doi:10.1038/nature14441

1. Material and device stack optimization

Lower electroforming voltages reduce the electrical stress as well as current overshoot during the forming, which is a known risk factor contributing to device variability[1]. It had been noticed that lower forming voltages may be achieved in devices with higher conductivity, obtained by a combination of oxide layer thickness reduction and stoichiometry adjustment[2,3]. In our experiments, the oxygen concentration was continuously reduced by controlling the oxygen flow rate during the layer growth to the point at which resistive switching nearly ceased. The switching voltages of such devices were close to, or less than the electroforming voltage. However, many devices could not be turned off after switching and instead were shunted, or in some cases exhibited spontaneous complementary resistive switching. The observed nonlinearity of the *I-V* curves of such devices was also not large enough for crossbar operation. To address these issues, an Al_2O_3 layer was added to the device stack. (Multilayers and bilayers had been used previously to improve device performance[4,5].) An excessive increase of the Al_2O_3 barrier thickness causes an increase of the forming voltage, so that an optimum thickness had to be selected. As a result of material and stack optimization, the most suitable values of thicknesses for TiO_{2-x} and Al_2O_3 layers have turned out to be close to respectively, 30 nm and 4 nm (Fig. S1b).

2. Electrical characterization of single devices

To perform electrical characterization, single 200 nm × 200 nm devices of the dog-bone geometry (Fig. S1b) were fabricated first. (The inset in that figure shows the material stack parameters; note that they are slightly different from those in the crossbar-integrated devices – cf. the inset in Fig. 1b of the main text.) Figure S1a shows typical switching curves of such single devices, obtained by applying bipolar voltage sweeps. (To exhibit the *I-V* curve nonlinearity better, Fig. S1a also shows Ohmic currents for several resistance values.)

Figure S1c shows representative endurance test results for single devices. The data are obtained by repeatedly (over 5,000 times total) applying a sequence of set (-2 V, 500 μs), readout (0.1 V, 1 ms), reset (2 V, 500 μs), and again readout voltage pulses to a single device. The figure does not reflect the fact that approximately 7% of the negative pulses failed to switch off the device; however, this behavior may be attributed to an imperfect endurance setup rather than any deep device problem. This conclusion is supported by the fact that the device was being switched after each failure event. Some aging (in the form of a slight degradation of the ON/OFF dynamic range) is also visible, but generally the devices are rather robust. For example, during our experiments with the crossbar circuit, each of its memristors had been subjected, on the average, to 200,000 set/reset pulses even prior to the successful classification experiment.

As customary in nonvolatile memory technology, the retention test shown in Fig. S1d was carried out with a sample heated to 350 K. The device was first switched into a certain resistive state, and then its resistance was measured repeatedly by applying a 100 mV bias every

doi:10.1038/nature14441

2 seconds during a 50,000-second time interval. Such retention measurements were carried out for ON (highly conductive), OFF (highly resistive), and some intermediate states. Based on the measurements, the retention at room temperature is expected to exceed 10 years.

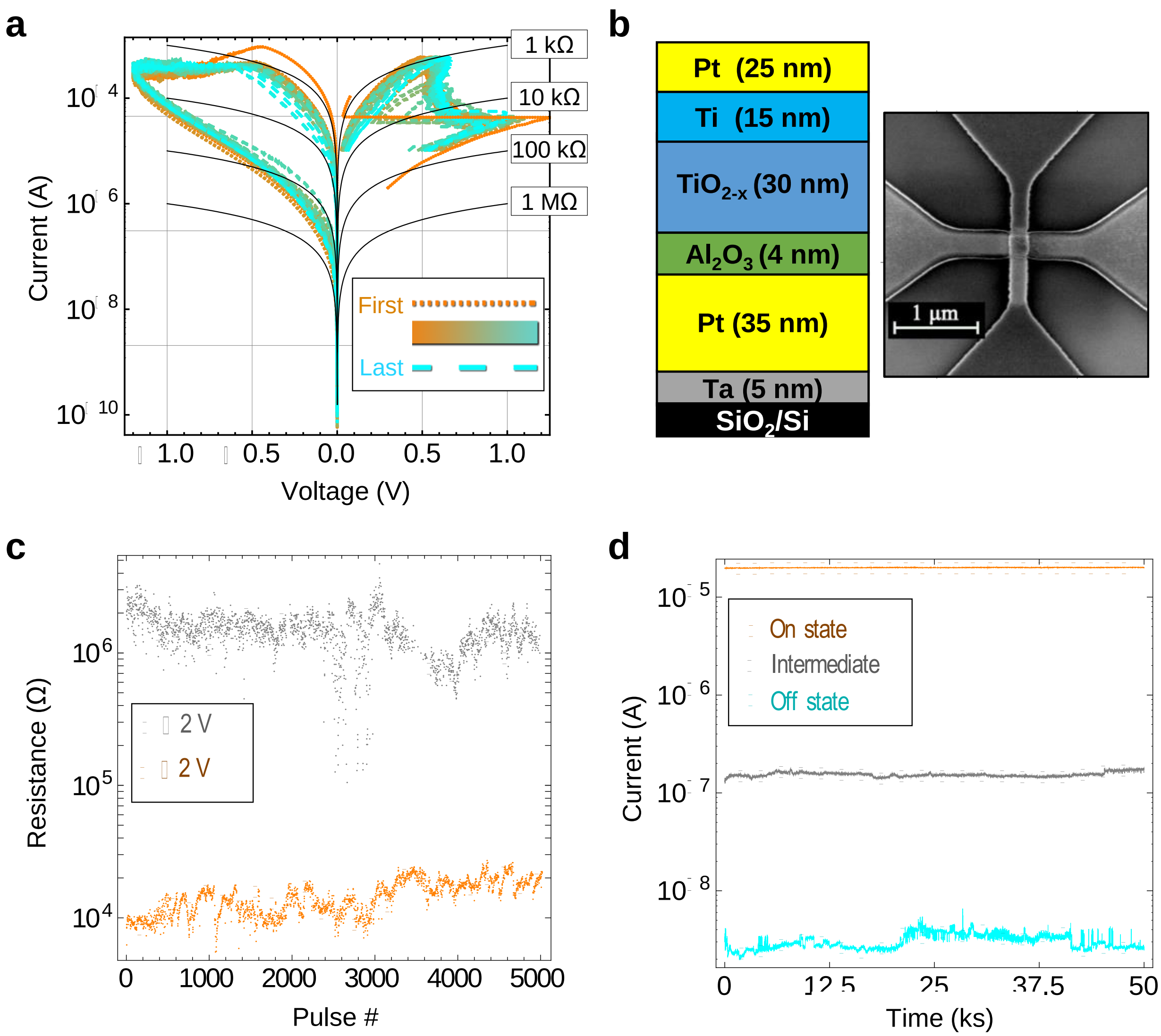


Figure S1. Isolated Pt/TiO_{2-x}/Al_2O_3/Pt memristive devices: (a) typical switching and electroforming behavior of a single device, (b) micrograph of a single device and its stack's structure, (c) switching endurance under a stream of ±2 V, 500-μs pulses, and (d) retention of 3 initial states at 350 K. To highlight the data trends, large markers on panels (c) and (d) show the results of every 100^{th} measurement at the endurance test, and of every $1{,}000^{th}$ measurement at the retention test. Note that panel (d) shows significantly lower OFF state currents (and hence much higher ON/OFF current ratio) as compared to those of panels (a) and (c) due to higher voltages applied to reset the device.

