

# Classification with a disordered dopant-atom network in silicon

<https://doi.org/10.1038/s41586-019-1901-0>

Received: 4 December 2018

Accepted: 13 November 2019

Published online: 15 January 2020

Tao Chen<sup>1</sup>, Jeroen van Gelder<sup>1</sup>, Bram van de Ven<sup>1</sup>, Sergey V. Amitonov<sup>1</sup>, Bram de Wilde<sup>1</sup>, Hans-Christian Ruiz Euler<sup>1</sup>, Hajo Broersma<sup>2</sup>, Peter A. Bobbert<sup>1,3</sup>, Floris A. Zwanenburg<sup>1</sup> & Wilfred G. van der Wiel<sup>1\*</sup>

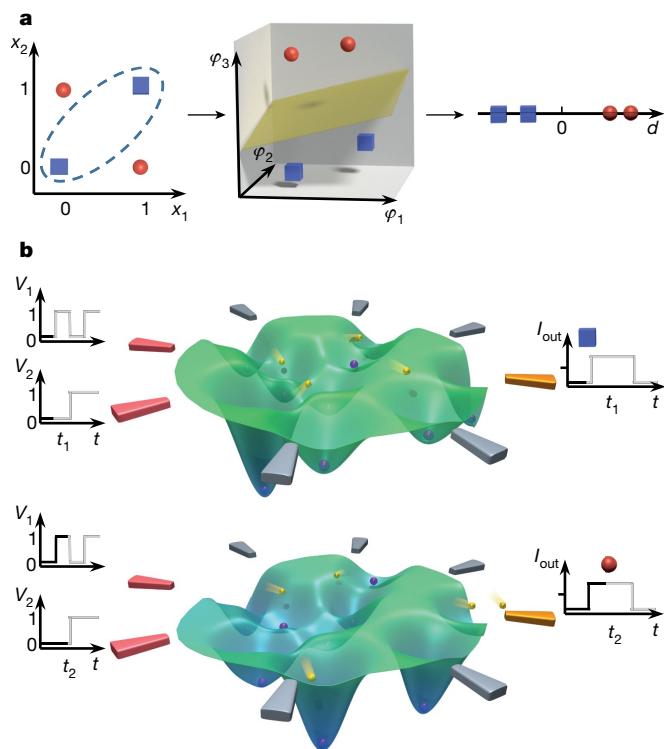
Classification is an important task at which both biological and artificial neural networks excel<sup>1,2</sup>. In machine learning, nonlinear projection into a high-dimensional feature space can make data linearly separable<sup>3,4</sup>, simplifying the classification of complex features. Such nonlinear projections are computationally expensive in conventional computers. A promising approach is to exploit physical materials systems that perform this nonlinear projection intrinsically, because of their high computational density<sup>5</sup>, inherent parallelism and energy efficiency<sup>6,7</sup>. However, existing approaches either rely on the systems' time dynamics, which requires sequential data processing and therefore hinders parallel computation<sup>5,6,8</sup>, or employ large materials systems that are difficult to scale up<sup>7</sup>. Here we use a parallel, nanoscale approach inspired by filters in the brain<sup>1</sup> and artificial neural networks<sup>2</sup> to perform nonlinear classification and feature extraction. We exploit the nonlinearity of hopping conduction<sup>9–11</sup> through an electrically tunable network of boron dopant atoms in silicon, reconfiguring the network through artificial evolution to realize different computational functions. We first solve the canonical two-input binary classification problem, realizing all Boolean logic gates<sup>12</sup> up to room temperature, demonstrating nonlinear classification with the nanomaterial system. We then evolve our dopant network to realize feature filters<sup>2</sup> that can perform four-input binary classification on the Modified National Institute of Standards and Technology handwritten digit database. Implementation of our material-based filters substantially improves the classification accuracy over that of a linear classifier directly applied to the original data<sup>13</sup>. Our results establish a paradigm of silicon-based electronics for small-footprint and energy-efficient computation<sup>14</sup>.

