Synaptic and neuromorphic functions: general discussion

Berg, Sanne; Brivio, Stefano; Brown, Simon; Burr, Geoffrey ; Deswal, Sweety; Deuermeier, Jonas; Gale, Ella; Hwang, Hyunsang ; Ielmini, Daniele ; Indiveri, Giacomo

Total number of authors:
25

Published in:
Faraday Discussions

Link to article, DOI:
10.1039/C8FD90065E

Publication date:
2019

Document Version
Peer reviewed version

Faraday Discussions C8FD90065E

We have presented the graphical abstract image and text for your article below. This briefly summarises your work, and will be presented with your article online. It will not appear in the print edition.

DISCUSSIONS

1

Synaptic and neuromorphic functions: general discussion

Sanne Berg, Stefano Brivio, Simon Brown, Geoffrey Burr, Sweety Deswal, Jonas Deuermeier, Ella Gale, Hyunsang Hwang, Daniele Ielmini, Giacomo Indiveri, Tony Kenyon, Asal Kiazadeh, Itir Köymen, Michael Kozicki, Yang Li, Daniel Mannion, Themis Prodromakis, Carlo Ricciardi, Sebastian Siegel, Maximilian Speckbacher, Ilia Valov, Wei Wang, Stanley Williams, Dirk Wouters and Yuchao Yang

Please check this proof carefully. **Our staff will not read it in detail after you have returned it.**

Proof corrections must be returned as a single set of corrections, approved by all co-authors. No further corrections can be made after you have submitted your proof corrections as we will publish your article online as soon as possible after they are received.

Please ensure that:

- The spelling and format of all author names and affiliations are checked carefully. Names will be indexed and cited as shown on the proof, so these must be correct.
- Any funding bodies have been acknowledged appropriately.
- All of the editor's queries are answered.
- Any necessary attachments, such as updated images or ESI files, are provided.

Translation errors between word-processor files and typesetting systems can occur so the whole proof needs to be read. Please pay particular attention to: tables; equations; numerical data; figures and graphics; and references.

Please send your corrections preferably as a copy of the proof PDF with electronic notes attached or alternatively as a list of corrections – do not change the text within the PDF file or send a revised manuscript. Corrections at this stage should be minor and not involve extensive changes.

Please return your **final** corrections, where possible within **48 hours** of receipt, by e-mail to: faraday@rsc.org. If you require more time, please notify us by email.

Researcher information

If any authors have ORCID or ResearcherID details that are not listed below, please provide these with your proof corrections.

Please check that the ORCID and ResearcherID details listed below have been assigned to the correct author. Authors should have their own unique ORCID iD and should not use another researcher's, as errors will delay publication.

Please also update your account on our online manuscript submission system to add your ORCID details, which will then be automatically included in all future submissions. See here for step-by-step instructions and more information on author identifiers.

First (given) name(s)	Last (family) name(s)	ResearcherID	ORCID
Sanne	Berg		
Stefano	Brivio		
Simon	Brown		
Geoffrey	Burr		
Sweety	Deswal		
Jonas	Deuermeier		
Ella	Gale		
Hyunsang	Hwang		
Daniele	Ielmini		
Giacomo	Indiveri		
Tony	Kenyon		
Asal	Kiazadeh		
Itir	Köymen		
Michael	Kozicki		
Yang	Li		
Daniel	Mannion		
Themis	Prodromakis		
Carlo	Ricciardi		
Sebastian	Siegel		
Maximilian	Speckbacher		
Ilia	Valov		
Wei	Wang		
Stanley	Williams		
Dirk	Wouters		
Yuchao	Yang		

Queries for the attention of the authors

Journal: **Faraday Discussions**

Paper: **C8FD90065E**

Title: **Synaptic and neuromorphic functions: general discussion**

For your information: You can cite this article before you receive notification of the page numbers by using the following format: (authors), Faraday Discuss., (year), DOI: 10.1039/C8FD90065E.

Editor's queries are marked on your proof like this **1**, **2**, etc. and for your convenience line numbers are indicated like this 5, 10, 15,

Please ensure that all queries are answered when returning your proof corrections so that publication of your article is not delayed.

Query Reference	Query	Remarks
1	[INFO-1] Please carefully check the spelling of all author names. This is important for the correct indexing and future citation of your article. No late corrections can be made. FOR YOUR INFORMATION: You can cite this paper before the page numbers are assigned with: (authors), Faraday Discuss., DOI: 10.1039/c8fd90065e.	

Faraday Discussions
Cite this: DOI: 10.1039/C8FD90065E

DISCUSSIONS

Synaptic and neuromorphic functions: general discussion

Sanne Berg, Stefano Brivio, Simon Brown, Geoffrey Burr, Sweety Deswal, Jonas Deuermeier, Ella Gale, Hyunsang Hwang, Daniele Ielmini, Giacomo Indiveri, Tony Kenyon, Asal Kiazadeh, Itir Köymen, Michael Kozicki, Yang Li, Daniel Mannion, Themis Prodromakis, Carlo Ricciardi, Sebastian Siegel, Maximilian Speckbacher, Ilia Valov, Wei Wang, Stanley Williams, Dirk Wouters and Yuchao Yang

DOI: 10.1039/C8FD90065E

(600:[600]600) **Daniel Mannion** opened discussion of the paper by Hyunsang Hwang: This is a question regarding your devices relating to Fig. 21 of your paper (DOI: 10.1039/C8FD00127H). The caption states the potentiation and depression observed in Fig. 21a and 21b respectively are for separate devices with different composition. Is it that each device exhibits only one of the behaviours shown or do the two devices exhibit both potentiation and depression as well? Also, what is the retention time of this potentiation and depression?

Hyunsang Hwang responded: Thank you for your comment. It was a typo. We measured potentiation and depression using a Mo/TiO_x/TiN device. I will revise the manuscript. Although we did not perform an in-depth study on retention characteristics, we cannot guarantee long-term retention using the Mo/TiO_x/TiN device.

(601:[601]601) **Dirk Wouters** asked: Please allow me two questions on your interesting talk: firstly, as it is not obvious to develop an ideal analog device, what do you feel is the potential to use more than one "bipolar digital" switching device for emulating one matrix coefficient? Secondly, you showed the potential of ferroelectric devices. However, what about scaling, when the device dimensions become comparable to those of a single ferroelectric domain?

Hyunsang Hwang replied: Thank you for your comments. In response to your first question, using a bipolar digital device, we can implement a simple binary neural network with limited pattern recognition accuracy and inference-only application. Although the applications are limited, I think there is a market if we could implement a nano-scale X-point array device with only 1-bit storage. Regarding your second question, we have not yet studied the device dimensions

issue of a ferroelectric device. I think there is a significant variation issue with scaling the device dimensions.

(602:[602]602) **Tony Kenyon** commented: Thanks for the nice talk. You mentioned in your requirements for an ideal synapse that linearity is needed. I can see why that's required for applications such as vector-matrix multiplication in which you would need to reliably programme many discrete resistance levels efficiently, but in biology synapses and neurons are very non-linear, and nonlinearity is required for many processes such as habituation. Biology is able to process unstructured data with extreme efficiency, which suggests that complex systems that are very non-linear are potentially very valuable. For neuromorphic behaviour that more closely mimics some of the processes we see in biology, we should maybe be looking at embracing nonlinearity.

(603:[603]603) **Yang Li** asked: Regarding the requirements of an ideal synaptic device, which one do you believe is the most important one? How do you predict these specific requirements and what are your criteria?

Hyunsang Hwang answered: Thank you for your comments. It depends on the application. For on-chip training, symmetric conductance changes under a potentiation/depression pulse condition directly affect the pattern recognition accuracy. To evaluate pattern recognition accuracy, we need to simulate neural networks where each parameter is varied using a standard data set (MNIST, CIFAR).

(604:[604]604) **Geoffrey Burr** remarked: We mentioned linearity in two different contexts. In one case, we discuss a linear change in conductance as a function of the number of pulses. This is important for the backpropagation algorithm, because the algorithm assumes that N small increases of conductance followed by N small decreases of conductance will leave the device in the original conductance state. In the other case, we discuss a linear I–V characteristic (in other words, an "ohmic" device). This is useful if one is encoding data onto a crossbar array using analog voltage, so that read current corresponds to the product of the input "voltage" data times the stored "conductance" data. However, by encoding data into temporal duration and integrating the read current, a non-ohmic device can be used (the read voltage is constant, yet we still get the product of input "duration" data times the stored "conductance" data).

Hyunsang Hwang answered: Thank you for your comments. I agree with your comments. We consider I–V linearity to implement matrix multiplication using an analog input voltage. If we cannot meet ideal I–V linearity, we can integrate the read current with multiple pulses which require additional circuits.

(605:[605]605) **Asal Kiazadeh** commented: Thanks for sharing these interesting results in the Faraday Discussion. Concerning Fig. 5 of your paper (DOI: 10.1039/C8FD00127H), in order to obtain the gradual set, various schemes are introduced and case 'c' was the optimized one. My question is, out of curiosity, have you ever tried to set the device gradually with application of bias at different frequencies,

such as using a lower SET voltage amplitude with shorter intervals between pulses?

