window of 1 V. Ultrafast interfacial trapping and detrapping of electrons in the forward and reverse voltage scans due to surface adsorbates or presence of defects at the semiconductor-dielectric interface explained for the small clockwise hysteresis in the global-gated mode.[1] In the local-gated mode, the anticlockwise hysteresis was attributed to migration-relaxation kinetics of mobile ions in the dielectric, which in turn doped and de-doped the channel based on the direction of voltage scan.[2] Forward scan resulted in migration of cations towards the ion-gel – semiconductor interface, accumulating electrons in the IWO layer to form a conductive channel. This resulted in lower voltage requirements for the subsequent channel formation in the reverse scan, accounting for the anticlockwise hysteresis.

Note 2: Short-term Plasticity

Paired-pulse facilitation (PPF) refers to a short-term form of homosynaptic facilitation in which the postsynaptic response to the second action potential is increased dramatically relative to the first due to the accumulation of residual Ca^{2+} in the presynaptic terminal from the two pulses.[3] This increase in EPSCs extends beyond the two presynaptic activations leading to synaptic facilitative responses to high-frequency stimulation as seen in Figures 1 B-D. An indication of low initial probability of neurotransmitter release, the PPF behaviour is believed to underlie information processing in the human brain. The degree of facilitation is greatest when the Ca^{2+} ions are not allowed to return to the baseline concentration prior to the second stimulus, that is, when the pulse interval is kept shortest.[4] Analogous to this, action potentials separated by minute pulse intervals (< 50 ms), triggered higher excitatory post synaptic currents (EPSCs) in the second presynaptic spike, resulting in paired pulse facilitation (PPF) indices well above 100 %. While the global-gated mode

of operation exhibited PPF indices of ~138%, the local-gated mode depicted much higher indices of ~181% for a pulse interval of 10 ms. Increasing intervals resulted in an exponential reduction of the facilitation indices in accordance with the Ca^{2+} residual hypothesis (Figures S2 A-C Table T1). Higher facilitation indices and higher retention of the local-gated mode indicated larger residual carriers in the channel due to slower ion relaxation mechanism when compared to ultrafast detrapping mechanisms depleting the channel in the global-gated mode.