3. Crossbar circuit fabrication and packaging

Crossbar lines, 200 nm wide and separated by 400 nm gaps, were formed on 4” silicon wafers covered by 200 nm of thermal SiO_2. After the standard cleaning and rinse, fabrication started with an e-beam evaporation of Ta (5 nm)/Pt (60 nm) bilayer over a patterned photoresist to form the bottom electrodes (“rows”). After liftoff, the wafer was descum by active oxygen dry etching at 200°C for 10 minutes. Then, a blanket film consisting of a 4-nm sputtered Al_2O_3 barrier and a 30-nm TiO_2 switching layer was deposited from a fully oxidized target and a partially oxidized target, respectively. This bilayer was then removed by etching in an ICP chamber using CHF_3 plasma, while preserving it in the future crossbar area by a pre-deposited negative photoresist. After stripping the photoresist in the 1165 solvent for 3h at 80°C, the wafer was cleaned using a mild descum procedure performed in a RIE chamber for 15 seconds with 10 mTorr oxygen plasma at 300 W. Next, the top electrodes (“columns”) consisting of 15 nm Ti and 60 nm Pt were deposited and patterned using e-beam evaporation and liftoff. Finally, the wire bonding pads were formed by e-beam deposition of Cr (10 nm) / Ni (30 nm) / Au (500 nm). All lithographic steps were performed using a DUV stepper using a 248 nm laser. After fabrication and dicing, the dies were annealed in a reducing atmosphere (10% H_2, 90% N_2) for 30 minutes at 300°C.

A single dye was wire-bonded (with gold wires) to the DIP40 package, using a thermosonic bonding process. The process was simplified due to thicker Au metallization of the outer contacts, which also helped to reduce overall wire resistance. Figure S2a shows an optical image of a dye mounted onto the package. It shows, in particular, 6 gold wires bonded to the pads of each chip side, with a total of 12 wires for the columns (top and bottom sides of a crossbar) and 12 wires for the rows (left and right sides). Figure S2b shows an SEM image of the crossbar area of the chip.

4. Electrical characterization and forming of crossbar circuits

A detailed electrical characterization was performed on a 10×8 section of the crossbar, which was later utilized in the classification experiment. All electrical characterizations were performed using the Agilent B1500A parameter analyzer. In addition, the Agilent B5250A switching matrix was employed for testing packaged crossbar circuits and carrying out the pattern classification experiment. The parameter analyzer and the switching matrix were controlled by a personal computer via a GPIB interface using a custom C code. All write and read pulses were 500 μs long. For the memristor adjustment, we used the “V/2 scheme”, in which the selected rows and columns were voltage-biased at -V/2 and +V/2, respectively. For device state readout, we voltage-biased the selected column, connected the selected row to a virtual ground, and physically grounded all the other lines.

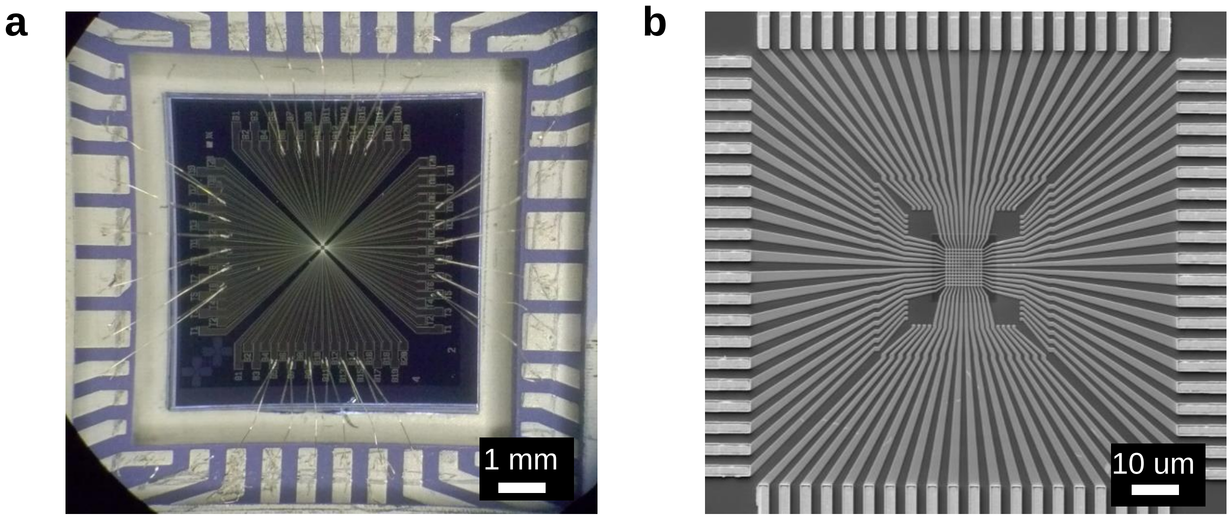


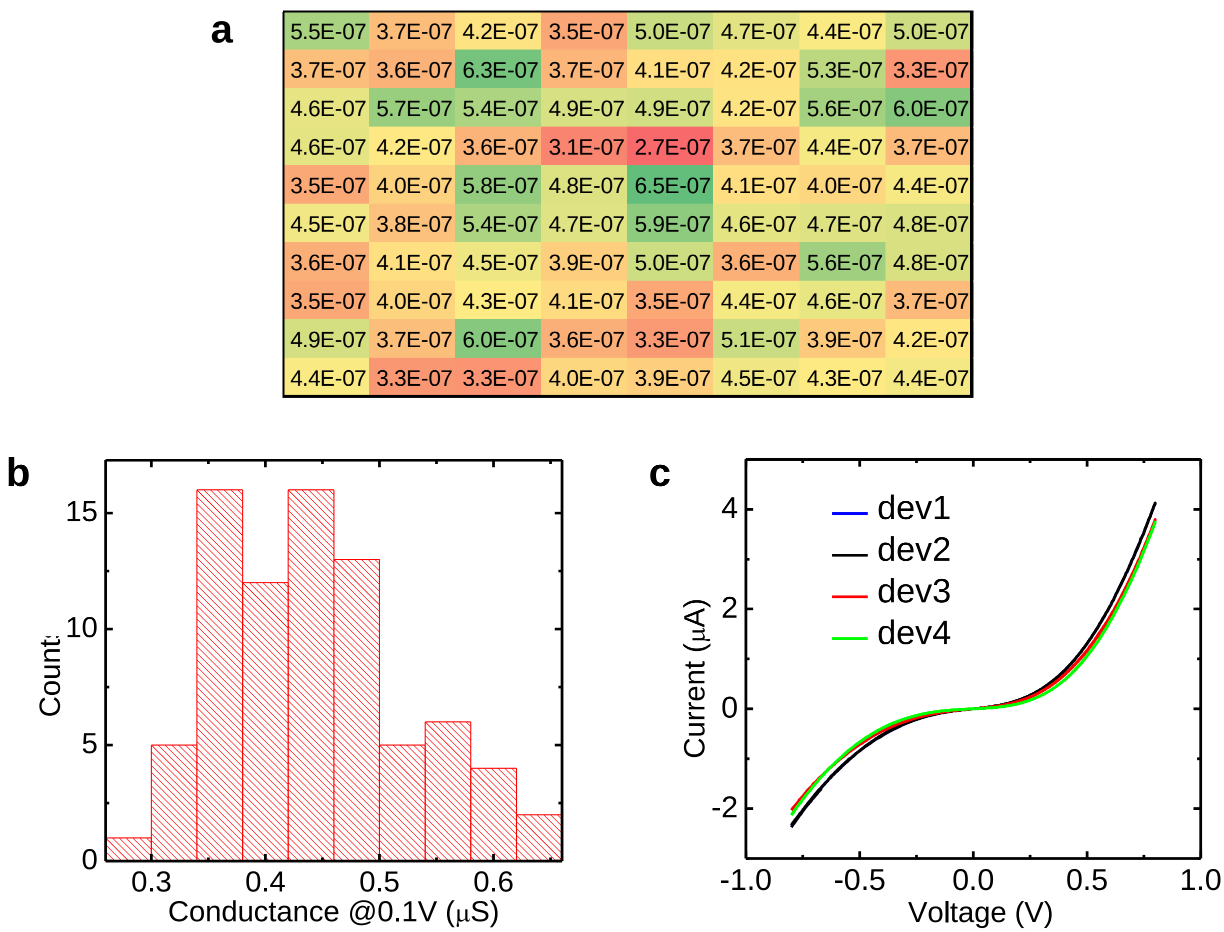
Figure S2. Microphotographs of the 12×12 crossbar with integrated Pt/Al_2O_3/TiO_{2-x}/Pt memristors: (a) the bonded chip, and (b) zoom-in on the crossbar area. Figure 1a in the main text shows the further zoom-in on the crossbar.