Hyunsang Hwang responded: Thank you for your comments. We have not yet studied the frequency dependence of the SET pulse. However, it might be difficult to control SET pulse conditions for a filamentary switching device.

(606:[606]606) **Asal Kiazadeh** said: You made a summary of the general characteristics of synaptic devices studied to date and, in addition, a definition of required components of an ideal synapse. My first question is: I would like to know if there is any reason that you didn't consider second-order memristors, *e.g.* a diffusive memristor or threshold-type resistive switching element, as a synapse device in your summary table? Of course these memristors are volatile, but there are some interesting studies concerning their ability to bio-realistically implement synaptic plasticity. As an example, please have a look at ref. 1 in *Nature Materials*. The second question is: would you think a synapse element should have all of the required elements as a programmable memory (the ones that are mentioned in the definition of an ideal synapse)?

1 Z. Wang *et al.*, Memristors with diffusive dynamics as synaptic emulators for neuromorphic computing, *Nature Materials*, 2017, **16**, 101–108.

Hyunsang Hwang replied: Thank you for your comments. Regarding your first question, I have investigated synapse devices for a HW deep neural network based on a backpropagation algorithm. A synapse device without retention is not a practical device for matrix multiplication. I think memristors with diffusive dynamics might be OK for SNN. In response to your second question, we need all the synapse device requirements for DNN with on-chip training capability. However, we may not need all the requirements for an inference-only application.

(607:[607]607) **Tony Kenyon** addressed Hyunsang Hwang and Geoffrey Burr: Responding to Geoff's comment regarding your slide of requirements for an ideal neuromorphic memristor: it isn't just the nonlinearity of *I*/*V* characteristics that seems to be important in biology, but also the nonlinearity of the conductance *vs.* number of programming pulses. This is how habituation works.

Hyunsang Hwang answered: Thank you for your comment. Since a neural network requires significant power consumption to implement matrix multiplication, I just focused on HW devices' requirements for matrix multiplication. I am not interested in biology and mimicking brain function.

Geoffrey Burr responded: I agree that nonlinearities are important, but in biology these would be implemented by assumably fairly complex attention mechanisms, so that when the neural system (brain) is "paying attention", weight change is stronger and/or more likely than otherwise. I suspect this kind of thing is what permits both strong tolerance to noise and spurious observations as well as one-short learning.

(609:[609]609) **Stefano Brivio** remarked: The authors mention that one of the requirements for electronic synapses is that they undergo a linear conductance change as a function of the number of pulses, and that such a conductance change should be symmetric for conductance increase and decrease. I do agree that the mentioned properties are really required for synapses to be employed in a system for deep neural network acceleration in which operations of conductance increase and decrease must be carefully balanced. On the contrary, in systems with a different kind of learning scheme, this requirement can be relaxed and non-linearity and asymmetry can even be beneficial, as shown in the papers by La Barbera *et al.*[1] who employ an STDP-based network and by Brivio (myself) *et al.*,[2] who employ a learning rule that corresponds to a biologically plausible generalization of STDP.

1 S. La Barbera *et al.*, *Adv. Electron. Mater.*, 2018, **4**, 1800223.
2 S. Brivio *et al.*, *Nanotechnology*, 2018, **30**, 015102.

Hyunsang Hwang responded: Thank you for your comment. I only focused on synapse devices for HW DNN applications. I do agree with your comments on SNN application. Although SNN is a very interesting approach, the application for pattern recognition using SNN is limited. It might be difficult to recognize complex images (for example the CIFAR data set).

(612:[612]612) **Stanley Williams** addressed Hyunsang Hwang and Geoffrey Burr: A quick comment: constraining the weight changes to small increments potentially does away with one-shot learning, a very important thing that humans can do. A three-year-old girl sees a cat and that's a kitty. The next day she sees a tiger and that's also a kitty; she only needs one example to learn and can generalize from that, which is an extremely important learning mode. This is a potential argument in favour of having large weight changes in a training system.

Geoffrey Burr answered: Agreed. Particularly important here will be an "attention" mechanism, so that other parts of the system can gate when large synaptic changes are desirable (one-shot learning) and when they are not (noise).

(613:[613]613) **Giacomo Indiveri** remarked: Concerning the linearity *versus* non-linearity discussion, I fully agree with the argument that, in the context of mixed analog-digital memristive/CMOS computing architectures, embracing non-linearities and non-idealities has a very large potential for developing compact and efficient brain-inspired computing systems. Concerning the differences between the one-shot learning features of biological neural processing systems and the slow Spike-Timing Dependent Plasticity (STDP) learning of computational models typically used in memristive device publications, one can reconcile these two behaviors by using more elaborate spike-driven learning methods that go beyond the basic STDP scheme. It has long been argued that the original STDP learning mechanism first observed in neurophysiology experiments of 1997–1998 is not sufficient to explain the wealth of learning data obtained experimentally in biology. Similarly, the computational neuroscience community has long proposed variants and extensions of the basic STDP mechanism that dramatically

improve its learning performance. Finally, these new and more powerful spike-based learning mechanisms are easier to implement in hybrid memristive/CMOS neuromorphic electronic circuits. The general class of extended STDP learning mechanisms that can be used to go beyond the plain STDP one is that of "Three Factor Learning" rules. A recent (2018) review paper by Emre Neftci presents convincing arguments about their benefits. Concerning the question I meant to ask, which is related to the linear and non-linear characteristics of the memristive devices used in these learning experiments, have you considered using the memristive devices with their non-linear properties as basic synapses (without making any effort to obtain linear behaviors out of them), but then changing the probability of updating the synapse in a linear way? Using memristive synapses as stochastic elements and controlling their stochastic behavior might be better/more efficient than re-engineering the devices and changing their physics to make them linear.

Hyunsang Hwang responded: Thank you for your comment. We have not yet evaluated the stochastic behavior of each synapse device. Considering an X-point synapse array device for HW DNN, it might be difficult to adopt stochastic behavior of each synapse device.

(614:[614]614) **Yuchao Yang** said: Regarding the summary of requirements for synapses, what is the requirement for the dynamic range or on/off ratio? And does the requirement on neurons depend on the neuron models?

Hyunsang Hwang answered: The dynamic range or on/off ratio depends on the variability and multi-bit states of synapse conductance. If the conductance uniformity is not so good, we need a large on/off ratio to meet MLC characteristics. Although we have not yet evaluated the effect of the neuron model on the synapse on/off ratio, I think both parameters might affect pattern recognition accuracy.

(805:[615]615) **Sweety Deswal** communicated: Among the devices with an abrupt change and those with an analog change of conductance, which would give a better option for a synapse device and why? How does the conduction mechanism vary with the two devices? How important is the switching mechanism for synapse devices, compared to memory devices?

Hyunsang Hwang communicated in reply: In response to your first question, an analog synapse is better because we can store multi-bit information. Regarding how the conduction mechanism varies with the two devices: for a filament switching device this is metallic, whilst it is Ohmic conduction for an interface switching device. Regarding your final point, understanding the switching mechanism is very important for synapse devices. Controlling the variability of synapse conductance is one of the key issues.

(617:[617]617) **Stefano Brivio** opened discussion of the paper by Wei Wang: According to the learning scheme proposed in the work (DOI: 10.1039/C8FD00097B), the spiking neural network (SNN) employs pulses whose duration is as long as the timescales the network is sensitive to. That is to say, if the

SNN must find the correlation in time among events occurring, *e.g.*, in the ms timescale, the length of the pulses driving the transistor gate must be at least 1 ms or even more. Therefore my questions are as follows: do the authors envisage any possible issue in scaling the implementation of this learning rule up to huge cross-bars of synaptic elements? Secondly, if a network is designed to be sensitive to events *e.g.* in the ms timescale, how can it manage more than two events that are occurring, by chance, much closer in time with one another? Would this be a problem?

Wei Wang responded: One issue for the scaling of the implementation to large cross-bars of synaptic elements might be the 1T1R structure we used here. In this work (DOI: 10.1039/C8FD00097B), we applied an exponentially decreasing signal to the gate terminal of the transistor, which is essential for the processing of temporal correlation among spikes. So, actually, we need three terminals (transistor gate, transistor source and RRAM top terminal) to be accessed for each synaptic element, which might hinder the higher density integration within a cross-bar architecture. We currently have some ideas to overcome this issue; for instance, to process the temporal spiking information by utilizing the volatile behavior of a volatile RRAM device.[1] However, a better uniformity of the device and an elaborate learning methodology are both needed. The exponentially decreasing signal we used and the scale of the recognizable temporal information is indeed on the millisecond scale. We believe it is biologically plausible. The action potential in the brain is on the millisecond scale, and the typical temporal information received by two ears for sound azimuth location, *i.e.*, the interaural time difference (ITD) is also on the millisecond scale (-0.6 ms $\sim$ 0.6 ms).[2] For spike events in a much longer time interval, our system is not able to process them; however, on the other hand, they might not be considered as a "temporal pattern", since they are discreted to each other and are less time-related. These cases might be suitable to be solved with a system which is able to store and process long-term memory with memory cells.[3]

1 W. Wang, A. Bricalli, M. Laudato, E. Ambrosi, E. Covi and D. Ielmini, Physics-based modeling of volatile resistive switching memory (RRAM) for crosspoint selector and neuromorphic computing, in the International Electron Devices Meeting (IEDM), 2018, 40.3.1–40.3.2.
2 W. Wang, G. Pedretti, V. Milo, R. Carboni, A. Calderoni, N. Ramaswamy, A.S. Spinelli and D. Ielmini, Learning of spatiotemporal patterns in a spiking neural network with resistive switching synapses, *Sci. Adv.*, 2018, **4**, eaat4752.
3 S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.*, 1997, **9**, 1735–1780.