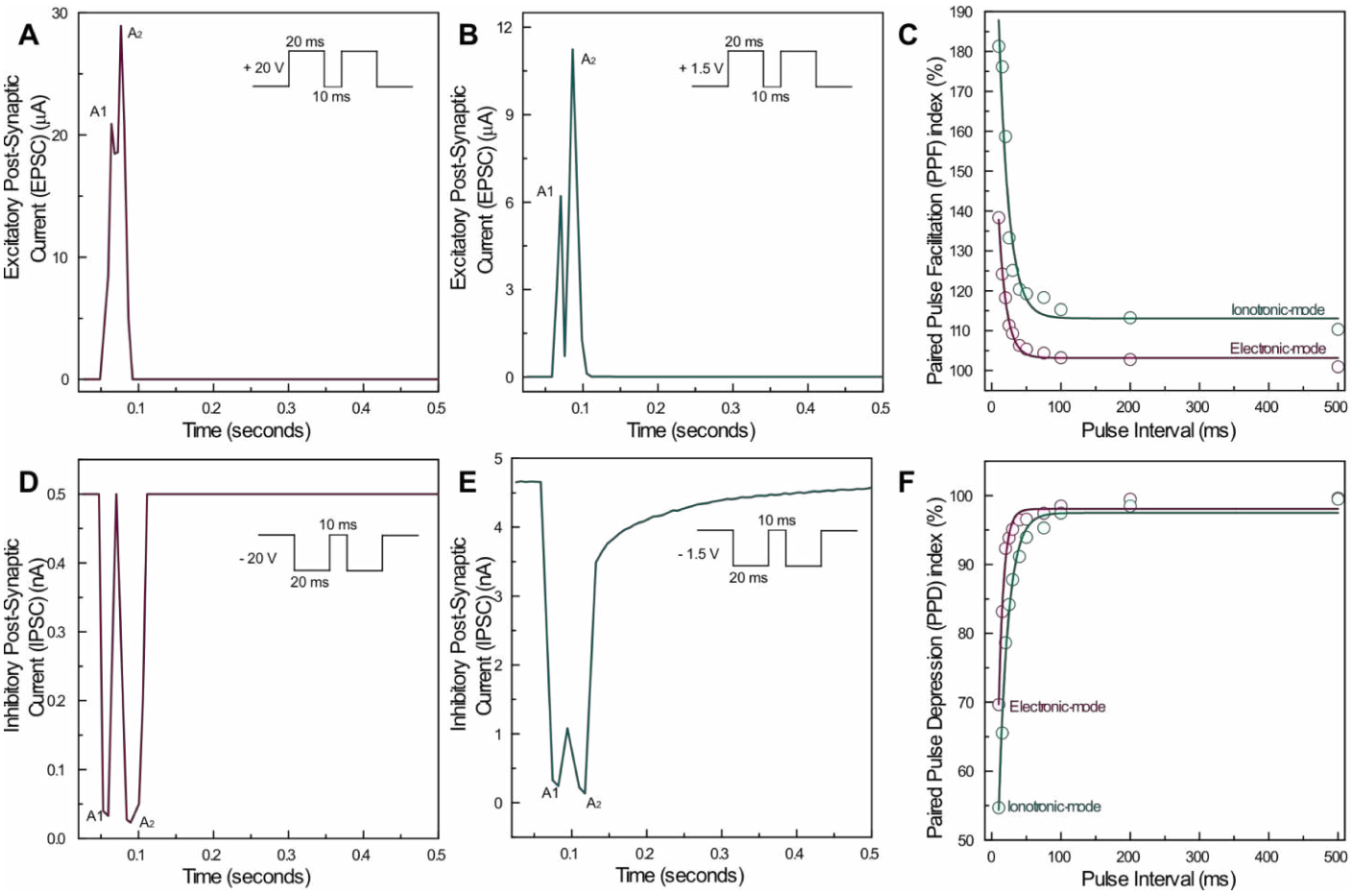

Figure S2. A pair of presynaptic action potentials (pulse width = 20 ms, interval = 10 ms, input waveforms are shown as inset) triggered a pair of excitatory postsynaptic currents (EPSCs) with increasing amplitude due to hysteresis. This phenomenon known as paired pulse facilitation (PPF) reflects the number of residual carriers during trapping-detrapping mechanisms in the global-gated mode (**A**) and ion migration relaxation kinetics in the local-gated mode (**B**). (**C**) PPF index, defined as $[PPF = \left(\frac{A_2}{A_1}\right) * 100\%)]$ is plotted as a function of inter-spike interval to demonstrate the decay process. It quantifies the degree of facilitation and indicates the vesicular release probability, in agreement with the residual Ca^{2+} hypothesis by Katz and Miledi.[4] (**D-F**) Reversal of polarity of presynaptic action potentials result in short-term depression or paired pulse depression, with the indices dependent on pulse width and interval of the presynaptic action potentials, similar to facilitation.

In resemblance to the coupling of biological neurons, PPF variation with pulse interval fitted well with an exponential decay function as shown below.

$$y = B_1 * \exp\left(-\frac{x}{t_1}\right) + y_0$$

where x is pulse interval time, y_0 is resting facilitation magnitude, B_1 is facilitation constants, and t_1 is characteristic time constants. Detailed comparison of the decay time constants is presented in table T1 below.

Configuration	y_0 (facilitation/depression)	B_1 (facilitation/depression)	t_1 (facilitation/depression)	R-Square (facilitation/depression)
Electronic	103.15/98.06	84.22/-109.32	11.27/-7.41	99.11/98.49
Ionotronic	113.1/97.49	147.87/-89.59	14.66/-13.67	94.98/98.98

Table T1. Best fit values of PPF decay as a function of pulse interval.

On activation with presynaptic pulses of opposite polarity, a decrease in short-term conductance was observed, equivalent to **synaptic depression** (Figures S2 D-F). An indication of high vesicular release probability, this stems from inactivation of voltage-gated Ca^{2+} channels or decreased pool of releasable synaptic vesicles.[5,6] The local-gated mode depicted higher depression reflected by lower values of depression indices ~55%, when compared to the global-gated mode ~70%, again indicating the slower relaxation of ions in this mode.