Specifically, we first characterized the "virgin" devices (not electroformed yet). Figures S3a and S3b show the recorded map of conductances measured at 0.1 V and the corresponding histogram, while Fig. S3c shows a representative set of dc *I*-*V* curves. After this characterization, the electroforming procedure was performed. For the forming, a quasi-DC current ramp was applied to the selected column line, while the selected row line was grounded and all the remaining (unselected) lines were kept floating. Such fixed-current technique prevents devices from an excessive stress during its electroforming, when device's resistance drops sharply. To minimize the current leakage during forming, the already formed devices were switched to their low-conductive state. More particularly, the devices of a 2×2 subarray were formed first. Then the devices in additional rows and columns were formed, so that the subarray of formed devices was gradually increased: first to 3×3 devices, then to 4×4, and so on. Figure S4a shows the map of forming voltages for the working section of the crossbar, while Fig. S4b shows the corresponding histogram. Additionally, Figure S4c shows the electroforming process dynamics (for the diagonal devices, which were formed last to complete forming of the corresponding subarray) on the [I, V] plane.

Following the electroforming procedure, we characterized the effective switching thresholds for all devices in the working section of the crossbar array (Fig. S5). The threshold set / reset voltages were measured by first programming devices to their high / low resistive states, and then applying a sequence of 500-μs pulses of the appropriate polarity with gradually increasing amplitude to measure the smallest voltage that caused a resistance change by more than 2 kΩ. The evolution of device conductance during the reset and set switching are shown in Figs. S5a and S5b, respectively, while Figs. S5c and S5d show maps of the corresponding

doi:10.1038/nature14441

effective reset and set threshold voltages. Panel (e) of Fig. S5 exhibits the data from its panels (c) and (d) in the form of histograms. The devices marked with 'X' in the threshold maps could not be switched with largest applied voltages.



5.5E-07	3.7E-07	4.2E-07	3.5E-07	5.0E-07	4.7E-07	4.4E-07	5.0E-07
3.7E-07	3.6E-07	6.3E-07	3.7E-07	4.1E-07	4.2E-07	5.3E-07	3.3E-07
4.6E-07	5.7E-07	5.4E-07	4.9E-07	4.9E-07	4.2E-07	5.6E-07	6.0E-07
4.6E-07	4.2E-07	3.6E-07	3.1E-07	2.7E-07	3.7E-07	4.4E-07	3.7E-07
3.5E-07	4.0E-07	5.8E-07	4.8E-07	6.5E-07	4.1E-07	4.0E-07	4.4E-07
4.5E-07	3.8E-07	5.4E-07	4.7E-07	5.9E-07	4.6E-07	4.7E-07	4.8E-07
3.6E-07	4.1E-07	4.5E-07	3.9E-07	5.0E-07	3.6E-07	5.6E-07	4.8E-07
3.5E-07	4.0E-07	4.3E-07	4.1E-07	3.5E-07	4.4E-07	4.6E-07	3.7E-07
4.9E-07	3.7E-07	6.0E-07	3.6E-07	3.3E-07	5.1E-07	3.9E-07	4.2E-07
4.4E-07	3.3E-07	3.3E-07	4.0E-07	3.9E-07	4.5E-07	4.3E-07	4.4E-07

Figure S3. Pre-formed (virgin sample) characterization of a 10×8 section of the crossbar: (a) color-coded conductance map of device conductances measured at 0.1 V, (b) the corresponding histogram, and (c) dc *I*-*V* curves of four representative devices. The average conductance and the standard deviation of conductances represented in panels (b) and (c) are 0.43 μS and 0.08 μS, correspondingly.

Of these data, the fact most important for applications is that the spread in effective switching voltages is narrow enough to avoid the infamous half-select problem.[6] For example, application of voltage +1.4 V to any device (besides those marked with X) ensures its set adjustment, while the half of that voltage (i.e. +0.7 V) is below the smallest observed set threshold voltage. This fact prevents disturbance of half-selected devices, connected to just one of voltage-biased lines. The switching threshold data were also important to identify a range of reasonable switching voltages, following two main criteria: balancing the set and reset dynamics and avoiding permanent damage of the devices and hence prolonging their life.