(618:[618]618) **Dirk Wouters** remarked: It is interesting that in your paper (DOI: 10.1039/C8FD00097B) you go beyond the classification of static images, and that your proposed circuit is successful in recognizing temporal series. I do have a question though about the learning methodology. Do you still use supervised learning, and can you go to a self-learning system, too?

Wei Wang answered: The learning strategy we showed in this paper (DOI: 10.1039/C8FD00097B) is supervised learning. We can conduct the learning in an unsupervised fashion, as we did in the STDP learning work in our group's previous work.[1] Supervised learning for a spatiotemporal neural network is much

quicker than the unsupervised learning.[2] Therefore, we chose the supervised learning to more efficiently demonstrate the learning and recognition of temporal information. Unsupervised learning can also implement the idea of processing temporal information. Actually, we cannot say that unsupervised learning is a self-learning methodology while the supervised learning is not. In the learning strategy shown in this work, the "outside world" only provides the input spikes and supervised signal; the "learning" of the system, or the update of the synapse weights (that is, which synapse should be potentiated or depressed and when to apply voltage to potentiate or depress the synapse), is all determined and conducted by the memristive learning system itself.

1 G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli and D. Ielmini, Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity, *Sci. Rep.*, 2017, **7**, 5288.
2 Q. Yu, H. Tang, K. C. Tan and H. Li, Precise-Spike-Driven Synaptic Plasticity: Learning Hetero-Association of Spatiotemporal Spike Patterns, *PLoS One*, 2013, **8**, e78318.

(619:[619]619) **Ella Gale** asked: I'm looking at Fig. 9, section 2.7 of your paper (DOI: 10.1039/C8FD00097B), with the 'X' that moves around. Did you actually build this system? If so, how did it work? It would be great if it does work. Also, you used unsupervised learning in the first layer; what sort of unsupervised learning do you use, which algorithm? Did you find any effect of the shape, *i.e.* the X's compared to other shapes?

Wei Wang responded: At this stage, the detection of a moving object shown in Fig. 9 of our paper (DOI: 10.1039/C8FD00097B) is just a proposal. We haven't made the hardware implementation. However, our group has experimentally shown the spiking pattern recognition system[1] which is needed in the first layer, and here, we show the experimental temporal neural network which will be used in the second layer. We can envision that the combined network should work well. Yes, we propose to use unsupervised learning in the first layer with an STDP algorithm.[1] The system can learn to recognize other patterns, *e.g.* 'O', or another complex object if we have a larger-scale system. The recognition depends on what pattern has been learned by the first layer.

1 G. Pedretti, V. Milo, S. Ambrogio, R. Carboni, S. Bianchi, A. Calderoni, N. Ramaswamy, A. S. Spinelli and D. Ielmini, Memristive neural network for on-line learning and tracking with brain-inspired spike timing dependent plasticity, *Sci. Rep.*, 2017, **7**, 5288.

(620:[620]620) **Ella Gale** said: You did look at other patterns as well as the X, then. Is it the case that the network would follow any moving pattern, or is it specific to the recognised shape?

Wei Wang responded: It depends on the pattern it learned. If the learned pattern is X, then it will recognize the moving pattern of X, while if it learned another pattern, it will be specific to that pattern.

(621:[621]621) **Ella Gale** remarked: A quick comment; you probably know this, but if you don't it'd be interesting for you to look at. There are some results from neuroscience demonstrating that neurons can change the thickness of the myelin sheath layer, and controlling by the thickness of that layer, the speed of signal

transfer along the neuron can be altered. There is the suggestion that this helps neurons control when the neural signals arrive and synapses, and thus might be involved in temporal coding? It might be an interesting place to look at for more bio-inspiration (as we are), if you weren't aware of that.

Wei Wang responded: Thanks for the comment. I totally agree with you that the thickness of the myelin sheath layer would control the speed of action potential along the axon. This is another source of the complexity of the spatio-temporal coding in biological neural systems. There are surely many mechanisms that need to be explored in our brain, which might need efforts both from us and from neuroscientists.

(622:[622]622) **Carlo Ricciardi** opened discussion of the paper by Simon Brown: Why did you choose tin clusters instead of more standard electromigrating materials like Ag or Cu?

Simon Brown answered: Over the years we have seen similar switching behaviour in percolating films made with several other materials, I think the first time we reported this was in ref. 1. Recent work on similar behaviour in devices made using gold particles was reported in ref. 2 and 3. We have continued to use tin particles simply because the process is straightforward and has been well optimised in our lab, but we believe that similar devices can be made using many other metals.

1 M. Schulze *et al.*, *Eur. Phys. J. D*, 2003, **24**, 291.
2 C. Minnai *et al.*, *Sci. Rep.*, 2017, **7**, 7955.
3 C. Minnai *et al.*, *Nano Futures*, 2018, **2**, 011002.

(623:[623]623) **Carlo Ricciardi** said: What about the retention of your states? Since you are dealing with atomic switches of quantum point resistance, don't you experience quick segregation of the conductive filaments?

Simon Brown answered: Retention was discussed in a previous paper:[1] essentially we find that resistance of the devices is unchanged for weeks at a time when either a read voltage or no voltage is applied, *i.e.* the filaments are stable for at least weeks at a time.

1 S. K. Bose *et al.*, *IEEE Trans. Electron Devices*, 2017, **64**, 5194–5201.

(624:[624]624) **Maximilian Speckbacher** commented: Thanks a lot for the interesting work presented in this discussion. I have a question regarding the charge transport mechanism in your network: in lines 23 to 26 in your paper (DOI: 10.1039/C8FD00109J) you state that the "increase in activity of the network for stronger excitation pulses" is either due to "increased activity of the same synapses or activation of new previously inactive synapses" (see also Fig. 1d in your paper). My question is: what is the reason you do not see a "winner takes it all"-type conductance often observed in unordered networks (such as in ref. 1) and why are you sure charge indeed flows over multiple pathways? Could it not also be that you have a single, rather unstable filament showing the same features?

1 H. G. Manning *et al.*, *Nat. Commun.*, 2018, **9**, 3219.

Simon Brown replied: I will answer each part of the question sequentially. Firstly, as discussed at the end of the same paragraph, we believe that the increased activity at higher voltages largely occurs because additional sites are activated. The higher voltage obviously increases the electric field in each tunnel junction, so that the threshold for activation is exceeded in more junctions. These effects will be discussed in more detail in a future paper. Secondly, in ref. 1 we have previously modeled the activation of junctions across our network and indeed see a "winner takes all"-type conduction in our model when junctions are only allowed to turn on (*i.e.* in a version of the model in which atomic wires are formed across tunnel gaps but cannot be disconnected). In the experiments, and in the main (probabilistic) version of the model discussed in ref. 1, when atomic wires are disconnected in one location alternate pathways through the system are activated. We believe that the more complex switching behaviour observed across our whole networks is advantageous for useful computation. Thirdly, it is possible to make devices in which the conductance of a single atomic wire/filament dominates the conductance of the device, leading to the observation of quantised conduction (see ref. 2). However, in the present devices, made using the techniques discussed in ref. 3, it is clear that there is a qualitative change in behaviour (compare archetypes A and B discussed in the paper, DOI: 10.1039/C8FD00109J) and that at high voltages multiple sites in the network are active (see also my first point). We have a significant amount of additional data that supports this view – this will be published in due course.

1 S. Fostner and S. A. Brown, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2015, **92**, 052134.
2 A. Sattar *et al.*, *Phys. Rev. Lett.*, 2013, **111**, 136808.
3 S. K. Bose *et al.*, *IEEE Trans. Electron Devices*, 2017, **64**, 5194–5201.

(625:[625]625) **Ella Gale** commented: Firstly a comment: Erokhin *et al.*'s group took a conductive polymer, PANI (polyaniline), an electrolyte, PEO (poly(ethylene oxide)), and Au nanoparticles – these are the ingredients for the 3-terminal memristor – and mixed it together to make similar types of networks to the ones you've shown here. It was featured in Chemistry World.[1] I mention it as they found a way to do rewritable learning and permanent learning in this 'soup' of stuff dependent on how you charge the different electrodes with respect to each other – I suggest you look at their work. And a question – I appreciate you've been talking about criticality here; what sort of networks do you think you have, as in small world, nearest neighbour, random? And how can you tell that from the switching dynamics?