Doping is a crucial process in semiconductor electronics, where impurity atoms are introduced to modulate the charge carrier concentration. Conventional semiconductor devices operate in the band regime of charge transport, where the delocalization of the charge carriers gives rise to high mobility and a linear response to an applied electric field. At sufficiently low doping concentration and temperature<sup>9,15</sup>, however, delocalization is lost, and carriers move sequentially from dopant atom to dopant atom. This is referred to as the hopping regime<sup>10,11,16</sup>, which exhibits higher resistivity and nonlinearity. Nonlinearity is often undesired, but it is a valuable asset for unconventional computing, that is, for systems that do not follow the Turing model of computation<sup>6–8,17–19</sup>. Rather than excluding nonlinearity, we can exploit it<sup>12</sup> and manipulate our physical system with artificial evolution to solve computational problems<sup>17</sup>. This evolution in matter has been used, for example, for frequency distinguishing by liquid crystals<sup>18</sup> and robot control with carbon nanotubes<sup>19</sup>. We recently showed that a disordered network of gold nanoparticles acting as single-electron transistors can be evolved

into any Boolean logic gate at sub-kelvin temperatures<sup>12</sup>. By exploiting the physics of materials for computation at the nanoscale through evolution, we may realize systems with unprecedented computational density and efficiency that are too complex to design<sup>20</sup>.

Here, we fundamentally advance our previous work<sup>12</sup> by expanding the functionality, exploiting the well established platform of silicon technology and demonstrating operation up to room temperature. According to Cover's theorem<sup>4</sup>, complex, linearly inseparable classification problems, when nonlinearly and sparsely mapped to a higher-dimensional space, can transform into linearly separable problems. The essence of this nonlinear mapping is illustrated in Fig. 1a for the XOR classification problem. To save resources, this projection is often done implicitly by using kernel functions in machine learning, that is, without explicit computation of high-dimensional coordinates<sup>3</sup>. In artificial neural networks (ANNs), the nonlinear projection is learned by adjusting internal weights, traditionally through back-propagation, leading to powerful feature extractors<sup>2</sup>. However, emulating ANNs with

<sup>1</sup>NanoElectronics Group, MESA+ Institute for Nanotechnology and BRAINS Center for Brain-Inspired Nano Systems, University of Twente, Enschede, The Netherlands. <sup>2</sup>Programmable Nanosystems and Formal Methods and Tools, MESA+ Institute for Nanotechnology, DSI Digital Society Institute and BRAINS Center for Brain-Inspired Nano Systems, University of Twente, Enschede, The Netherlands. <sup>3</sup>Molecular Materials and Nanosystems and Center for Computational Energy Research, Department of Applied Physics, Eindhoven University of Technology, Eindhoven, The Netherlands. \*e-mail: W.G.vanderWiel@utwente.nl



**Fig. 1 | Simplifying classification by nonlinear projection.** **a**, In the XOR classification problem two classes of data (red circles for  $(1, 0)$ ,  $(0, 1)$  and blue squares for  $(0, 0)$ ,  $(1, 1)$ ) cannot be linearly separated in two dimensions ( $x_1, x_2$ ; left). When nonlinearly transformed to three dimensions ( $\varphi_1, \varphi_2, \varphi_3$ ; middle), the data can be linearly separated according to their distances  $d$  (right) to a decision boundary (yellow plane in the middle panel). **b**, Schematic representation of the potential landscape of the dopant network. In the hopping regime, the potentials of  $N$  dopants (purple spheres) span a high-dimensional feature space. Yellow spheres represent charge carriers. The voltage–time ( $V$ – $t$ ) diagrams on the left schematically show the voltage combinations applied to the input electrodes (red), affecting the potential landscape and projecting information nonlinearly to the feature space. Note the difference between the potential landscapes in the top and bottom panels for different input voltages. The characteristics of the output current (yellow electrode) are tunable by the control voltages (grey electrodes).

conventional complementary metal–oxide–semiconductor (CMOS) technology is known to be power-inefficient<sup>21</sup>, and CMOS scaling is not keeping pace with ANNs<sup>14</sup>. To avoid the area and power costs of emulating neurons and synapses, reconfigurable<sup>2</sup> material systems with intrinsic complexity and diversity of nonlinear operations<sup>6,22,23</sup> are strongly sought after.