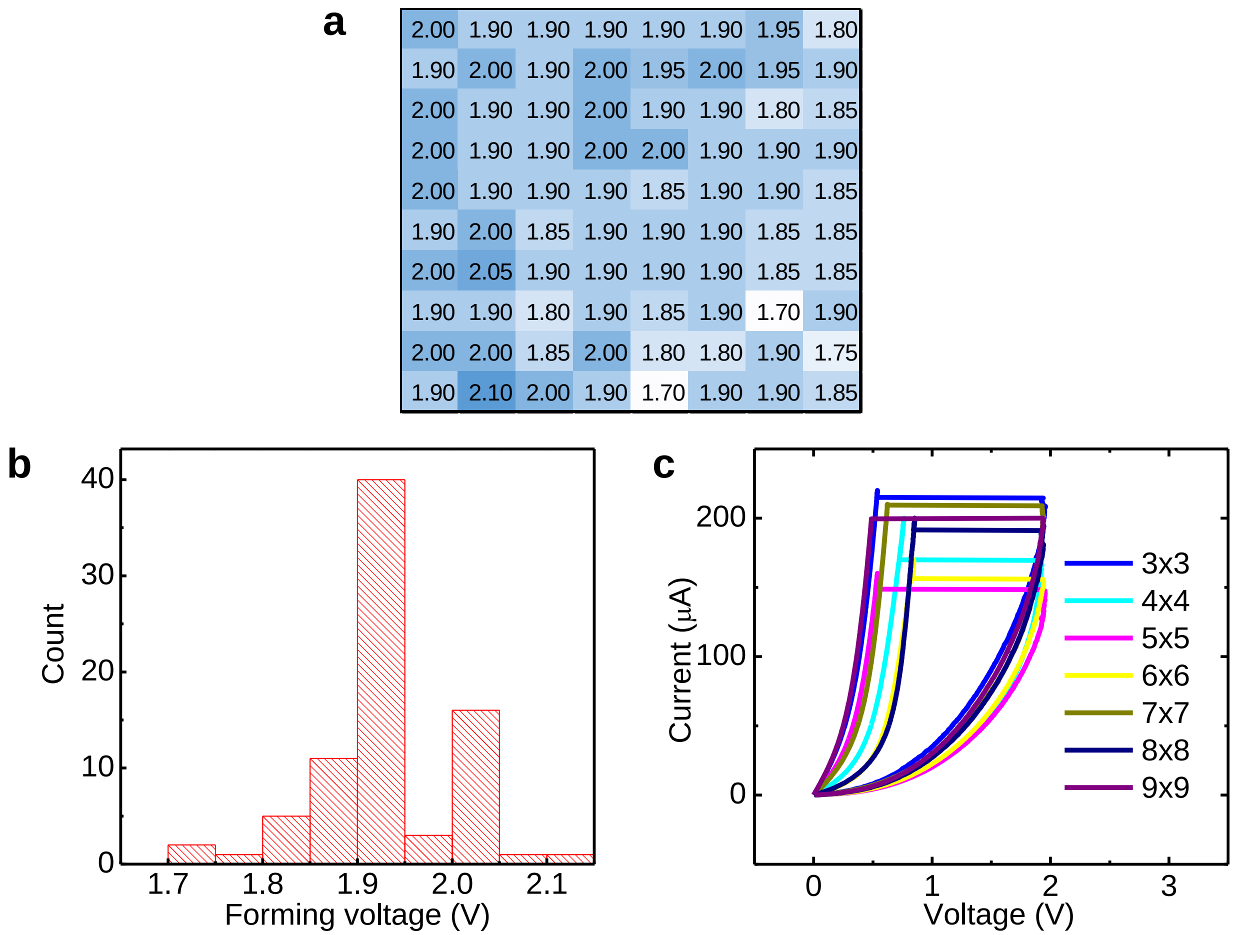
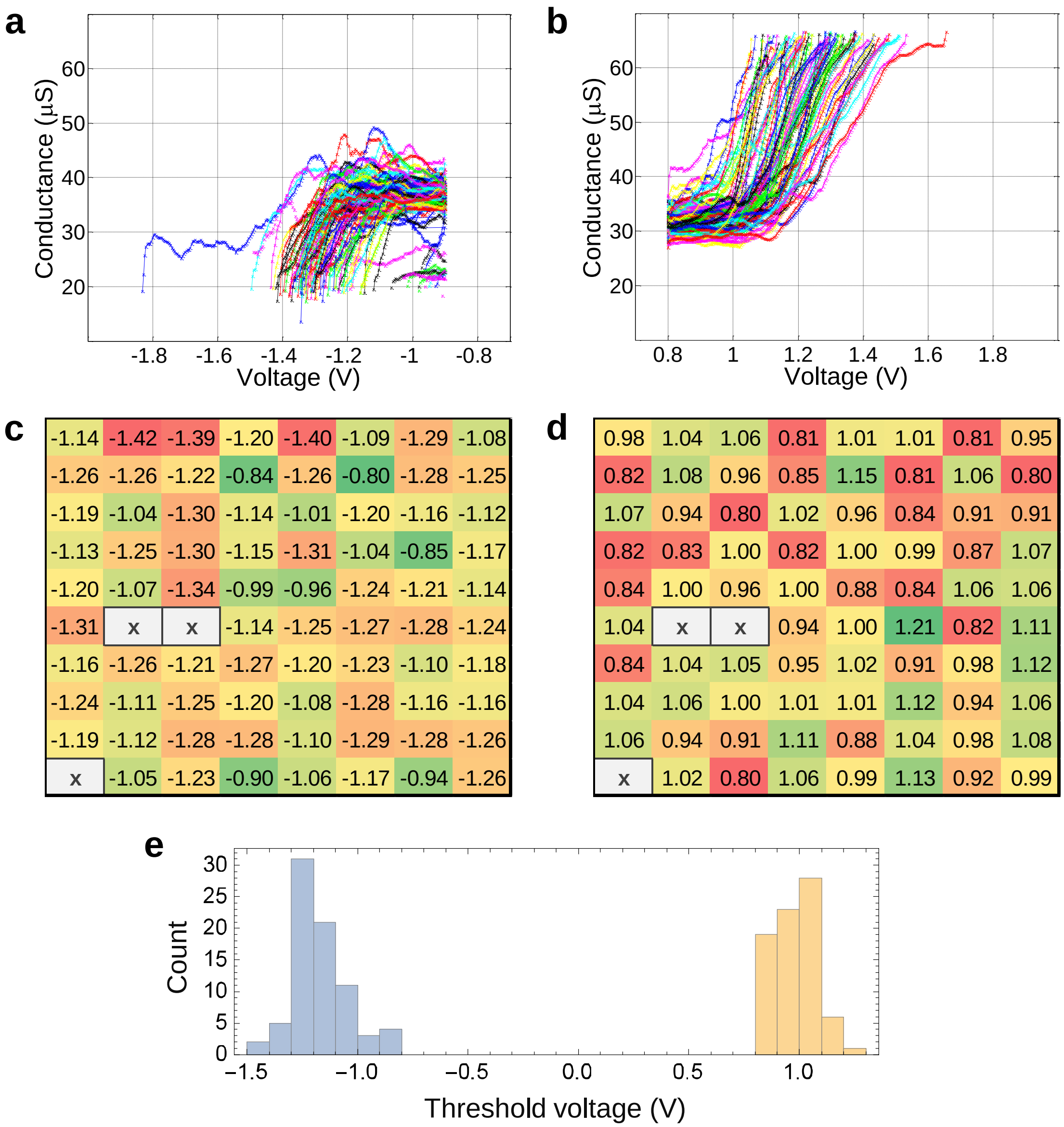


Figure S4. Characterization of memristor electroforming in a 10×8 section of the crossbar. (a) Color-coded map of the forming voltages, and (b) their histogram. (c) Typical forming switching curves of the last-formed devices in each partial array (of the size indicated in the legend). The average forming voltage and the standard deviation for the data shown on panels (b) and (c) are, respectively, 1.91 V and 0.07 V.

The switching behavior of crossbar devices was further characterized in stress conditions similar to those imposed by the used neuromorphic network training procedure. In particular, we have studied the evolution of device conductance due to a train of rectangular pulses of the same voltage magnitude. Figure S6 shows typical results of such an experiment, for a specific device. The experiment was repeated three times, with different pulse magnitudes, for both set and reset transitions. The plots show clearly a saturation effect: the conductance change under the effect of a single voltage pulse is gradually reduced (on the average), so that the conductance reaches a certain value for each particular pulse magnitude. (This effect is summarized in Fig. 1c of the main text, which shows the final conductance change as the function of the initial conductance.) The saturation effect needs to be taken into account when conducting the neural network experiment: as stated in the main text, it is much better, in the beginning of network training, to initialize devices in the middle of their dynamical range, to achieve substantial conductance changes in the beginning of the training process.



c

-1.14	-1.42	-1.39	-1.20	-1.40	-1.09	-1.29	-1.08
-1.26	-1.26	-1.22	-0.84	-1.26	-0.80	-1.28	-1.25
-1.19	-1.04	-1.30	-1.14	-1.01	-1.20	-1.16	-1.12
-1.13	-1.25	-1.30	-1.15	-1.31	-1.04	-0.85	-1.17
-1.20	-1.07	-1.34	-0.99	-0.96	-1.24	-1.21	-1.14
-1.31	x	x	-1.14	-1.25	-1.27	-1.28	-1.24
-1.16	-1.26	-1.21	-1.27	-1.20	-1.23	-1.10	-1.18
-1.24	-1.11	-1.25	-1.20	-1.08	-1.28	-1.16	-1.16
-1.19	-1.12	-1.28	-1.28	-1.10	-1.29	-1.28	-1.26
x	-1.05	-1.23	-0.90	-1.06	-1.17	-0.94	-1.26

d

0.98	1.04	1.06	0.81	1.01	1.01	0.81	0.95
0.82	1.08	0.96	0.85	1.15	0.81	1.06	0.80
1.07	0.94	0.80	1.02	0.96	0.84	0.91	0.91
0.82	0.83	1.00	0.82	1.00	0.99	0.87	1.07
0.84	1.00	0.96	1.00	0.88	0.84	1.06	1.06
1.04	x	x	0.94	1.00	1.21	0.82	1.11
0.84	1.04	1.05	0.95	1.02	0.91	0.98	1.12
1.04	1.06	1.00	1.01	1.01	1.12	0.94	1.06
1.06	0.94	0.91	1.11	0.88	1.04	0.98	1.08
x	1.02	0.80	1.06	0.99	1.13	0.92	0.99