When pulsed with positive gate voltages for short time durations like in the case of PPF, the increase in the 2nd EPSC possibly indicates carrier trapping at the semiconductor-dielectric interface on removal of the pulse, i.e. during the pulse interval; which then detraps/releases upon subsequent pulsing and adds on to the current response to the 2nd voltage pulse resulting in a higher 2nd EPSC or PPF. Shorter intervals do not allow sufficient time for driving this process to completion and hence, the 1st EPSC is not allowed to return back to its original state. When the 2nd voltage spike is applied before complete relaxation of the 1st EPSC, the subsequent detrapping process adds on carriers to the 2nd current response, resulting in a higher 2nd EPSC and hence PPF. On reversing the gate polarity, the same phenomenon happens in the opposite direction, resulting in a lower 2nd EPSC and hence, paired pulse depression. This behaviour has been observed before in conventional

thin film transistor configurations, but on a longer time scale.[5,7] The depression behaviour in our case is observed at both short and long time scales probably because of the nature and number of available trap states at the semiconductor-dielectric interface. Similar trapping-detrapping behaviour has been observed before in 2D TMDC based conventional transistor configurations.[1,8]

Note 3: Energy Consumption

Reference	[9]Yang, Chuan Sen, et al. *Advanced Materials* 29.27 (2017): 1700906.	[10]Yang, Jing-Ting et al, *Advanced Materials* (2018): 1801548.	[11]Xu, Wentao, et al. *Nano Energy* 48 (2018): 575-581.	[12]Zhu, Jiadi, et al. *Advanced Materials* 30 .21 (2018): 1800195.	[13]Sun, Linfeng, et al. *Nano letters* 18.5 (2018): 3229-3234.	This work
Device	MoO_3 transistor	WO_3 transistor	Conjugated polyelectrolyte based memristor	WSe_2 transistor	MoS_2 device with metal-insulator transition	IWO transistor
Energy Consumption per spike (fJ)	9600	36000	10	30	72	9.3

Table T2. Comparison of Energy Consumption with state of the art.

E_{on} was calculated from the equation $E_{on} = I_{peak,first} \times t \times V_{ds}$; where I_{peak} is the peak value of the 1st generated EPSC for a single spike event, t is the spike duration, and V_{ds} is the applied drain voltage.[14] Scaling down of device dimensions and reducing spike durations to sub-millisecond levels could be utilized as strategies to decrease the energy consumption per event even further (Figure S3).

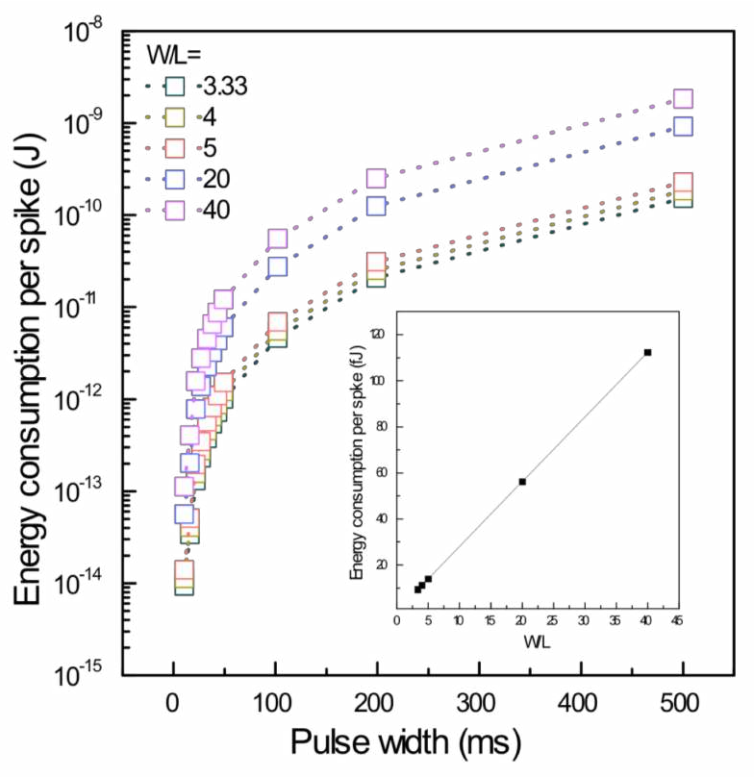

Figure S3. Variation of energy consumption with device scaling

Note 4: Long-term Plasticity

Long-term synaptic plasticity is a fundamental property of the nervous system, defined by long-lasting, activity-dependent changes in synaptic efficacy. Widely considered a primary mechanism for learning and memory, long-term plasticity can bi-directionally modify synaptic strength—either potentiation or depression.[15] Long-term synaptic plasticity was mimicked in our neuristors *via* persistent training sequences lasting up to several seconds, creating permanent changes in channel-conductance. Figure S4 depicts the long-term weight changes associated with the two independent modes of operation under different polarities of training. While persistent pulsing at – 20 V activated long-term potentiation in the global-gated mode, + 1.5 V activated long-term potentiation in the local-gated mode. Application of 10 facilitating pulses of pulse width 500 ms resulted in a larger weight change of 57 % for the local-gated mode when compared to 18 % in the global-gated mode.