1 https://www.chemistryworld.com/news/training-a-memristive-network/5455.article

Simon Brown answered: There has been quite a lot of previous work on complex networks of many kinds. I guess that what could be viewed as similar depends a lot on your point of view. We believe that the nanoparticles deposition processes used in our work are particularly valuable because they allow devices to be made controllably close to the percolation threshold. The connectivity in our percolating networks is probably best described in ref. 1. Groups of interconnected particles are connected to neighbouring groups through tunnel gaps/

atomic scale switches, but one has to be careful when using terminology like "nearest neighbour" as to many people this would create an image of regular nodes that connect only to their neighbours – such networks have a very different structure to ours.

1 S. Fostner and S. A. Brown, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 2015, **92**, 052134.

(626:[626]626) **Stanley Williams** commented: Along the lines of trying to obtain inspiration from more recent discoveries in biological sciences, I saw an exciting talk at the DARPA Electronics Resurgence Initiative (ERI) Summit a couple of months ago,[1] in which Prof. Surya Ganguli of Stanford University postulated that a single synapse is not just a scalar, but is a tensor, and thus stores multiple memories in a single synaptic junction. Your system looks like it could be utilized for such a tensor-like synapse. The interesting question is how you address it, to enhance one element of the tensor as you're making the connections. In a different talk, which was complementary, it came up that recent research from the Technion in Israel has shown that spikes are not just spikes. Their shapes carry information in the brain, not just a digital entity but actually an information packet. Tensor-based synapses may be capable of unpacking and repacking the information in such a spike. This enables the quantity of information flowing through a brain to be orders of magnitude larger than has been estimated in the past. Your quantum-dot systems or nanoparticles, are you thinking of them in terms of a computational system? You need some type of gain and/or the ability to combine signals in a nonlinear fashion to achieve computation. A passive network, no matter how complex the connections, really doesn't do you much good. If your system has the capability of multi-state or tensorial storage, it could be a very interesting and powerful addition to the neuromorphic computing toolchest.

1 https://www.darpa.mil/news-events/electronics-resurgence-initiative-summit

Simon Brown answered: Thanks for these interesting comments, I was not aware of this specific work but obviously will look into it now. But there are a variety of existing computational schemes that we think can be implemented using our networks. For example, the concept of multistate storage is similar to that in several versions of reservoir computing, which uses a complex network to generate higher dimensional outputs, which can then be used for pattern recognition (among other things).

(627:[627]627) **Stanley Williams** remarked: I do not see how your multi-particle system can have gain, which you need to amplify a signal and to perform an actual computation. You have essentially fluctuation-dissipation in your resistor network, which is nice to have for introducing noise, but it doesn't give you a nonlinear means for a computation. When Leon Chua talks about neurons in the brain, he uses the phrase poised on the edge of chaos – a neuron has an internal resting potential such that a very small amount of additional energy that comes in will release a larger amount of energy – *i.e.* you've got gain. However, you don't have that ability to amplify a signal in a passive network like this.

Simon Brown responded: We think that the idea of the edge of chaos, or criticality as it is often called, is a really important one for our networks, and we hope to publish a detailed study soon. But in regard to your specific points – firstly, there are multiple types of computation that are possible with our devices, and in, for example, reservoir computing, what is required is a mapping of the inputs into a high-dimensional space, which we think will be possible using devices with multiple output contacts (although there are other options). Secondly, in regard to gain, one has to think about the type of signal that is important in this type of network. In fact when the network is biased in the right way, a small additional input causes a cascade of signals in much the way that you describe in your question. We will discuss this in our forthcoming paper.

(628:[628]628) **Yuchao Yang** asked: This is a type of physically evolving network, dynamically responding to external stimuli. When it is used for computing, what aspect of the network is used to represent information and computing results? Either the distribution of the particles or electrical outputs?

Simon Brown answered: We expect that the electrical outputs from devices with multiple output contacts will provide the required mapping into a higher dimensional space, but there are several other approaches that we are in the process of exploring.

(629:[629]629) **Yuchao Yang** commented: In the form of reservoir computing, the complex spatial and/or temporal features are mapped into the high-dimensional space. In this physically evolving network, are the mapping results the topology, connectivity of the clusters or other network properties?

Simon Brown replied: We expect that the electrical outputs from devices with multiple output contacts will provide the required mapping into a higher dimensional space, but there are several other approaches that we are in the process of exploring.

(630:[630]630) **Sanne Berg** said: Could you please elaborate a bit more on the self-assembly of the network during the fabrication process. Is it possible to control the structure of the network during growth?

Simon Brown responded: The self-organisation process is discussed in ref. 1 and 2. For the devices discussed here, 7 nm particles were deposited under controlled conditions (ref. 2) that result in a ~20 nm typical structure size.

1 A. Sattar *et al.*, *Phys. Rev. Lett.*, 2013, **111**, 136808.
2 S. K. Bose *et al.*, *IEEE Trans. Electron Devices*, 2017, **64**, 5194–5201.

(631:[631]631) **Michael Kozicki** enquired: Regarding the fractality of these networks – do you have any idea about the fractal dimension at criticality of the network? Have you measured it from the images or perhaps even extracted it from the measured electrical characteristics?

Simon Brown answered: One has to be careful to distinguish the exact property of interest, as, for example, the percolating "backbone", "hull" *etc.* have different fractal dimensions, but these are all typically in the range you mentioned in your previous question. These properties were investigated extensively in the early 1980s. A few of the standard references include ref. 1–3. That said, the literature still appears to have some gaps, and, for example, there is room for more work to be done to clarify the fractal dimensions of the relevant structures in continuum percolation as a function of density of the particles. Some of the above references measured the fractal dimensions of similar metal films from SEM images, but we have not done this ourselves because we believe it is near impossible to be sure whether connections exist in narrow tunnel gaps that are below the resolution of SEM or even AFM. We will soon publish evidence for self-similarity in the electrical characteristics of our devices.

1 H. E. Stanley and A. Coniglio, *Phys. Rev. B*, 1984, **29**, 522.
2 R. F. Voss, *J. Phys. A: Math. Gen.*, 1984, **17**, L373.
3 A Kapitulnik *et al.*, *J. Phys. A: Math. Gen.*, 1983, **16**, L269.

(632:[632]632) **Michael Kozicki** commented: Many structures in nature have a similar fractal dimension (1.5–1.8), particularly dendritic fractals that are nature's solution to optimization of flow (of mass, energy, information) in a system.

Simon Brown answered: I agree, and this is one of the reasons we focus on percolating structures: they typically have similar fractal dimensions to those found in nature.

(633:[633]633) **Dirk Wouters** asked: Your system is a conductive network, involving many different percolation paths. However, are you certain that the switching occurs inside the nanoparticle conglomerate, or could it also not occur by making and/or deleting connections between the (large) electrodes and this nanoparticle conglomerate? Could the asymmetry of such a connection structure at the two electrodes also not explain the simultaneous forming or deleting of contributing conduction channels at the same imposed voltage polarity? How would you know where the switching happens?

Simon Brown replied: We have thought about this carefully and have accumulated a lot of evidence that the switching is distributed across the network. This evidence includes SEM images of networks that have been stressed at high voltages which show that the network is "burnt out" in a very large number of locations spread across the network. It also includes results of modelling work that show that the distribution of measured conductance changes is consistent with the self-similar nature of the percolating network. There are numerous other measurements that are consistent with the distributed nature of the switching, and our next paper will present some key results that cannot be explained in other ways.

(634:[634]634) **Daniel Mannion** commented: Have you looked into the small signal frequency response of this setup? For example, once the device has

stabilised to a DC voltage, have you superposed a small signal AC to investigate any potential filtering?

Simon Brown answered: Yes, the system has interesting responses over a range of frequencies. We intend to publish these data in due course.

(637:[635]635) **Jonas Deuermeier** asked: From a materials science perspective – do you have any data on the oxidation state of Sn and how it may change during switching? It makes a large difference for the conduction mechanism, since SnO_2 is n-type whereas SnO is p-type. The use of gating voltages, which you are suggesting in the paper, requires knowledge on whether the junctions (the current limiting parts of the device) are semiconducting or not and if the contacts to Sn are rectifying or Ohmic. With this in mind, what is the role of the Sn oxidation state in the stability and (non-)volatility of your device?

Simon Brown answered: Perhaps the most important thing to make clear here is that while the devices are fabricated in a controlled atmosphere, they are measured in vacuum,[1] and so we believe there is only a thin 'crust' of SnO_x on the surface of the metallic tin particles that form the percolating structure and that the oxide does not directly contribute to the transport properties of the devices. We will discuss gating and other device geometries in future papers.

1 S. K. Bose *et al.*, *IEEE Trans. Electron Devices*, 2017, **64**, 5194–5201.

(639:[636]636) **Tony Kenyon** commented: Looking at the agglomeration of these particles, when you agglomerate do you have an oxide between the individual 7 nm particles within the agglomerate or is the oxide outside the agglomerate, or both?