Figure S5. Characterization of the set and reset thresholds of formed memristors: (a, b) device conductance dynamics under applied voltage ramps of opposite polarities; (c, d) the effective threshold voltage maps for set and reset transitions, and (e) their histograms. The histograms on panel (e) do not include the 3 devices that could not be switched to the OFF state with voltages above -1.5 V; these devices are marked with crosses on panels (c) and (d). The average effective set/reset switching threshold voltages and their standard deviations, for the data shown on panel (e) are, respectively 0.9 V/-1.17 V and 0.1 V/0.12 V.

doi:10.1038/nature14441

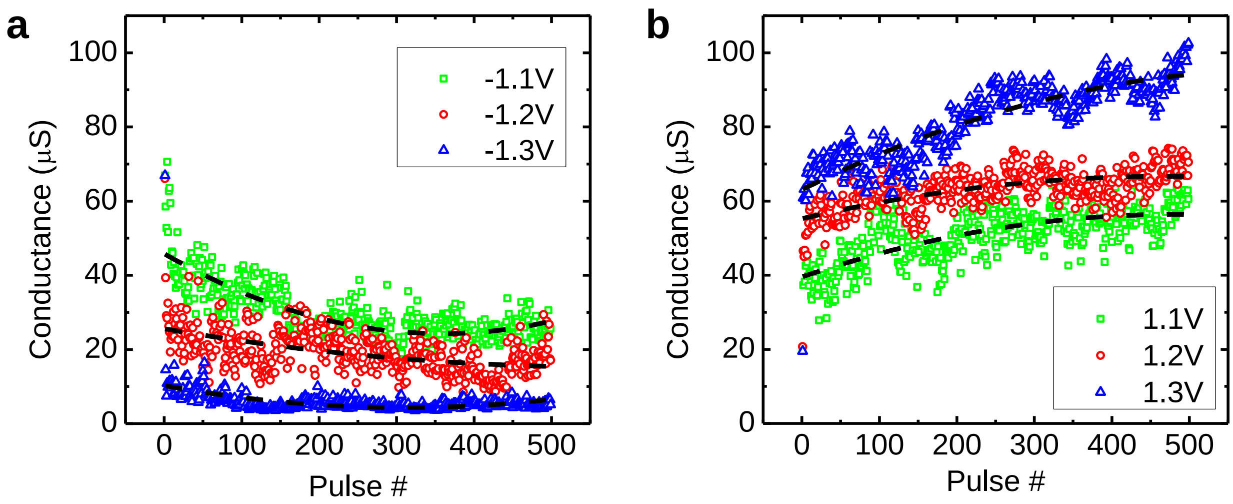


Figure S6. Evolution of memristor's conductance (measured at 0.1 V) under the effect of 500-μs pulse trains of several magnitudes, for the (a) reset and (b) set switching. Note that unlike Figures S5a and S5b, this figure shows the change in conductance for a fixed-magnitude pulses applied repeatedly to the same device. The dashed lines are just guides for the eye.

5. Network training result analysis

The key metric of neural network classifier training is the difference it creates between network outputs induced by input patterns belonging to different classes. Two panels of Figure S7a show such difference for our memristor crossbar network. Namely, these are the histograms of the differences between currents in the "correct" and "incorrect" outputs for each of 30 input patterns. (Evidently, there are 60 such differences: for each of 30 inputs in 3 classes, there is one correct output and two incorrect ones.) In a perfectly trained classifier, all these differences have to be positive. The top panel shows that before training, the differences are small, and have random signs. During the training the differences increased, so that, as the lower panel shows, after a sufficient number (in this particular case, 21) of training epochs, all of them have become positive, signaling the 100% classification fidelity.

Figure S7b shows another useful way of understanding network dynamics during its training, namely the distributions of conductances at three phases of training process. (Note that the same information, but for only two phases, is presented in the inset of Fig. 4a of the main text.) For convenience, Fig. S7c shows the distributions of the effective weights *W*, i.e. the differences between the conductances of devices that form each differential pair, for the same three phases of training. As the data show, the difference of the weight distributions created by the training is less dramatic than that of the output signals, shown in panel (a), i.e. the training imposes mostly relative weight changes (which are important for the correct classification), rather than their global change.

doi:10.1038/nature14441

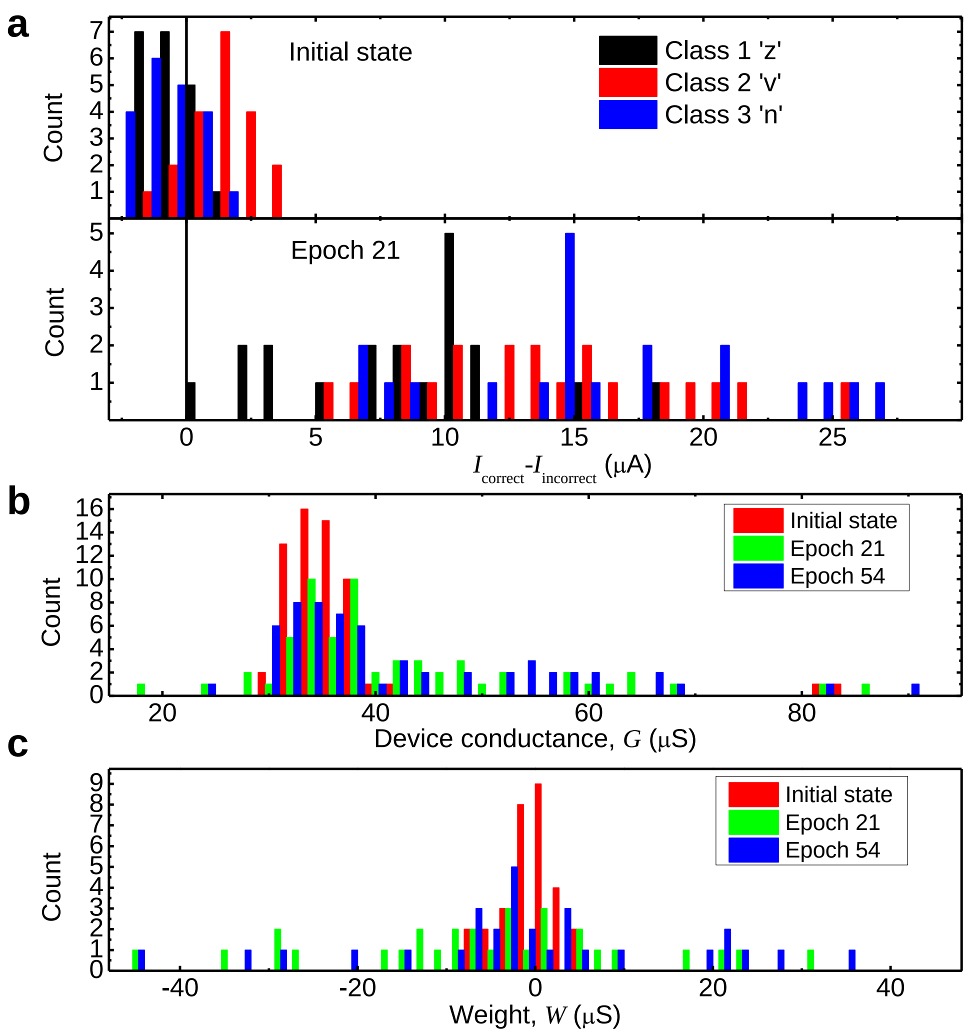


Figure S7. Additional data from a particular pattern classification experiment ("Run 1") for which the perfect classification has been achieved, for the first time, after 21 training epochs. (a) Histogram of differences between currents in the "correct" output (corresponding to input pattern's class) and other two channels. (b) Histogram of device conductances at the initial state, measured after training epoch 21, and after epoch 54 (for which classification is also perfect). (c) Histogram of the differential weights W defined by Eq. (3) of main text. In panel (b), the average values / standard deviations of the measured conductances are 36.3 μS / 9 μS, 41.9 μS / 13 μS, and 42.4 μS / 13.4 μS, when measured in the initial state, after epoch 21, and epoch 54, respectively. The corresponding averages / deviations of the weights in panel (c) are -0.24 μS / 2.83μS, -5.36 μS / 16.8 μS, and -1.17 μS / 17.1 μS.