For long-term weight analysis the following approach was used in the manuscript throughout.

1. Read the channel conductance with V_{ds} = 0.1 V.

2. Apply necessary voltage waveforms to induce non-volatile weight change/ long-term plasticity.

3. Monitor the channel conductance 30 minutes after application of the waveforms and compare it to the initial conductance state to calculate weight %.

Presynaptic polarities of opposite polarity reversed the direction of conductance change as expected, with the local-gated mode depicting a larger depression when compared to the global-gated mode. In the global-gated mode of operation, non-volatile changes in conductance indicated permanent trapping and detrapping of electrons at the semiconductor-dielectric interface upon persistent pulsing/stressing at the gate terminal. On the other hand, permanent pinning of ions in the ion gel or creation of additional oxygen vacancies accounted for larger non-volatile changes in channel conductance and hence higher weight changes in the local-gated mode. Such time-dependent channel hysteretic behaviour due to charge-trapping dynamics has been previously observed in transistors *via* current transient measurements.[1,16]

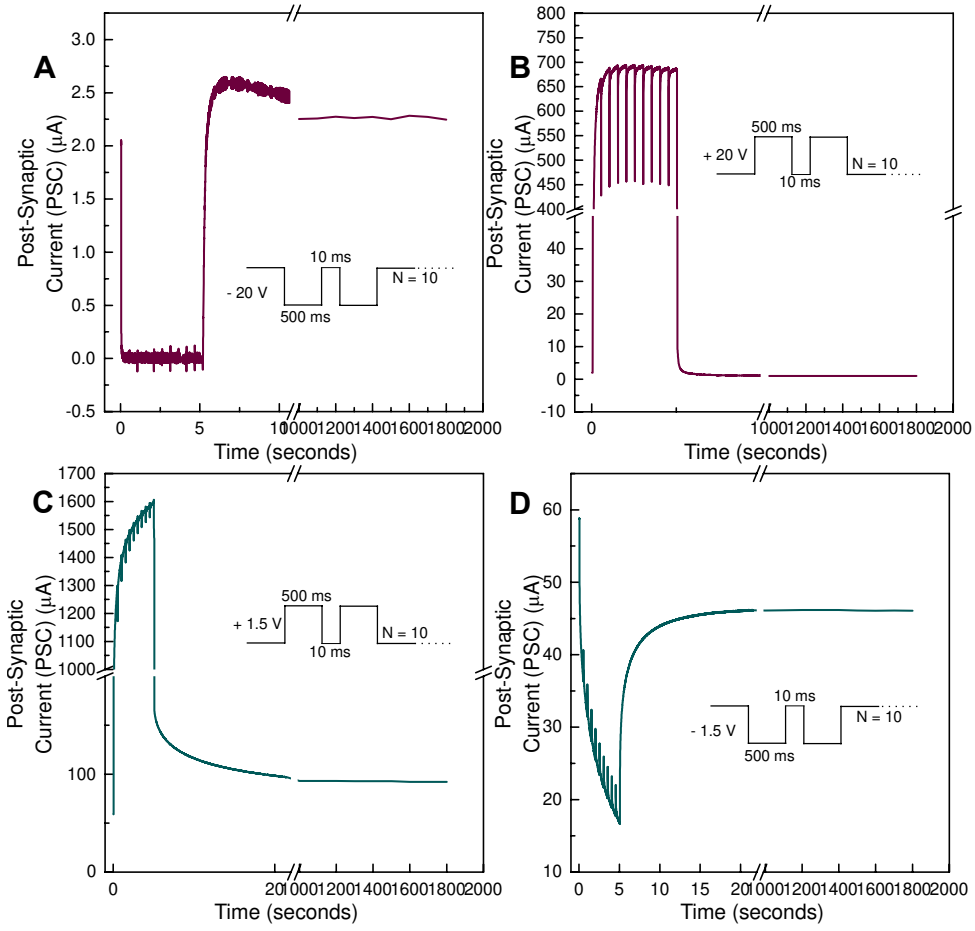

Figure S4. Representative Long-term potentiation (LTP) (**A**) and depression (LTD) (**B**) in global-gated mode. LTP (**C**) and LTD (**D**) in the local-gated mode. The devices were pre-programmed to a common initial conductance state prior to these measurements for a fair analysis.