Simon Brown responded: Direct measurement of the location of the oxide is obviously difficult, because once devices are removed from the vacuum system additional and more comprehensive oxidation takes place, masking the original oxide position. Nevertheless, the observed stabilisation of the structures (see ref. 1) suggests the oxide forms a thin/partial shell around the final network.

1 S. K. Bose *et al.*, *IEEE Trans. Electron Devices*, 2017, **64**, 5194–5201.

(636:[637]637) **Ella Gale** opened a general discussion of topics discussed in the session: A question for everyone: Chua said in 1976 that memristors are synapses and there has been a lot of work on using them as synapses. Later, Chua compared memristors to ion pumps in the Hodgkin–Huxley model. Other than Professor Haga's work which was obviously inspired by that, why is no one else currently looking at the proton pump angle? Is it that they think it doesn't work? Why are we concentrating so much on synapses rather than on proton pumps?

Stanley Williams responded: There has actually been a lot of work in this area. For example, see ref. 1–3.

1 L. Chua, V. Sbitnev and H. Kim, Hodgkin–Huxley axon is made of memristors, *Int. J. Bifur. Chaos*, 2012, **22**, 1230011.

2 M. D. Pickett, G. Medeiros-Ribeiro and R. S. Williams, A scalable neuristor built with Mott memristors, *Nature Materials*, 2013, **12**, 114–117.
3 W. Yi, K. K. Tsang, S. K. Lam, X. Bai, J. A. Crowell and E. A. Flores, Biological plausibility and stochasticity in scalable VO_2 active memristor neurons, *Nature Communications*, 2018, **9**, 4661.

(638:[638]638) **Tony Kenyon** addressed Ella Gale: We published a paper[1] that showed that Si oxide memristors can behave in ways that mimic some of the behaviours of neurons. They can generate spikes whose frequency depends on bias current; they can perform integration, and they show thresholding behaviour. We are very much interested in using memristors as neurons as well as synapses.

1 A. Mehonic and A. J. Kenyon, *Frontiers in Neuroscience*, 2016, **10**, 57.

Ella Gale responded: Brilliant, I would like to see more of this work.

(701:[701]701) **Itir Köymen** opened discussion of the paper by Giacomo Indiveri: I have two connected questions. Firstly, when you talk about the normalized circuit in the paper (DOI: 10.1039/C8FD00114F), if I understood correctly, you're suggesting that you're compensating for the variability of the devices by almost taking a mean function of the signal, then feeding into the system. So if you were to use actual memristors, you would utilize one normalizer circuit for however many memristors you might have to compensate for the variability of the devices, to ensure the variability doesn't get in the way of functionality? Later on in the paper you talk about exploiting mismatch. Instead of using the normalizer circuit you use the variability in your favor. Could you please explain this a little bit?

Giacomo Indiveri replied: The best way to understand the normalizer circuit's function is to consider the case of two memristive devices per synapse, that are updated in a push-pull fashion (*i.e.* when the conductance of one is increased, the conductance of the other is decreased and *vice versa*). In the extreme case when one is set, the other is reset and *vice versa*. In these conditions, the output current of the normalizer will be proportional either to the maximum conductance divided by the sum or maximum plus minimum, or the minimum conductance divided by the same quantity (maximum plus minimum). If the on/off ratio is large enough (*e.g.* 10 or more), then this ratio will be approximately equal to either one or "zero" (*i.e.* a very small number). This divisive normalization dramatically reduces the effect of variability. The larger the on/off ratio, the better is the mismatch attenuation. Later on, the other message we try to convey is that variability can actually be useful for computation. Within this context, the role of the normalizer is to provide a circuit that can reduce the amount of device variability, so that the designer has a knob or tool to control the amount of variability desired for the given computational task at hand.

(704:[704]704) **Geoffrey Burr** commented: In the context of this discussion of harnessing randomness, a very useful device for the memristor community to explore would be a compact source of true random numbers. Desired features would be a high rate of random numbers, full independence, control over the bias (between 0 and 1), and low power.

Giacomo Indiveri replied: I agree completely. But in my experience, it is very difficult to design true random number sources using compact devices. Usually imperfections or intrinsic properties of the device bias the source of randomness. One very elegant solution to this problem I remember was proposed by colleagues using neuromorphic analog circuits and floating gates.[1]

1 P. Xu, Y. L. Wong, T. K. Horiuchi and P. A. Abshire, Compact floating-gate true random number generator, *Electronics Letters*, 2006, **42**, 1346.

(705:[705]705) **Yuchao Yang** enquired: How much biological principle do you think should be considered for the development of neuromorphic hardware? To be more specific, are excitatory and inhibitory neurons and synapses all necessary?

Giacomo Indiveri replied: Our approach at the university is to emulate the biophysics of neural processing systems using the physics of electronic devices as faithfully as possible, for basic research purposes. However, in our experience, every mechanism we emulated, such as excitatory and inhibitory synapse properties, short-term plasticity, homeostatic adaptation, *etc.* has turned out to have an important computational role in the artificial neural processing systems we built. So in my opinion, for the design and exploratory phase of the development of beyond von Neumann brain-inspired computing systems, it is very important to include as many biological principles as possible. For a more practical development phase, when a specific product has to be made, then it might be possible to reduce the number of biological principles to the bare necessary ones.

(706:[706]706) **Itir Köymen** opened discussion of the paper by Themis Prodromakis: Looking at Fig. 3b of your paper (DOI: 10.1039/C8FD00130H), you use a RESET signal to make sure that the memristors do not reach R_{off}/saturate, right? You also mention that as a way to circumvent having to apply a RESET signal, volatile memristors could be utilised. I realise this is the topic of another paper, but what is the trade-off? Do you need to wait for a certain time period for the memristors to "forget" and reset to benefit from the volatility? And by equal measure, do volatile memristors require faster spike trains?

Themis Prodromakis replied: Indeed the trade-off lies with the ability of the technologies to sustain a measure of the overall activity. Thus, one would have to optimise the timing of recordings for evaluating the overall activity carefully so that spiking events are not missed and/or sufficient "forgetting" allows for appropriate detection rates.

(708:[708]708) **Ella Gale** asked: In ref. 1 we published experiments where we connected memristors to neural cell cultures, and we saw that the cells were capable of occasionally switching the memristors (see Fig. 1 below). I wonder if this sort of action might remove the necessity of you resetting your devices; what are your thoughts on that? We were excited as we thought that this perhaps showed that neural cell spikes could directly affect the behaviour of memristors. Furthermore, what are your thoughts on how we could combine your

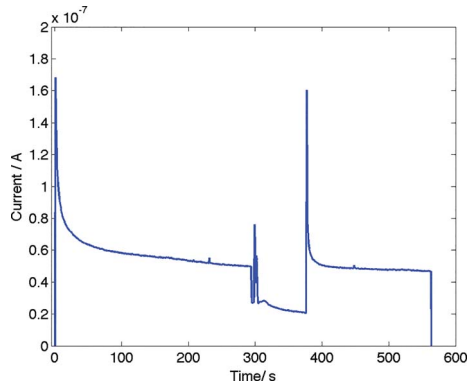


Fig. 1

spike-sorting algorithm with this type of set-up, to allow for a practical proof-of-concept of memristor spike-sorting hardware?

1 Deborah Gater *et al.*, Connecting spiking neurons to a spiking memristor network changes the memristor dynamics, 2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS), IEEE, 2013.

Themis Prodromakis replied: Fig. 1 that you have provided shows some arbitrary spiking activity, and it is difficult to evaluate the source of this behaviour in a clear and undisputed manner. The suggestion would be to employ the methodology we presented in this work (for spike sorting, as well as previously[1] for spike detection) for evaluating reliably the impact of neuronal spiking activity to resistance modulation.

1 I. Gupta *et al.*, *Nat. Commun.*, 2016, **7**, 12805.

(709:[709]709) **Geoffrey Burr** enquired: Given that these memristors might be on a chip implanted either inside the patient's skull (with the electrode array), or perhaps outside with the data collection electronics, what is the required endurance of the memristors in this application, in terms of RESET-SET cycles they must be able to achieve?

Themis Prodromakis answered: Endurance of memristive technologies is always of concern, especially in applications where devices are programmed many times, such as in our case. Clearly a technology with a rather high endurance capability will be required; nonetheless, our methodology allows the biasing signals (recorded spikes) to be optimised, such that they can have a minimum impact on the change in resistive state that will in turn allow such technologies to be utilised for a much longer time.

(712:[712]712) **Daniele Ielmini** asked: Is the spike sorting sensitive to the shape of the spike, or only to the highest amplitude? Can it be interpreted as a weighted integration of the time-dependent signal?

Themis Prodromakis answered: This is indeed a weighted integration of the time-dependent signal, and thus not only dependent on the spike's amplitude.

(713:[713]713) **Itir Köymen** opened discussion of the paper by Ella M. Gale: Could you please explain what is meant by "bounce-back"? To me it looks like the opposite of the habituation observed for the positive current range is happening for the negative current range in Fig. 3 of your paper (DOI: 10.1039/C8FD00111A).