Finally, Figures S8 and S9 show the weight maps measured before and after training for 6 separate training experiments ("runs"). (The convergence graphs for these experiments are shown in Fig. 4a of the main text). The comparison of weight maps in Fig. S8 shows that even though the initialization procedure was similar in all runs, there were some run-to-run variations of the state of same device. (To quantify the variations, Figs. S8b and S9b show relative standard deviations for each of 30 weights.)

doi:10.1038/nature14441

a

Run 1

z	v	n
-1.62E-06	-1.30E-06	-1.78E-06
3.75E-06	2.52E-06	-4.46E-06
-2.36E-06	6.50E-07	-6.15E-06
2.87E-06	-2.41E-06	-3.00E-07
-4.73E-06	3.30E-07	1.93E-06
7.50E-07	1.59E-06	-6.23E-06
8.60E-07	4.16E-06	-2.80E-07
-2.85E-07	1.72E-06	-3.18E-06
-8.85E-07	-1.65E-06	2.65E-07
5.19E-06	1.60E-06	2.24E-06

Run 2

z	v	n
-3.98E-06	-1.22E-06	-1.66E-06
1.42E-06	1.50E-07	-7.87E-06
-3.10E-06	6.49E-08	-6.25E-06
1.97E-06	-4.31E-06	-3.41E-06
-5.35E-06	-1.49E-08	1.49E-06
3.31E-06	-5.65E-07	-9.75E-06
-9.75E-07	3.16E-06	-1.37E-06
-5.45E-07	1.46E-06	7.00E-08
-1.35E-06	-3.65E-07	-3.80E-07
1.36E-05	3.69E-06	1.12E-06

Run 3

z	v	n
-2.57E-06	-1.61E-06	-3.35E-06
3.01E-06	-3.95E-06	4.82E-06
2.70E-07	1.62E-06	-2.30E-07
-6.55E-07	-3.87E-06	-2.11E-06
-5.46E-06	3.12E-06	-1.20E-07
4.60E-07	-4.89E-05	6.10E-07
1.33E-06	-3.25E-06	4.26E-06
3.20E-07	-1.70E-06	1.09E-06
-5.51E-08	2.95E-06	-3.77E-06
5.50E-06	-5.87E-06	1.33E-06

Run 4

z	v	n
-2.35E-07	-5.67E-06	-2.38E-06
-1.05E-07	-4.06E-06	7.04E-06
-1.20E-07	5.15E-07	-1.96E-06
3.16E-06	-7.09E-06	9.50E-08
-4.36E-06	2.78E-06	-1.44E-06
1.39E-06	-4.91E-05	-1.15E-06
7.15E-07	-1.20E-06	3.01E-06
1.42E-06	-2.43E-06	1.16E-06
-9.95E-07	-1.30E-06	-9.65E-07
4.37E-06	-4.78E-06	1.33E-06

Run 5

z	v	n
-9.15E-07	-1.13E-06	-7.15E-06
-3.26E-06	-1.85E-06	-3.50E-06
1.63E-06	-2.60E-06	-2.66E-06
1.26E-06	-5.10E-07	-4.58E-06
-6.36E-06	4.19E-06	-5.80E-07
-6.92E-06	-4.61E-05	-5.60E-07
0.00E+00	3.50E-07	2.68E-06
3.61E-06	-2.52E-06	8.00E-08
-1.24E-06	1.27E-06	-1.44E-06
6.56E-06	-5.67E-06	1.12E-06

Run 6

z	v	n
-4.06E-06	4.10E-07	-5.77E-06
-1.74E-06	-7.60E-06	-1.41E-06
1.17E-06	-2.28E-06	1.51E-08
4.98E-06	-5.17E-06	-1.38E-06
-4.55E-06	4.91E-06	-1.96E-06
-1.28E-05	-4.60E-05	5.50E-08
-1.36E-06	1.23E-06	2.18E-06
8.50E-07	-1.41E-06	3.55E-07
-1.42E-06	-2.77E-06	1.77E-06
1.12E-05	-3.66E-06	-6.35E-07

b

0.6	1.1	0.6
4.8	1.3	5.8
4.2	4.6	0.9
0.8	0.5	0.8
0.1	0.7	12.3
2.5	0.7	1.3
10.1	3.4	1.1
1.5	2.2	21.1
0.5	6.2	2.2
0.4	1.5	0.8

Figure S8: Initial weight values: (a) the maps of initial weights W_{ij} (in siemens) measured in 6 separate training experiments ("runs"), and (b) the corresponding relative variation, i.e. (standard deviation)-to-(average value) ratio for each weight. On panel (a), red and blue background colors indicate, respectively, negative and positive values of W_{ij}, while on panel (b), the colors are used just to emphasize lower (greenish) and higher (reddish) values. Narrow columns with gray / white cells show positive / negative input signals from noiseless images of the classes corresponding to the particular synaptic columns - see Fig. 3b in the main text.

a

Run 1

z	v	n
9.72E-06	1.61E-05	-4.48E-05
5.65E-06	2.32E-05	-2.08E-06
-6.41E-06	6.31E-06	3.12E-05
2.02E-05	-6.32E-06	-9.09E-06
1.96E-06	-9.87E-06	-2.93E-05
-3.96E-06	1.56E-06	-1.18E-05
9.85E-07	4.09E-06	-2.91E-05
-5.28E-06	-1.65E-05	-1.38E-05
-6.55E-07	-3.55E-05	-1.49E-05
-2.63E-06	-1.33E-05	-2.64E-05

Run 2

z	v	n
5.84E-06	8.16E-06	-2.48E-05
2.40E-07	1.24E-05	1.10E-05
-3.29E-06	-2.85E-07	2.54E-05
1.06E-05	-1.44E-05	-6.10E-07
-5.11E-06	-2.07E-06	-5.95E-06
1.74E-06	-2.76E-06	-1.58E-05
-1.93E-06	4.44E-06	-2.69E-05
-2.77E-06	-5.96E-06	-1.44E-05
1.21E-06	-2.95E-05	-2.80E-05
2.78E-06	-1.54E-05	-3.13E-05

Run 3

z	v	n
6.89E-06	-6.34E-06	1.93E-05
-7.47E-06	-2.06E-06	1.39E-05
-1.57E-05	1.69E-06	-1.05E-06
9.32E-06	7.05E-07	-1.97E-05
-5.17E-06	-1.02E-06	5.95E-07
-1.59E-06	-4.19E-05	-6.25E-06
-2.38E-06	-1.66E-06	4.29E-06
-4.42E-06	1.21E-05	-1.57E-05
1.50E-07	2.88E-05	-3.44E-05
5.61E-06	2.32E-06	-7.50E-06

Run 4

z	v	n
-5.66E-06	-4.15E-06	1.39E-06
-1.73E-05	1.35E-05	5.78E-06
1.12E-06	7.05E-06	-9.03E-06
6.34E-06	1.28E-05	-1.58E-05
9.08E-06	-1.84E-06	-7.69E-06
-9.35E-06	-4.30E-05	-3.99E-06
-5.02E-06	7.84E-06	3.38E-06
-6.58E-06	1.49E-05	-1.32E-05
3.59E-06	2.41E-05	-2.90E-05
2.40E-06	-7.90E-07	-7.01E-06