Note 5: Spike-timing dependent plasticity (STDP)

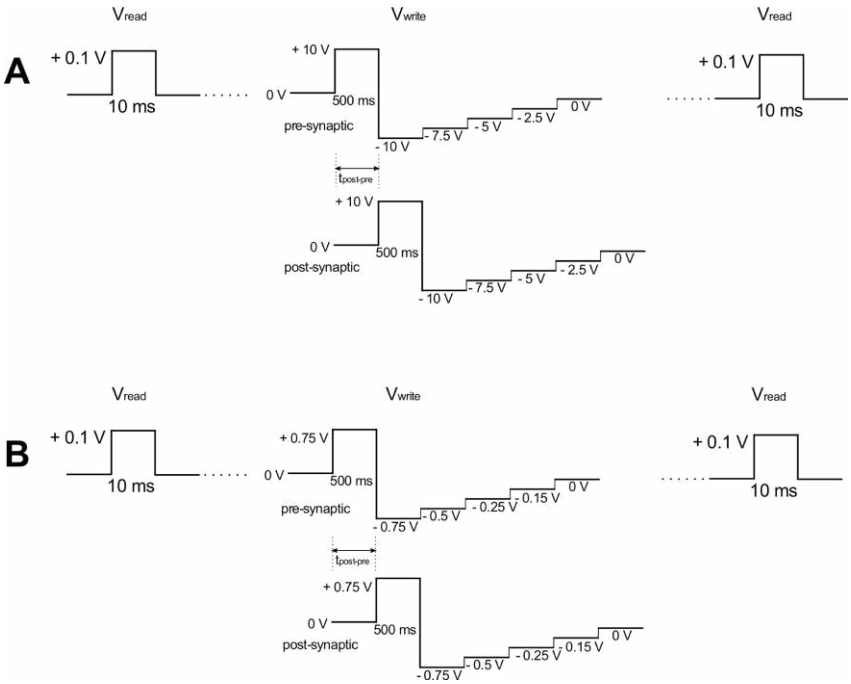

Figure S5. Input waveforms used for STDP measurements in the (**A**) global-gated and (**B**) local-gated modes respectively. A read pulse of + 0.1 V is applied at the drain terminal to read the memconductance states before and after the STDP write operations.

Spike-timing-dependent plasticity (STDP): A refinement of Hebb's theory, STDP is considered to be the first law of synaptic plasticity and forms the basis of associative learning.[17,18] In pristine excitatory connections, precedence of presynaptic action potentials results in LTP whereas presynaptic activity following postsynaptic spikes causes long-term depression (LTD) [Antisymmetric Hebbian]. For excitatory to inhibitory connections, LTP and LTD can be induced in the opposite manner [Antisymmetric anti-Hebbian], while in neocortex and neuromuscular junctions, the order does not play a role [Symmetric anti-Hebbian and Hebbian].[17] The precise relative timing of pre- and postsynaptic spikes significantly affects the sign and magnitude of long-term synaptic modification. Portraying a variety of functional consequences in neural information processing, these different STDP forms reflect the complexity of the underlying cellular mechanisms. Initially, the channel conductance was read by a reading spike (V_{read} = + 0.1 V, 10 ms). Spike patterns (V_{write}) corresponding to Figure S5 were then applied and the change in conductance/weight was recorded as a function of the pulse interval between pre- and postsynaptic

spikes. The timing difference created effect writing voltages ($f(V_{pre}\text{-}V_{post}),\ t$) across the device, which when crossed the threshold voltage, created long-term weight changes in the channel. The resultant conductance change was finally read again with the V_{read} pulse. Weight changes were predominant at small pulse intervals, and weakened with increase in the interval, reflecting strong temporal correlations between the pre- and postsynaptic spikes. The STDP time windows shown here in milli-seconds and weight changes are comparable to biological values and could be further tuned by modulating the pulse width, number and shape of the input spikes.[19]