Ella Gale answered: There are two things being confused here: "bounce-back" is my name for the response when you zero or reverse voltage polarity, and Fig. 3b of my paper (DOI: 10.1039/C8FD00111A) is the negative current response to the habituation experiment, which is really a measure of the bounce-back effect over the course of the experiment. I'm not the best person to name things, I think. By "bounce-back" I merely mean the response of the device to a change in voltage polarity (where 0 counts as a change in polarity), which is observed in the opposite direction. So if I turn a device on with a positive voltage, $+V_A$, I get a positive current response of $+I_A$. If I then either change the voltage to 0 or give it a negative voltage, $-V_B$, I see a response current, $-I_B$, which is in the opposite direction: it is this response that I call bounce-back. I use that term as that describes the observed process, and avoids naming the cause of that process (*i.e.* memristance, capacitance, inductance, some combination of the three) as I do not know the precise cause. In regards to the habituation example in Fig. 3 of my paper, the "bounce-back" response (Fig. 3b) changes over the habituation experiment, and yes, it goes in the opposite direction to the positive current habituation. So, in the positive current the memristor first responds largely to the stimuli, with the response decreasing with further presentations of the stimuli (habituation). On the negative side, the "bounce-back" response increases with presented 'stimuli' (although the stimuli in this case is the return to the 0 V).

(714:[714]714) **Itir Köymen** said: Did you model the logic applications or did you conduct experiments with actual devices?

Ella Gale responded: Thank you for your question; for some reason, this often causes confusion when I present my work. All the reported data (specifically, *I-t* plots) in my paper (DOI: 10.1039/C8FD00111A) came from real devices (those presented in ref. 1). The currents in the tables are those from real devices, in some cases rounded up to integers to make it easier to demonstrate design principles. The only 'modelling' in the paper is the drawing of perceptron models to describe the physical processes.

1 Ella Gale *et al.*, The effect of changing electrode metal on solution-processed flexible titanium dioxide memristors, *Materials Chemistry and Physics*, 2015, **162**, 20–30.

(715:[715]715) **Stanley Williams** said: I'm afraid the devices you show in your paper are not memristors at all, but are in fact leaky capacitors. Just putting some form of titanium dioxide between two electrodes does not mean that you will obtain memristor behavior. Generally, a dielectric between two electrodes is a capacitor – you only get memristive behavior if there is some type of ionic species, *e.g.* O vacancies, that acts as a mobile electronic dopant, or some type of

phase change. You have neither in your system, as can be seen in Fig. 1 of your paper (DOI: 10.1039/C8FD00111A), which shows a classical transient response of a capacitor after abruptly turning on a voltage. Sometimes one can see a similar transient in a circuit containing a memristor because there is an effective capacitor in parallel with a memristor fabricated from a dielectric between two electrodes – in that case, one needs to construct a model for the physical device that has a memristor in parallel with a capacitor to model it correctly. There is an effective parallel resistance in your system, either from leakage through your capacitor or from the measurement circuit. In Fig. 2 below, I have taken the Fig. 1 from your paper and shown that you can compute the RC time constant from the 1/e decay time of your capacitor, which is about 5 s (there appears to be a second and somewhat longer time scale in your system, which probably comes from electrons escaping from charge traps in your sol–gel oxide). You can then calculate the resistance of the effective parallel resistor from $R = V/I$, where V is the DC bias voltage (0.1 V? – it was not clear to me what the applied voltage was for Fig. 1) and I is the asymptotic current (~2.7 nA). If my guess for the voltage is correct, then I compute that the parallel resistance is ~37 M and thus the capacitance would be ~135 nF. This is a fairly large number, but I did not see the size of your devices in the paper so it is at least plausible. In principle, you can now build a nearly identical equivalent circuit from a set of standard linear capacitors and resistors, with no memristors involved.

Ella Gale answered: Stanley Williams raises an interesting point: is this observed spike response of a memristor to a voltage step merely explained by a capacitor and resistor in parallel (as he suggests) or is it all explained by memristance, or is it explained by a memristor and capacitor in parallel? As my memristor devices are physical devices, and not ideal circuit elements, they are highly unlikely to be modelled by pure memristor theories. So the question really is, are these devices (and thus my observed behaviour) due to a capacitor-resistor (RC) circuit, or a memristor-capacitor (MC) circuit? I will attempt to summarise the information I have on this point. Regarding the methodology points he

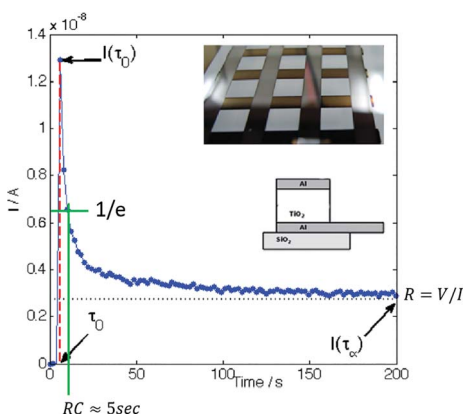


Fig. 2 Modified from DOI: 10.1039/C8FD00111A, reproduced with permission from the Royal Society of Chemistry.

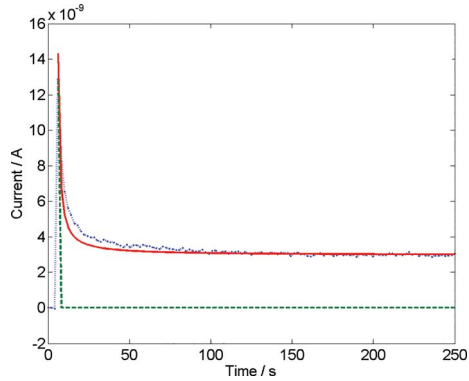


Fig. 3 A positive spike with the memory conservation theory fit in red and exponential in green.

makes: my devices are described in ref. 1 and 2, and are based on the work in ref. 3 (this contains the size *etc.*), and there is no parallel resistance from my measurement circuit as I used the top of the range Keithley electrometer which is specifically designed for testing devices to high accuracy; the manufacture was done professionally for me by HP. I am confused as to his comment that "Just putting some form of titanium dioxide between two electrodes does not mean that you will obtain memristive behaviour if there is some type of ionic species, *e.g.* O vacancies, that acts as a mobile electronic dopant, or some type of phase change. You have neither in your system", as I do have oxygen vacancies in my system and am using the type of set-up and materials where you expect them. Also my devices require aluminium electrodes in order to work, which suggests that they are a source/sink of oxygen ions (in an Al_2O_3 layer on the electrodes). I actually get both filamentary memristors (triangular type in ref. 2) and what appear to be phase change memristors (curved type in ref. 2). Regarding whether

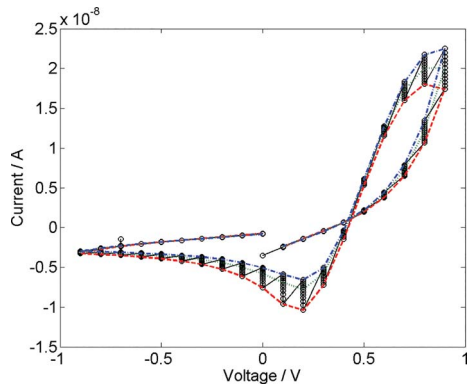


Fig. 4 An example of the type of experiment mentioned above in point 6, as a member of the audience asked me to include one; this demonstrates how the hysteresis changed with longer dwell time at each voltage, due to the spikes, and explains where the hysteresis frequency effect comes from.