Run 5

z	v	n
-1.00E-06	5.02E-06	-1.59E-05
-2.74E-05	8.73E-06	2.03E-05
-2.05E-05	6.95E-06	-1.43E-05
-3.27E-06	2.02E-05	-2.78E-05
-6.52E-06	1.41E-05	-1.44E-05
-8.91E-06	-4.56E-05	-3.92E-06
-1.63E-05	8.44E-06	8.70E-06
-2.11E-05	2.60E-05	-2.29E-05
-3.46E-06	2.69E-05	-3.39E-05
6.93E-06	2.26E-06	-4.97E-06

Run 6

z	v	n
2.58E-06	1.18E-05	-4.17E-06
-2.12E-05	1.37E-05	1.03E-05
3.60E-07	7.45E-06	-1.31E-05
3.11E-06	1.42E-05	-2.78E-05
-7.16E-06	1.02E-05	-1.00E-05
-7.39E-06	-4.25E-05	-1.09E-05
-1.34E-05	1.23E-05	2.06E-06
-2.52E-05	3.44E-05	-2.60E-05
1.31E-06	3.66E-05	-3.89E-05
-7.14E-06	1.46E-05	-2.48E-05

b

1.7	1.6	1.8
1.0	0.6	0.7
1.1	0.6	5.8
0.9	2.7	0.6
2.7	5.1	0.8
0.8	0.7	0.5
1.0	0.7	2.5
0.8	1.6	0.3
6.0	3.4	0.3
3.7	5.9	0.6

Figure S9: Final synaptic weight values after the network training process convergence: (a) the maps of weights W_{ij} (in siemens) measured after the perfect classification has been reached, for 6 separate runs (after, respectively, 21, 6, 33, 26, 35, and 18 training epochs for Runs 1, 2, 3, 4, 5, and 6 – see Fig. 4a); (b) the corresponding relative variation for each weight. The color coding is the same as in Fig. S8.

As Fig. S9 shows, after training the weights become markedly different - not only in their magnitude, but in many cases also in their sign, despite the fact that each final distribution provides perfect classification fidelity. This is in agreement with the well-known fact that the classifier training using the Delta Rule algorithm and its cousins (such as the Manhattan Update Rule used in this work) does not result in a unique synaptic weight distribution – unless the initial weight values are exactly the same in each training run.

6. Computer simulations

We have also carried out extensive computer simulations of our neural network, in order to understand the impact of various parameters, most importantly of the initial device conductances, on classifier's fidelity and convergence speed. To perform the simulation, an approximate device model was derived from the data shown on Fig. 1c of the main text. In this model, the conductance increase (ΔG_{set}) by a positive voltage pulse and its decrease ($\Delta G_{\mathrm{reset}}$) by the negative pulse are calculated as

$$\Delta G_{\mathrm{set}}(G) = +10^{-3}(10^6 G - 10^6 G_{\mathrm{min}} + 10^{\,vset\,/\,slope})^{-slope}, \qquad \text{(S1)}$$

$$\Delta G_{\mathrm{reset}}(G) = -\,10^{-3}(-10^6 G + 10^6 G_{\mathrm{max}} + 10^{\,vreset\,/\,slope})^{-slope}, \qquad \text{(S2)}$$

where constant *slope,* for our specific pulse amplitude and duration was taken equal to 2, while parameters *vset* and *vreset* were randomly chosen from the range [1, 5.5] for every memristor before each run. Constants G_{min} = 10 μS and G_{max} = 100 μS are the minimum and maximum conductance values; G is always clipped between these values after each update. Such simple model captures two main features of the memristors, namely the saturation of the switching dynamics (Fig. 1c) and switching threshold variations (Fig. S5).

Fig. S10 shows the most important results of these simulations. First, they have confirmed that the classifier fidelity and convergence speed are the best when the initial conductances of the devices are in the middle of their dynamical range. The second important simulation result is that the performance is rather insensitive to the choice of parameter β, with the optimal value close to $\beta = 2\times10^5$. As was stated in the main text, these results had been used at network training.

7. Toward practical applications

We believe that our work is the first proof-of-concept that passively integrated memristive crossbar circuits can be used to perform classification task. However, due to its small size, this network is not by itself practical, and several major steps have to be made toward larger, useful neuromorphic networks.

Besides the primary, self-evident task of fabrication much larger crossbars with smaller memristors of (at least) similar quality and reproducibility, there is also a challenge of their efficient training. For a large crossbar, the batch training algorithm described in the paper would come with a substantial circuit overhead, if it is fully implemented on the same chip as the network. In fact, in the batch mode, the training circuit has to hold at least ~$M_1 \times M_2$ continuous intermediate values, for example numbers Δ_{ij}, participating in Eq. (5), until the next weight update. (Here $M_{1,2}$ are the linear sizes of the crossbar array.) Such training circuit, implemented in the usual CMOS technology, would have a much larger chip footprint than the crossbar itself.

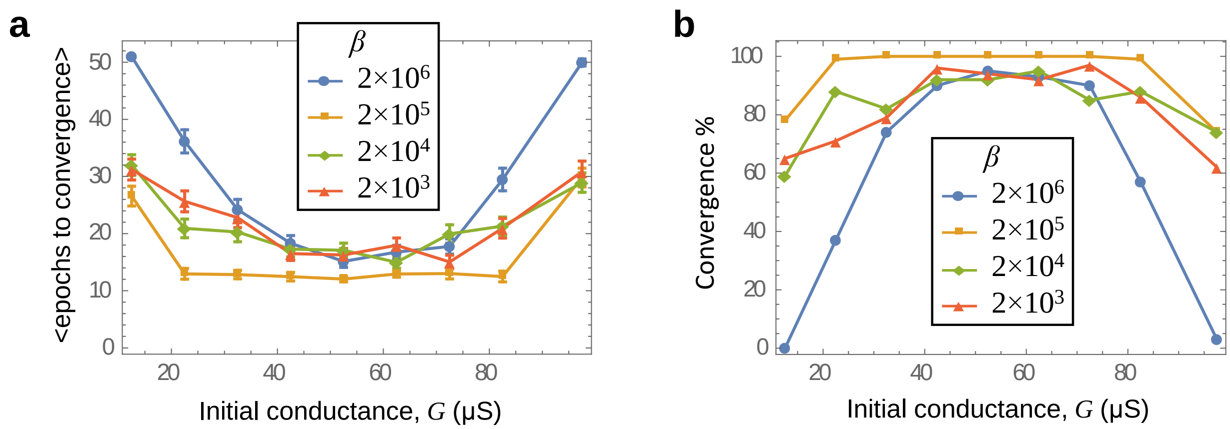


Figure S10: Major results of computer simulation of our neural network, using a realistic memristor model. (a) The number of training epochs required to achieve the perfect classification, and (b) the fraction of training experiments with the perfect convergence, as functions of the initial memristor conductances. Each point is an average over 100 runs; for each run the weights were randomly initialized within a 5-μS conductance window around the value indicated on the horizontal axis. If an experiment took more than 50 epochs for convergence, for the purposes of panel (b), it was considered a failure. These "failed" experiments were excluded when calculating the error bars on panel (a).

However, several factors make this task less hopeless than it may look. First of all, for most important practical applications, a neural network is used repeatedly for classification of patterns of the same type. In these cases it needs to be trained only in the beginning, and then may be used with the same set of weights for a long time. (In fact, this is the approach used in most advanced recent demonstrations of neural network hardware chips.[7,8]) Such rare training may be assisted by external (digital) computers which, in particular, would store all the intermediate values. According to our recent results,[9] this approach has substantial advantages over a purely ex-situ training in a “precursor” software network, with the subsequent synaptic weight import into the crossbar (as discussed, e.g., in Ref. 2 of the main text), because the former procedure allows to mitigate detrimental effects of memristor and also neuron circuits variability.