Note 6: Dual-gated approach enables ultra-fine modulation of weights

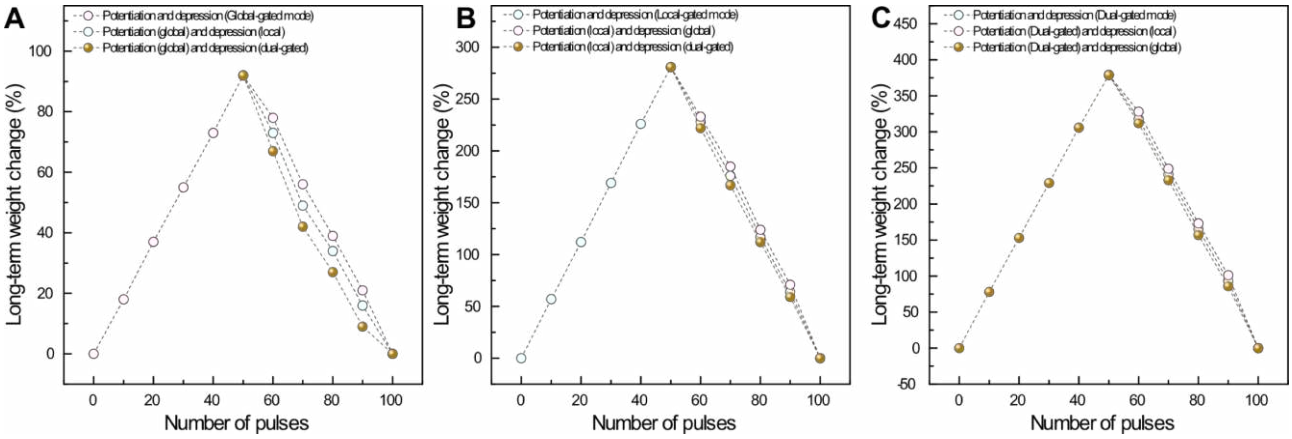

Figure S6. Modulations of strength of LTP-LTD realized *via* multiple gating. (**A**) Long-term potentiation achieved in the global-gated mode is regulated by depression in the global, local and dual-gated modes *via* selective gating. The slopes of depression depend on the operation mode activated to homeostatically stabilize the runaway synaptic potentiation. Long-term potentiation in the local (**B**) and dual-gated modes (**C**) are regulated by depression in the other modes. Potentiating spikes for the global-gated mode involved spikes of amplitude - 20 V and pulse width 500 ms, and + 1.5 V for the local-gated mode. For the dual-gated mode, both global and local presynaptic spikes were applied simultaneously.

The dual-gating approach enabled ultrafine modulation of synaptic weights with additive/subtractive operation. The net capacitive coupling defined the effective weight change and hence, the large electrical-double-layer capacitances in the local-gated mode dominated the global-gated mode at low biases. Additive operation of the two modes increased the slope of facilitation/depression and the memory window. The device could also be operated with one mode defining the facilitation window and the second gate defining the depression window as depicted in Figure S6 above.

Note 7: Input waveforms for heterosynaptic plasticity, coincidence detection and long-term weight changes as a function of gate voltage