Our system consists of a disordered network of boron dopants in silicon (Si:B) and is illustrated in Fig. 2a, b. The boron atoms were implanted in n-type silicon with a concentration of  $2 \times 10^{19} \text{ cm}^{-3}$  at the surface (Methods, Extended Data Fig. 1). A 300-nm-diameter active region was defined by eight electrodes. The central silicon region was etched (about 80 nm deep) so that the boron concentration at the recessed surface was reduced to about  $5 \times 10^{17} \text{ cm}^{-3}$ , as confirmed by secondary-ion mass spectroscopy. The current–voltage ( $I$ – $V$ ) characteristics (Fig. 2c, Extended Data Fig. 2) become increasingly nonlinear with decreasing  $T$ , and can be modelled as electric-field-activated hopping conduction at low temperatures (Supplementary Notes 1, 2). The network's potential landscape (Fig. 1b) depends in a highly nonlinear way on the input and control voltages, and spans a high-dimensional space. The output current is determined by this complex potential landscape. The nonlinear projection is realized when a combination of two or more input voltages is converted to an output current.

To identify the charge transport regimes, we focus on the low-bias conductance<sup>11</sup>  $G = dI_D/dV_{SD}|_{V_{SD}=-10 \text{ mV}}$ , where  $I_D$  is the drain current and  $V_{SD}$  is the source–drain voltage:

$$G(T) = G_b e^{(-\varepsilon_b/k_B T)} + G_h e^{-(T_h/T)^p} \quad (1)$$

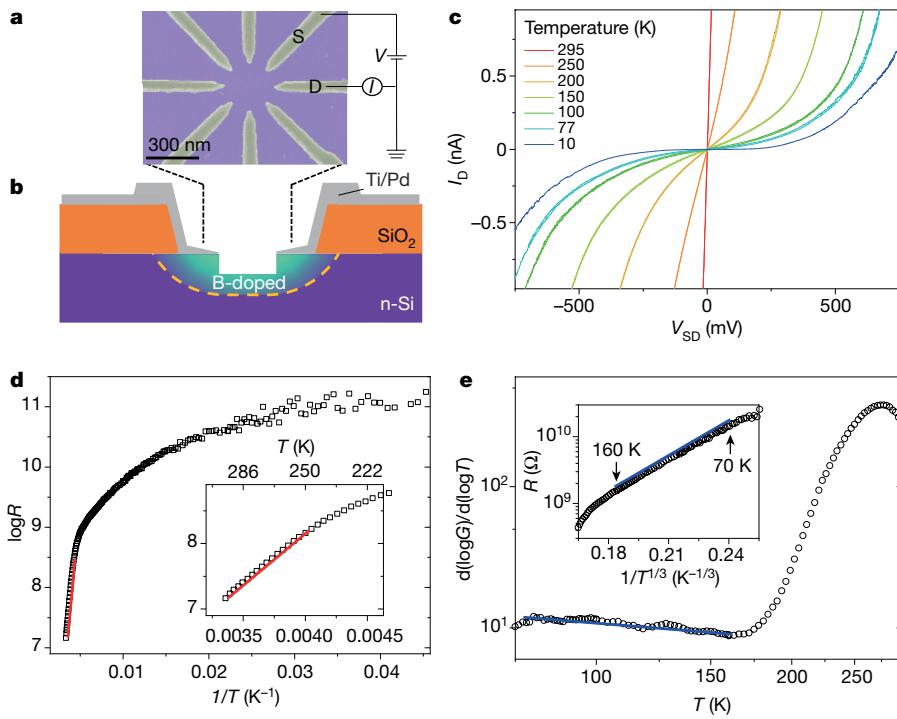
where the first term describes band (b) conduction and the second term describes hopping (h) conduction.  $G_b$  and  $G_h$  are pre-factors with a much weaker temperature ( $T$ ) dependence than the exponential terms,  $\varepsilon_b$  is the dopant ionization energy,  $T_h$  is a characteristic temperature of hopping conduction and  $k_B$  is the Boltzmann constant. The exponent  $p$  depends on the specific hopping model<sup>11</sup>. The resistance  $R = 1/G$  as a function of inverse temperature  $1/T$  is shown in Fig. 2d. At  $T > 250 \text{ K}$ , hole-band conduction dominates. The extracted  $\varepsilon_b$  is about 130 meV, three times larger than the value of boron in bulk silicon, about 45 meV. We attribute this increased ionization energy to dopant deactivation<sup>24,25</sup>: for hydrogen-like dopants near the silicon surface, the decreased dielectric screening leads to stronger electron confinement, and therefore a larger ionization energy.