my devices are memristors with some capacitance or resistors with some capacitance: firstly, I agree that it is possible to make capacitors with my set-up. I recorded that 3/64 of the devices were capacitors not memristors,[2] but I tested them to see which I had before experimenting with them, and discarded them for the spiking tests. Secondly, many of my devices exhibit the pinched hysteresis loop typical of memristors: see ref. 1, ref. 2 and my response to another question in this discussion. Thirdly, my curved-type memristors do indeed have some capacitance, and these devices produce pinched *V*–*I* curves over a large *V* range. To differentiate devices, I classify them based on the *V*–*I* curves observed over a low voltage range (to avoid forming), and I find a distinctive "jelly-bean"-type curve (Fig. 6 in ref. 2, and Fig. 2 in ref. 4). These curves do not cross zero, but are pinched towards it, and contain an asymmetry in the same direction as the larger pinched memristor hysteresis curves (the fabrication process causes non-symmetry in the electrodes, which shows up in the *I*–*V* curves as I always wire the devices the same way round during classification testing). These devices exhibit the (curved type) pinched hysteresis loops at larger voltage ranges. So I believe that my devices are non-ideal memristors which do include a little capacitance. A fourth point is that part of the "jelly-bean" shaped curves can be explained by a memristor theory, but may be missing a capacitance. I looked at devices with different electrode widths to compare a few memristor theories,[4,5] and found that the memory-conservation theory,[6,7] with added contact resistance, could explain the hysteresis size well, although it is clear to me now that adding in a typical, small capacitance, such as that measured in my capacitor-only devices (Fig. 6d in ref. 2) to the theoretically predicted curves (Fig. 16 in ref. 4) would make them resemble the "jelly-bean" curves (Fig. 2 in ref. 4). This is something I plan to add to the paper. I have now shown that my devices are memristors under A.C. voltage, which also possess some capacitance; now I need show that the D.C. response includes memristance, rather than just being resistance plus capacitance. Although, if my devices are memristors under A.C., I do not expect them to stop acting as memristors if I put a voltage spike into them (how can the memristor possibly know that an input voltage spike is part of a D.C. experiment rather than a discretised A.C. experiment of the type I use to get my pinched *I*–*V* curves?) Fifth, in ref. 8–10 I fitted the memristor spikes to the memory conservation theory of memristance, and found that this fitted the spikes better than an exponential (the type of which you'd fit in an RC description as Stanley Williams described). Finally, and most importantly, I will now show that these spikes are part of our definitions of memristance as they are behind the 'frequency effect' used by Chua as a 'fingerprint of the memristor'.[11] My sixth point is as follows: doing experiments with several recording time steps at each voltage in a stepped triangular voltage waveform reveals spikes that decay; if one were to join up the currents measured the same number of dwell timesteps after an input voltage change, one draws a pinched *I*–*V* curve with shrinking hysteresis. As the third fingerprint of the memristor is that the 'pinched hysteresis loop should shrink to a single valued function when the frequency tends to infinity',[11] the 'frequency effect', this demonstrates that the frequency effect is directly related to the D.C. spikes, as I concluded in ref. 12. The term 'current transient' only means that the current is short-lived and does not actually define the cause (capacitance or memristance). When I first noticed these spikes, I asked many memristor researchers if they had seen them in their devices: they had, many ignored them, so I know that my

devices are not unique in having this property (see my papers for when I found published memristor *I-t* curves). So, it is either the case that some previously observed 'current transients' are memristive in form, or that the frequency effect is due to the capacitance found in non-ideal memristors (which doesn't fit with Chua's equations, of course, as he modelled this effect in ideal memristors), or that Stanley Williams is correct, and my devices, and many other, possibly all, 'memristors' are just leaky capacitors. My usual response to this controversy is to say, well, I have the devices and they do what they do and we can do spiking computation with them, which is exciting, but it was probably a good idea to amass as much information as I could on this topic. I intend to post a longer form of this answer, along with more fits to more data, on arXiv in the near future.

1 Ella Gale *et al.*, Drop-coated titanium dioxide memristors, *Materials Chemistry and Physics*, 2014, **143**, 524–529.
2 Ella Gale *et al.*, The effect of changing electrode metal on solution-processed flexible titanium dioxide memristors, *Materials Chemistry and Physics*, 2015, **162**, 20–30.
3 Nadine Gergel-Hackett *et al.*, A flexible solution-processed memristor, *IEEE Electron Device Letters*, 2009, **30**, 706–708.
4 Ella Gale *et al.*, Which Memristor Theory is Best for Relating Devices Properties to Memristive Function?, arXiv:1312.4422v1.
5 Ella Gale, Benjamin de Lacy Costello and Andrew Adamatzky, The effect of electrode size on memristor properties: An experimental and theoretical study, *2012 IEEE International Conference on Electronics Design, Systems and Applications (ICEDSA)*, IEEE, 2012.
6 Ella Gale, The memory-conservation theory of memristance, *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, IEEE, 2014.
7 Ella Gale, The Memory-Conservation Theory of Memristance, arXiv:1106.3170.
8 Ella Gale, Ben De Lacy Costello, and Andrew Adamatzky, Observation and characterization of memristor current spikes and their application to neuromorphic computation, *AIP Conference Proceedings*, 2012, **1479**, 1898.
9 Ella Gale, Ben de Lacy Costello and Andrew Adamatzky, Observation, characterization and modeling of memristor current spikes, 2013, arXiv:1302.0771.
10 Ella Gale, Ben de Lacy Costello and Andrew Adamatzky, *Applied Mathematics and Information Sciences*, 2013, **7**, 1395–1403.
11 Shyam Prasad Adhikari *et al.*, Three fingerprints of memristor, *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2013, **60**, 3008–3021.
12 Ella Gale *et al.*, The short-term memory (dc response) of the memristor demonstrates the causes of the memristor frequency effect, *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2014.
13 Ella Gale, TiO_2-based memristors and ReRAM: materials, mechanisms and models (a review), *Semiconductor Science and Technology*, 2014, **29**, 104004.

(716:[716]716) **Stanley Williams** said: You're not doing any type of mathematics or logic with linear circuit elements; you're only doing a type of integration with a decay. You actually require more logic to analyze the signals you are observing than any type of computation that you may be doing.

Ella Gale answered: In response to your first point, "you're not doing any type of mathematics or logic with linear circuit elements", I'm using a non-linear circuit element called the memristor. Moreover, no one ever actually does mathematics or logic with physical devices; the devices act how they act as a result of their material properties and the laws of chemistry and physics (and some device and circuit design), we merely chose to interpret the output of this physical system as a type of logic/mathematics and give it some meaning. You stated "you're only doing a type of integration with a decay". Well, I said the devices were doing a non-linear summation, so yes, this is the underlying operation, but we can use that physical response to do a different type of logic. Regarding your

statement "you actually require more logic to analyze the signals you are observing than any type of computation that you may be doing", I disagree. To analyse my signals, all I need is something to implement a threshold on the output, a switch in the simplest case; this is a much simpler logic than what the memristor is doing. For example, to know if I have an output 1 in the XOR gate I need to implement the following mathematical operation: Abs[i] > 5.5 (and 0 would be Abs[i] < 5.5 and i is the output current), this is simpler than XOR[A,B]. Similarly, to know if I have an output sum of 2 I would need to implement the following mathematical operation: $x \leq i < y$, which is simpler to implement than a full adder (where x and y are the thresholds in the paper, and i is output current). I think these simple operations could be done with resistors and switches; the full adder requires quite a few transistors (six, I think).

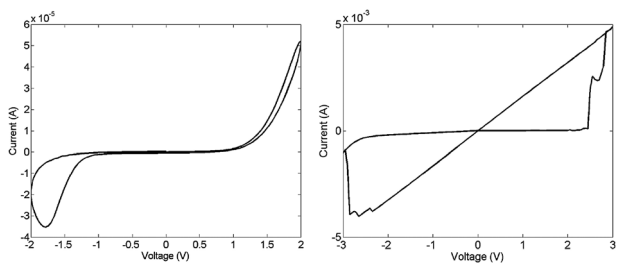
(719:[719]719) **Daniele Ielmini** said: Regarding the memristive effect in your device, do you see any forming operation to initiate the hysteresis behavior? Is it a reversible forming operation, or a destructive breakdown?

Ella Gale responded: No. The devices used for this spiking work are the ones from a previous paper,[1] specifically those with aluminium electrodes which didn't need forming. In that paper, I said that this was because the TiO_2 layer is an amorphous layer, so it already contains defects (and I think the act of sputtering aluminium on top of the layer might add them as well).

1 Ella Gale *et al.*, The effect of changing electrode metal on solution-processed flexible titanium dioxide memristors, *Materials Chemistry and Physics*, 2015, **162**, 20–30.

(720:[720]720) **Daniele Ielmini** commented: Is there any hysteresis in the *I–V* curves of your device? The presence of a quasi-static hysteresis, which is visible from a switching *I–V* curve, could provide the evidence for the memristance effect, as opposed to a reversible, capacitive charging/discharging effect. Then, one should prove that the memristive effect is functional for the observed XOR or full adder computation.

Ella Gale answered: The devices I used for this paper (DOI: 10.1039/C8FD00111A) were the ones I made and classified in a previous paper.[1] A figure from that paper is shown here as Fig. 5, and you can see that the devices do have



(a) (b)

Fig. 5

memristive *I–V* curves. In my work, I classified the devices into 'curved' switching and 'triangular'; I believe most of the work presented in this paper was from the 'curved' type of devices. Incidentally, the 'curved' type of switching I observed is what others call bulk switching, and the 'triangular' type has been called 'filamentary' by other authors.

1 Ella Gale *et al.*, The effect of changing electrode metal on solution-processed flexible titanium dioxide memristors, *Materials Chemistry and Physics*, 2015, **162**, 20–30.

(721:[721]721) **Sebastian Siegel** said: Is it possible to use the effect you described in a cascading circuit? What would be the FAN-OUT of such a device and can it drive further neurons?

Ella Gale answered: I hope that it is possible to build cascading circuits, but I must confess that I do not know how to do this yet. Obviously, the point of building a full adder-alike was to chain the circuits and thus build a novel processor, but at this point I need the help of electronic engineers to do so.