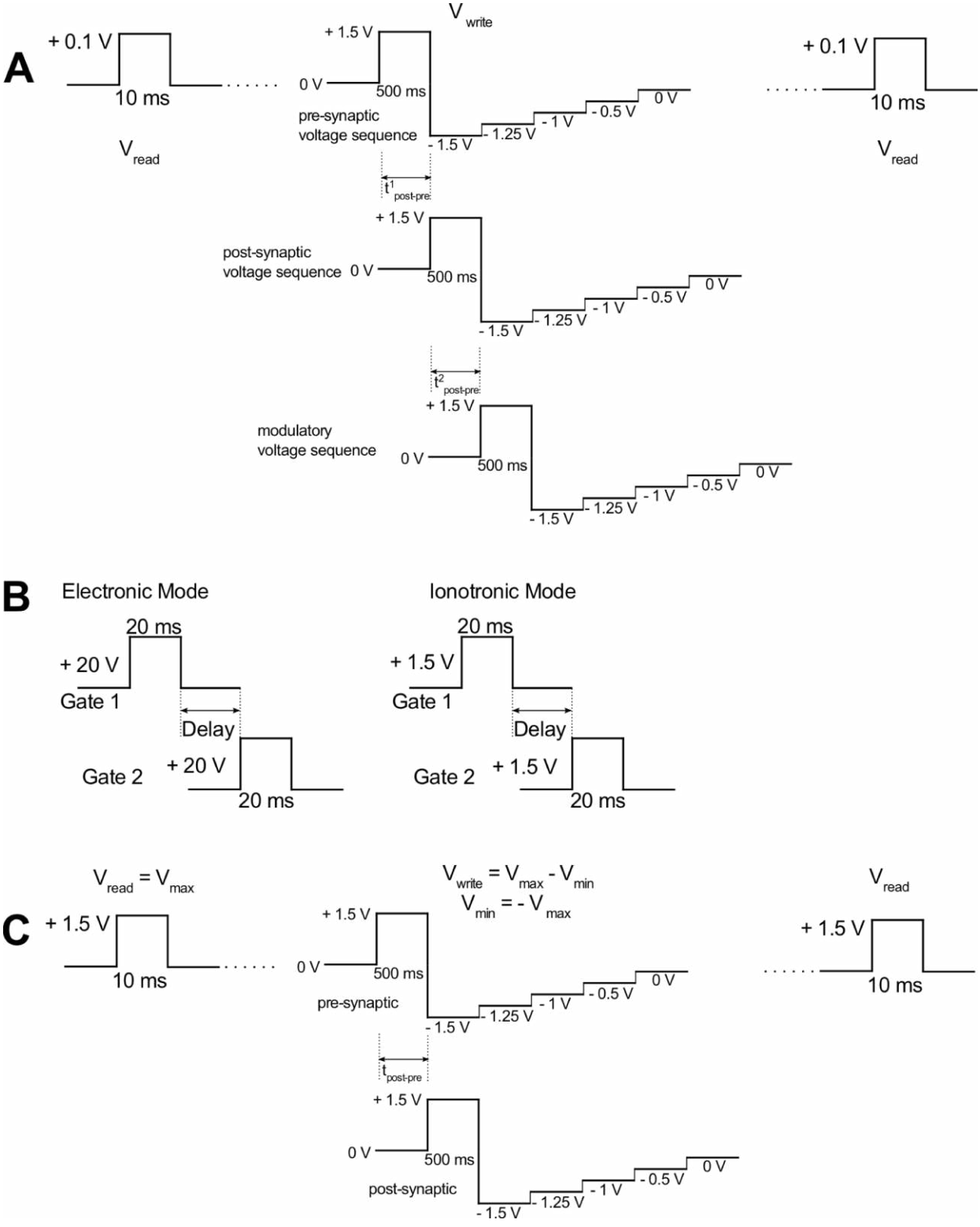

Figure S7. (**A**) Higher order temporal correlations were achieved *via* interaction between two different presynaptic action potentials and the postsynaptic action potential, creating real-time modulations of the standard STDP behaviour. (**B**) Spikes applied at two physically separate local and global gates with a delay created symmetric short-term plasticity behaviour which can be used for coincidence detection. The depicted waveforms represent an example of the above-

threshold region of operation. (**C**) Input waveforms used for STDP measurements at each step of gate voltage corresponding to the transfer characteristics of the neuristor. Weight changes were defined by effective write voltages ($V_{write}=V_{max}-V_{min}$) and read voltages equivalent to V_{min} were used to read the memconductance states.

References.

(1) Tian, H.; Guo, Q.; Xie, Y.; Zhao, H.; Li, C.; Cha, J. J.; Xia, F.; Wang, H. Anisotropic Black Phosphorus Synaptic Device for Neuromorphic Applications. *Adv. Mater.* 2016, 28, 4991–4997.

(2) Zhu, L. Q.; Wan, C. J.; Guo, L. Q.; Shi, Y.; Wan, Q. Artificial Synapse Network on Inorganic Proton Conductor for Neuromorphic Systems. *Nat. Commun.* 2014, 5.

(3) López, J. C. A Fresh Look at Paired-Pulse Facilitation. *Nat. Rev. Neurosci.* 2001, 2, 307.

(4) Katz, B.; Miledi, R. The Role of Calcium in Neuromuscular Facilitation. *J. Physiol.* 1968, 195, 481.