We adopt the method proposed by Zabrodskii et al.<sup>15</sup> to distinguish the hopping regime and extract  $p$  (Methods). For 70–160 K, we find  $p = 0.342 \pm 0.023$ , in agreement with  $p = 1/3$  predicted for two-dimensional Mott variable-range hopping (Mott-VRH)<sup>11,26</sup> (Fig. 2e). The two-dimensional nature implies that the dopants participating in transport are located close to the silicon surface, because the hopping resistance increases exponentially with inter-dopant distance<sup>11</sup>, which is lowest near the surface. This is consistent with the dopant deactivation observed in the band-conduction regime. Above about 160 K, band conduction starts to contribute, becoming dominant above about 250 K.

To demonstrate classification in the hopping regime (Supplementary Notes 3–7), we followed the evolutionary approach of ref. <sup>12</sup> (Methods) and configured the system into Boolean logic (Fig. 3a–c, Extended Data Figs. 3–5) at 77 K. The working-temperature window for a set of control voltages (about 30 K) is approximately 15 times wider than in our previous nanoparticle system<sup>12</sup> (about 2 K). The retention period of the gates is over two months in liquid nitrogen, and the device characteristics remain virtually unchanged after thermal cycling, indicating the robustness of the dopant network. Boolean logic represents a prototypical two-input binary classification problem<sup>3</sup>, and the XOR classification problem is a poignant example of a single-layer perceptron's inability to solve problems with linearly inseparable vectors<sup>27</sup>. Hence, solving the linearly inseparable X(N)OR problem demonstrates the system's separation ability<sup>3,22,23</sup> (Fig. 1a).

As realizing all Boolean logic gates with a standard ANN requires at least one hidden layer of two neurons<sup>3</sup> (corresponding to nine linear and three nonlinear operations), our dopant network can be considered to emulate at least such a neural network in hardware (Fig. 3d). Importantly, the dopant network has only a 300-nm-diameter footprint and an average power consumption of about 1  $\mu\text{W}$  (Methods). Using established monolithically integrated readout circuits (Methods, Extended Data Fig. 6), the bandwidth of the readout circuitry can be increased from 40 Hz in our current setup to over 100 MHz. With optimization (Methods and Supplementary Note 8), the energy efficiency of the dopant network at 77 K is projected to exceed 100 tera-operations per second per watt (TOPs<sup>-1</sup>W<sup>-1</sup>, where OP is one typical linear operation of a neural network<sup>28</sup>), one order of magnitude higher than a state-of-the-art customized CMOS neural network accelerator<sup>29</sup> (Supplementary Notes 8, 9, Extended Data Fig. 7b).