(722:[722]722) **Geoffrey Burr** opened a general discussion of the topics discussed in the session: One general comment/question is the discrepancy between the holistic analytical approach of the "system as a whole" that seems to be required in many of these new computing techniques – brain-inspired computing, analog computing, computing "at the edge of chaos", *etc.* – and the highly modular engineering approach that drives most of the IT and semiconductor industry. That hierarchical engineering approach requires breaking a huge and complex design down into many individual and fairly simple modules, each small enough to be tackled by a small team of technicians and/or engineers, each with a tightly-defined set of requirements, specifications and deliverables. This only works if the system is successful as soon as each team delivers on its modular task. One worry is that these new computing techniques lead to a "chicken-and-the-egg problem" – we cannot write down these specifications or even identify the necessary modules and their interdependences until we get the whole computing system working, yet we are unable to actually build the computing system at scale without the efforts of all these modularized engineering teams.

(723:[723]723) **Geoffrey Burr** said: Another general comment: because of the huge cost of making a "new" fab, the need for it to be large in size (for economies of scale), and then the need to keep it >90% full (to make enough product to make the investment pay off), these decisions cannot be gambles. Historically, this has meant that these fabs only get built when the return-on-investment is already seen as a sure thing, which means satisfying existing markets (memory, processors, displays, cell-phones) that are not only already large, but expected to grow further. This could lead one to have a pessimistic view of the future of semiconductor manufacturing. A further general comment that could lead one to have an optimistic view of the future of semiconductor manufacturing: to date, the semiconductor industry has always depended on building new fabs that, by the time that the capital equipment is fully depreciated and thus "paid for", have inevitably fallen "far behind the cutting-edge". However, as Moore's Law slows down, we can

expect to see fabs that are, for the first time, still very close to the "cutting-edge" even when fully depreciated. Together with the significant economic uncertainties introduced by the slow-down of Moore's Law ("where to invest now to ensure a profit in 4–5 years?"), this could change the way that some fab-owners make decisions, particularly in terms of what they might be willing to try in their fabs.

(727:[727]727) **Ilia Valov** addressed Geoffrey Burr, Giacomo Indiveri, Themis Prodromakis and Ella Gale: How strong in your opinion is the relation between memristors and neuromorphic computing. More clearly, how important/suitable are memristors for this type of application compared to other competitive technologies?

Themis Prodromakis answered: Memristive technologies offer significant competitive advantages over other commercially available technologies when used for emulating neuronal dynamics, specifically: two-terminal architecture, low power consumption and scalability. These attributes make memristors ideal for delivering the hardware demands of neuromorphic technologies.

Ella Gale responded: Neuromorphic computing was around before memristors, and there are plenty of people working in the area who are not working with memristors (for example, the huge neural network computing community happy to build software models than run on graphics chips), so I see memristors as just one technology/approach under the larger field of neuromorphic computing. After all, there are many ways to build a brain-like computer and it is not obvious that memristors are the best approach. However, there are some intriguing facts about memristors that make them interesting for neuromorphic computers: firstly, the fact that they've been compared to synapses; secondly, the fact that they've been compared to ion pumps in a model of neuronal operation;[1] thirdly, the fact that they natively spike;[2] fourth, the fact that the brain seems to be poised on the edge of chaos, and it is easy to build complex non-linear circuits from them[3] or demonstrate chaotic behaviour;[4] and fifth, the fact that they've been shown to communicate electronically with neural cell cultures,[5] which might suggest that they operate in a similar way, or could move the question from neuromorphic computing to a neural interface. Being more vague, you can also point to the nanoscale size of most of them and use of ionic currents to relate them to neural cell operation, which might suggest that neurons might use memristive principles, and thus that a brain-like computer built from memristors might be more brain-like as they operate on similar principles. From the other side, I often argue, when people point out my memristor's slow switching speed, that the brain is slow. Both the brain and memristors are low-power, we've all seen the comparison between a supercomputer that is as complex as a cat's brain with a huge power draw and a real cat, which requires much less power to operate. With this list, I cannot think of another type of hardware that is as attractive for doing spiking neuromorphic computing. However, there are other architectures that are further along (FPGAs for example) that work well as well, so then I think it comes down to power draw and possible similarity between the device's operation and the brain itself at the level of the physics, and there I really think the ease at which the memristors spike and the fact they can interact to either do standard

logic operations (as I showed in my paper here, DOI: 10.1039/C8FD00111A) or to do more emergent, highly non-linear and rich spiking behaviour (see ref. 3 and 4) is what makes them most attractive for building neuromorphic hardware.

1 L. O. Chua and Sung Mo Kang, Memristive devices and systems, *Proceedings of the IEEE*, 1976, **64**, 209–223.
2 Ella Gale, Ben de Lacy Costello and Andrew Adamatzky, Observation, characterization and modeling of memristor current spikes, 2013, arXiv:1302.0771.
3 Ella Gale, Ben de Lacy Costello, and Andrew Adamatzky, Emergent spiking in non-ideal memristor networks, *Microelectronics Journal*, 2014, **45**, 1401–1415.
4 Lucia Valentina Gambuzza *et al.*, Experimental evidence of chaos from memristors, *International Journal of Bifurcation and Chaos*, 2015, **25**, 1550101.
5 Deborah Gater *et al.*, Connecting spiking neurons to a spiking memristor network changes the memristor dynamics, 2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS), IEEE, 2013.

Geoffrey Burr replied: I think that for the forward-inference application with pre-trained weights, memristors have a good chance to play an important role if they can show long-term conductance stability at high resistance values with complete tunability within a decent resistance contrast (*e.g.* 300 kOhm to 10 MOhm). For training applications, as Stanley Williams mentioned at one point, filamentary switching is a significant problem. So memristors offering truly gradual and symmetric switching without being adversely affected by "sharp" filamentary-type switching also have a very real chance for this application in my opinion. For true neuromorphic applications beyond backpropagation using spikes, I think we need to wait for the development of a learning algorithm that actually works (converges to a solution, providing performance that transcends deep learning, on an application people care about, in a scalable way so that making the network and/or the dataset bigger invariably makes things better). After we have such an algorithm, then we can figure out if memristors can play a role in its implementation.

Giacomo Indiveri replied: The relationship is extremely strong. On one hand, memristive devices can be used as binary non-volatile memory elements for configuring latches and digital circuits typically present in multi-core neuromorphic devices to configure the neural network topology, and the properties of the neurons and synapses (*e.g.* excitatory or inhibitory). On the other, the very same technology can be used to directly emulate the plasticity of real biological synapses, for implementing on-chip on-line learning. In both cases, this relationship becomes crucial for large-scale neural processing systems, thanks to their non-volatility properties. Indeed, if one had to upload all the network parameters and all the synaptic weights of a very large-scale neural network each time a new experiment is set up, or each time the device is reset and powered up, this would take an unacceptably long amount of time.

(728:[728]728) **Daniel Mannion** addressed Giacomo Indiveri, Themis Prodromakis and Ella Gale: Returning to an earlier discussion on fabs and comments on how we as researchers need to convince industry to invest as a neuromorphic community: with Giacomo's work demonstrating techniques of mitigating for non-ideal memristive devices as well as talk of small-scale application specific technologies, could we see a shift in our approaches to fabs? Such as smaller-scale

setups with a wider range of technologies. The trade-off would be device quality, which we as circuit designers acknowledge and design into our systems.

Themis Prodromakis replied: Different applications would certainly have different requirements in terms of memristive device performance. Nonetheless, one has to be aware that only mature memristive technologies will in the end be integrated monolithically with CMOS technologies.

Ella Gale answered: Whilst I agree with Geoffrey Burr's points about getting memristors into fabs being a good way forward for this field, I think we should also keep an eye on some emerging areas, such as 3D printing, internet of things, sensors and circuitry on soft substrates and the like, as we may see memristors appear first in these areas as they do not require a huge investment from a semiconductor firm. There's also the approach of constructionist science, where instead of designing all the parts of a circuit, a system is built and then tested for useful 'emergent' properties, which could then be used. This approach would work well with memristor networks, or reservoir computing, and combined with modern computing techniques, it could be a useful area for memristor circuits.

Giacomo Indiveri answered: Given the economy of fabs, I don't think academic or even industrial "neuromorphic" research groups would be able to convince companies of VLSI fabs to make significant changes to their process parameters. However, these companies already do and most likely will keep on investing in research on neuromorphic technologies, because it is a very promising development that might allow the exploitation of variable and faulty components in computing architectures.

(729:[729]729) **Ella Gale** addressed Stanley Williams: In your closing remarks, you state very clearly that memristors are defined as a specific mathematical concept (based on Chua's 1971 paper, I presume) – I agree, but how exactly do you explain what magnetic flux is in the memristor equations?

Stanley Williams answered: The best answer to your question can be found in the on-line Chua Lectures.[1] Prof. Chua goes through the reasons for why he created nonlinear dynamical circuit theory and explains his rationale for his definition of charge and flux, the latter of which is simply the time integral of the voltage across a circuit element. He shows that this definition does not have to involve a magnetic field at all, and is more general than a definition that centers on the magnetic field in a coil.

1 https://www.youtube.com/watch?v=_qAcUDZQktQ.

Conflicts of interest

There are no conflicts to declare.