(5) Zang, Y.; Shen, H.; Huang, D.; Di, C.; Zhu, D. A Dual-Organic-Transistor-Based Tactile-Perception System with Signal-Processing Functionality. *Adv. Mater.* 2017, 29.

(6) Rubin, J. E.; Hayes, J. A.; Mendenhall, J. L.; Del Negro, C. A. Calcium-Activated Nonspecific Cation Current and Synaptic Depression Promote Network-Dependent Burst Oscillations. *Proc. Natl. Acad. Sci.* 2009, 106, 2939–2944.

(7) Xu, W.; Min, S.-Y.; Hwang, H.; Lee, T.-W. Organic Core-Sheath Nanowire Artificial Synapses with Femtojoule Energy Consumption. *Sci. Adv.* 2016, 2, e1501326.

(8) Arnold, A. J.; Razavieh, A.; Nasr, J. R.; Schulman, D. S.; Eichfeld, C. M.; Das, S. Mimicking Neurotransmitter Release in Chemical Synapses *via* Hysteresis Engineering in MoS2 Transistors. *ACS Nano* 2017, 11, 3110–3118.

(9) Yang, C. Sen; Shang, D. S.; Liu, N.; Shi, G.; Shen, X.; Yu, R. C.; Li, Y. Q.; Sun, Y. A Synaptic Transistor Based on Quasi-2D Molybdenum Oxide. *Adv. Mater.* 2017, 29, 1700906.

(10) Yang, J.; Ge, C.; Du, J.; Huang, H.; He, M.; Wang, C.; Lu, H.; Yang, G.; Jin, K. Artificial Synapses Emulated by an Electrolyte-Gated Tungsten-Oxide Transistor. *Adv. Mater.* 2018, 1801548.

(11) Xu, W.; Nguyen, T. L.; Kim, Y.-T.; Wolf, C.; Pfattner, R.; Lopez, J.; Chae, B.-G.; Kim, S.-I.; Lee, M. Y.; Shin, E.-Y. Ultrasensitive Artificial Synapse Based on Conjugated Polyelectrolyte. *Nano Energy* 2018, 48, 575–581.

(12) Zhu, J.; Yang, Y.; Jia, R.; Liang, Z.; Zhu, W.; Rehman, Z. U.; Bao, L.; Zhang, X.; Cai, Y.; Song, L. Ion Gated Synaptic Transistors Based on 2D van Der Waals Crystals with Tunable Diffusive Dynamics. *Adv. Mater.* 2018, 30, 1800195.

(13) Sun, L.; Zhang, Y.; Hwang, G.; Jiang, J.; Kim, D.; Eshete, Y. A.; Zhao, R.; Yang, H. Synaptic Computation Enabled by Joule Heating of Single-Layered Semiconductors for Sound Localization. *Nano Lett.* 2018, 18, 3229–3234.

(14) Shen, A. M.; Chen, C.-L.; Kim, K.; Cho, B.; Tudor, A.; Chen, Y. Analog Neuromorphic Module Based on Carbon Nanotube Synapses. *ACS Nano* 2013, 7, 6117–6122.

(15) Bliss, T. V; Collingridge, G. L. A Synaptic Model of Memory: Long-Term Potentiation in the Hippocampus. *Nature* 1993, 361, 31.

(16) Amit, I.; Octon, T. J.; Townsend, N. J.; Reale, F.; Wright, C. D.; Mattevi, C.; Craciun, M. F.; Russo, S. Role of Charge Traps in the Performance of Atomically Thin Transistors. *Adv. Mater.* 2017, 29.

(17) Li, Y.; Zhong, Y.; Xu, L.; Zhang, J.; Xu, X.; Sun, H.; Miao, X. Ultrafast Synaptic Events in a Chalcogenide Memristor. *Sci. Rep.* 2013, 3.

(18) D'Souza, P.; Liu, S.-C.; Hahnloser, R. H. R. Perceptron Learning Rule Derived from Spike-Frequency Adaptation and Spike-Time-Dependent Plasticity. *Proc. Natl. Acad. Sci.* 2010, 107, 4722–4727.

(19) Song, S.; Miller, K. D.; Abbott, L. F. Competitive Hebbian Learning through Spike-Timing-Dependent Synaptic Plasticity. *Nat. Neurosci.* 2000, 3, 919–926.