To investigate the correlation between the functionality of our devices and the transport mechanism, we performed random searches with 10,000 sets of control voltages as a function of temperature. We define the total abundance  $A$ , representing the overall probability of finding Boolean logic, with two fitness  $F$  thresholds for each logic gate<sup>12</sup> (Methods). For both fitness thresholds  $F > 1, 2$ , the total abundance drops to below 5% when band conduction sets in at around



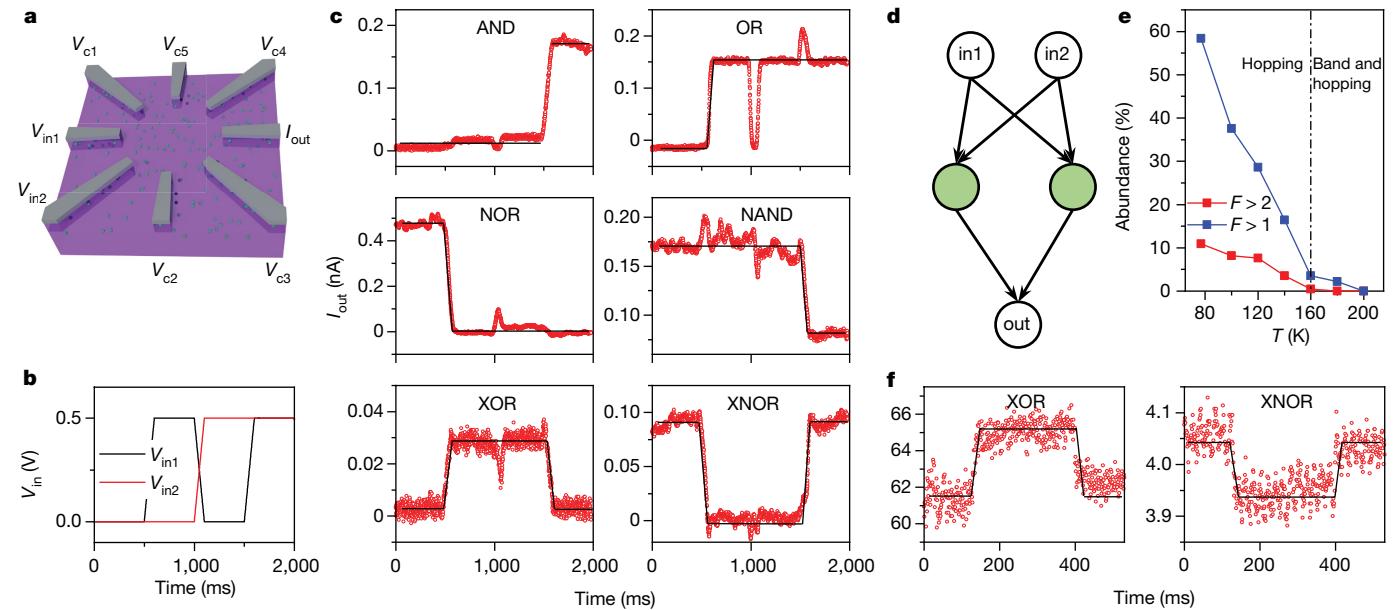
**Fig. 2 | Device structure and charge transport mechanism.** **a**, Scanning electron microscope image, indicating the source (S) and drain (D) contacts for  $I$ - $V$  measurements. **b**, Schematic cross-section, illustrating the doping profile and the p-n junction (yellow dashed line). **c**,  $I$ - $V$  characteristics at different temperatures ( $T$ ) showing nonlinear behaviour below about 250 K. **d**, Resistance  $R$  versus inverse temperature at  $V_{SD} = -10$  mV. Band transport is observed for 250–295 K (indicated by the red line in the main figure and the inset, which shows the high- $T$  region). **e**, Logarithmic derivative of the low-bias conduction  $G$  with respect to  $T$ . The linear segment for 70–160 K indicates hopping conduction (blue line). Inset, semi-logarithmic plot of  $R$  versus  $1/T^{1/3}$ , indicating two-dimensional variable-range hopping for 70–160 K (blue line) with  $T_h = 7.7 \times 10^4$  K, falling well within the range reported for Mott's VRH model<sup>16</sup>.