Another opportunity is to train the network in situ, using a local online (“stochastic”) training procedure, for which the requirements to external memory would be substantially reduced. In this case, just one circuit per each line and column of the memristive crossbar, i.e. ~(M_1+M_2) circuits per an $M_1 \times M_2$ memristors, may be sufficient to implement all training and neuron functions. As the crossbar is scaled up in future for more complex tasks (up to $M_{1,2}$ ~10^3-10^4 deemed necessary for some cognitive architectures), the relative size of this circuit overhead would be much decreased. Figure S11 shows some results of our preliminary attempt at implementing online training algorithm using the same memristive crossbar. Perfect classification has been achieved for the used small set of input patterns, but unfortunately, in its current form the online learning requires too many synaptic updates to achieve the perfect network performance.

Another important question which must be addressed in practical application is the generalization performance of the classifier. As Fig. S11 shows, for online training we could reach a perfect classification performance on a small separate test set. However, this training was too slow for our basic experiments with 30 images in 3 classes. As stated in the main text, these experiments were carried out without a separate test set, but using the measured values of weights recorded after every training epoch, we could use computer simulations to estimate network's generalization ability, i.e. its possible classification performance on a separate test set (not used at training). The set consisted of all possible patterns (36 patterns in each of 3 classes) obtained from the ideal images (Fig. 2c of main text) by flipping 2 pixels – see Fig. S12b. Though the classification improved during training (Fig. S12a), typically it is not perfect by its end. This is not surprising, giving the fact that the minimum Hamming distances between patterns of the test and training sets are 2 and 4, accordingly. Indeed, even a human would hardly be able to classify all these patterns correctly – please have one more look at Fig. S12b.

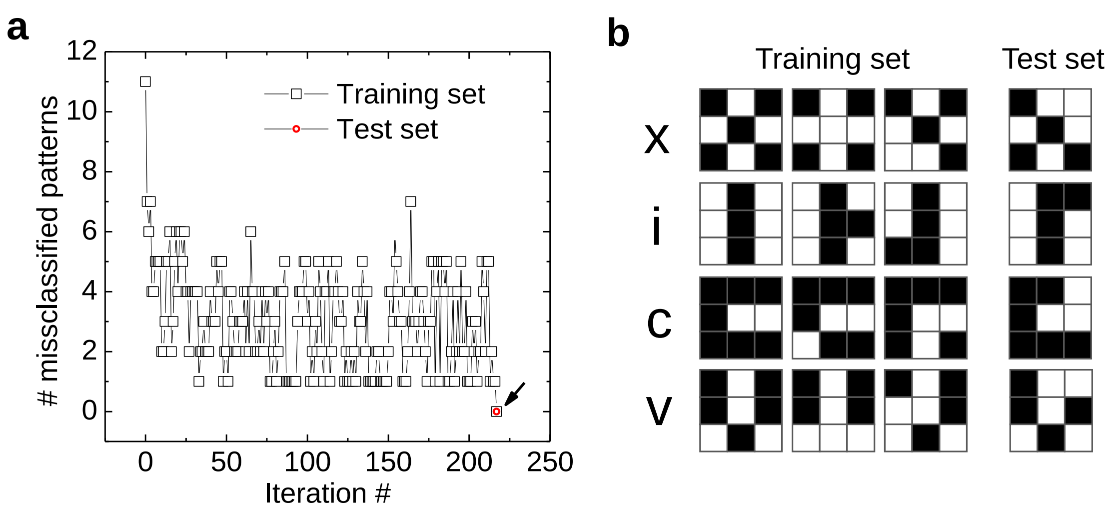


Figure S11: Preliminary results for pattern classification experiment using online ("stochastic") Manhattan Rule training: (a) Classification convergence for the training and test sets, and (b) used sets. On panel (a), "iteration" means an application of one pattern from a training set, followed by the corresponding weight update. (There were 12 iterations in training epoch.) Note that here perfect classification was achieved on the (admittedly, extremely small) test set. (The evaluation on this set was performed only at the end of training.)

As a result, we believe that a convincing demonstration of the generalization ability is only possible on larger input vectors, which in turn require larger networks (and hence larger crossbar arrays). Such a demonstration remains one of our major future goals.

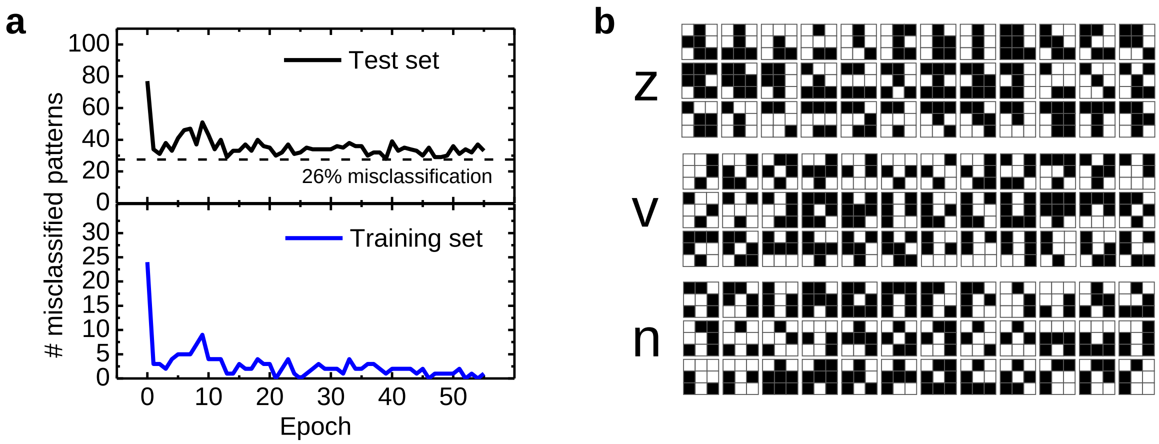


Figure S12: Pattern classification experiment (Run 1): (a) Classification convergence for the training and test sets, and (b) the used test set. (The training set and the training method were the same as described in main text). The bottom figure on panel (a) shows one of the curves (black one) from Fig. 4a, extended to the subsequent training epochs.

References

1. Young-Fisher, K. G. *et al*. Leakage current-forming voltage relation and oxygen gettering in HfO_x RRAM devices. *IEEE Electron Device Lett.* **34,** 750–752 (2013).
2. Yang, J. J. *et al*. The mechanism of electroforming of metal oxide memristive switches. *Nanotechnology* **20,** 215201 (2009).
3. Lentz, F. *et al*. Current compliance-dependent nonlinearity in TiO_2 RRAM. *IEEE Electron Device Lett.* **34**, 996–998 (2013).
4. Goux, L. *et al.* Ultralow sub-500nA operating current high-performance $TiN/Al_2O_3/HfO_2$ /Hf/TiN bipolar RRAM achieved through understanding-based stack-engineering. *VLSI Symp. Technol.'12*, 159–160 (2012).
5. Wu, H. *et al.* Resistive switching performance improvement of Ta_2O_{5-x}/TaO_y bilayer ReRAM devices by inserting AlO_δ barrier layer. *IEEE Electron Device Lett.* **35**, 39–41 (2014).
6. Strukov, D. B., Likharev, K. K. Reconfigurable nano-crossbar architectures. *Nanoelectronics and Information Technology*, Waser, R. (ed.), 3rd ed. (Wiley, Weinheim, Germany, 2012).
7. Merolla, P. A. *et al*. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668-673 (2014).

8. Farabet, C. *et al*. NeuFlow: A runtime reconfigurable dataflow processor for vision. *CVPRW'11*, 109-116 (2011).
9. Kataeva, I. *et al.* Efficient training algorithms for neural networks based on memristive crossbar circuits. *submitted to IJCNN'15*, (2015), available online at https://www.ece.ucsb.edu/~strukov/papers/2015/mlpclassifier.pdf .