160 K (Fig. 3e). Hence, functionality is highly correlated to the hopping regime.

Led by this correlation, we tried to increase the operating temperature by suppressing band conduction. With increasing temperature, dopants near the p-n junction (Fig. 2b) are expected to be ionized first, as they are less subject to deactivation than dopants far away from the junction. By depleting the junction using a backgate, we indeed observe nonlinearity, and can evolve all six major logic gates at room temperature (Fig. 3f, Extended Data Fig. 8). The confirmed correlation

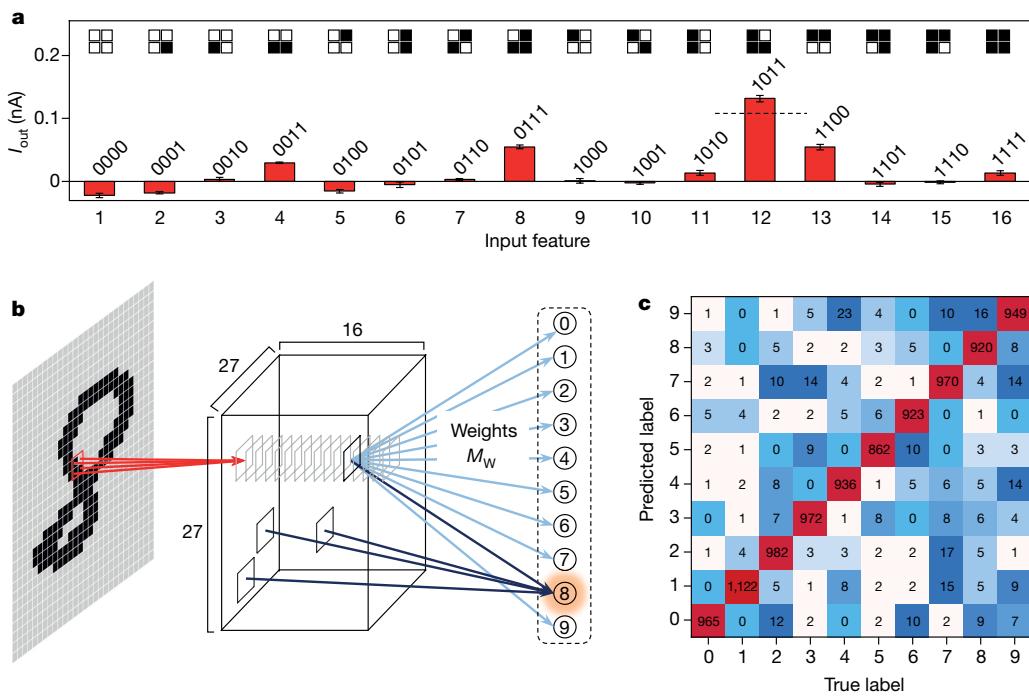
between functionality and the charge transport mechanism can serve as a guiding tool towards robust functionality at room temperature.

To demonstrate the ability of our device to perform more complicated classification tasks, we performed four-input binary classification in the form of filtering  $16 \times 2 \times 2$  black (1) and white (0) pixel features, as shown in the inset of Fig. 4a. The four pixel values are encoded as four input voltages to our dopant network, together with three control voltages and one output current. We use the three control voltages to evolve a single network into 16 different filters at 77 K. Each filter should



**Fig. 3 | Evolution of Boolean logic.** **a**, Schematic electrode configuration, indicating input voltages ( $V_{in1}, V_{in2}$ ), control voltages ( $V_{c1}-V_{c5}$ ) and output current ( $I_{out}$ ). **b**, Input waveforms. The logic 0 and 1 are represented here by two different voltages, 0 V and 0.5 V, respectively (see also Supplementary Note 6). **c**, Major Boolean logic gates at 77 K (experimental current values in red, desired output normalized to the experimental data in black). We reproduced all Boolean logic gates in seven devices. **d**, ANN with two hidden neurons (green

filled circles) emulated by the dopant network device. The ANN requires six (linear) weight multiplications, three (linear) summations and three (nonlinear) activations. **e**, Total abundance of logic gates (defined in Methods) as a function of temperature. The dashed line marks the onset of band conduction. The blue and red curves correspond to fitness thresholds of  $F > 1$  (noise level) and  $F > 2$ , respectively. **f**, XOR and XNOR gates evolved at room temperature with a backgate voltage of about 12 V.



**Fig. 4 | Feature filtering and handwritten digit classification.** **a**, Current response of one of the 16 filters. The  $2 \times 2$  pixel black/white patterns (inset) are represented by '0000', '0001', ..., '1111', with black (1) and white (0) mapped to input voltages  $0.5\text{ V}$  and  $-0.5\text{ V}$ , respectively. The output current of this filter is maximal when the '1011' pattern is presented. Error bars represent the standard deviation of ten tests. **b**, Feature mapping for digit recognition. Specific filters are activated (bold dark squares) depending on the features presented to them.

For clarity, most of the  $27 \times 27 \times 16$  filters are not shown. The output of the filters is obtained from the experimental data shown in **a** and Extended Data Fig. 9. The ten output nodes, representing digits 0 to 9, are connected to the filters through a weight matrix  $M_W$  of a linear classifier. **c**, Confusion matrix of classification with the 10,000 MNIST test dataset, showing that 96.0% of the digits are correctly classified.

make one of the 16 features distinguishable from all the others, which is realized by evolving the dopant network such that it yields the maximal or minimal output current for that specific feature (Fig. 4a, Extended Data Fig. 9). If we feed a feature to a group of 16 filters, each of which distinguishes one feature, then the 4-dimensional data are mapped to a 16-dimensional vector, and each feature vector is separated from the others in one of the dimensions (Supplementary Note 9).

Our approach allows the separation of data by evolving filters that are capable of processing data in parallel and with high throughput. Compared with optical networks, which also allow parallel processing, our dopant networks feature tunability and have much smaller dimensions: about 100 nm instead of centimetres<sup>7</sup>.

Taking advantage of the separation ability of our nanomaterial system, we used the evolved filters as the core ingredient to classify the Modified National Institute of Standards and Technology (MNIST) digits<sup>13</sup>. The whole classification procedure consists of a feature mapping layer of the evolved filters inspired by the convolutional neural network<sup>2</sup>, followed by a linear classifier in a traditional computer, which can in principle also be realized in materio<sup>30</sup> (Fig. 4b). The  $28 \times 28$  greyscale pixels of each MNIST digit are converted to black and white using a threshold and divided into  $2 \times 2$  pixel receptive fields (overlapping in one row/column with neighbours). The receptive fields feed their signal to the cluster of 16 filters, each filtering out one of the features. The  $(28 \times 28)$ -dimensional MNIST data are hence mapped onto  $(27 \times 27 \times 16)$ -dimensional feature vectors. The linear classifier then converts these high-dimensional feature vectors to a 10-dimensional output by a weight matrix  $M_W$  (Fig. 4b), obtained by pseudo-inverse learning<sup>31</sup> with the 60,000 MNIST training data (Methods). The largest of the ten outputs finally determines the recognized digit.

Application to 10,000 test digits shows 96.0% accuracy (Fig. 4c, Supplementary Note 9, Extended Data Fig. 10), which is better than the accuracy obtained with state-of-the-art physical reservoir computing<sup>8</sup> and optical networks<sup>7</sup>. We note that differences in the output current scales of the different filters are irrelevant, because the weight matrix will automatically compensate for those (Supplementary Note 9). We also simulated feature filters with ideal characteristics, which are only activated when presented with its corresponding feature (output 1 for target feature and 0 otherwise). The classification of the MNIST dataset with these ideal filters results in an accuracy of 96.2%. Therefore, as long as the data mapped to the feature space are sufficiently separated, a linear classifier can learn the decision boundaries. The underlying reason is that every complete set of independent vectors, be it orthogonal (ideal) or not, can represent other vectors by linear combination. This shows the power of our dopant network in making data linearly separable, owing to its intrinsic nonlinear transformation. The ability to separate data, when combined with an adaptable linear readout in a scaled-up system, can achieve universal computational power<sup>8,22,23</sup>. For instance, in ANNs, perceptrons can be cascaded to solve more complex problems<sup>3</sup>. This analogy strongly suggests that a system of interconnected dopant networks can address a much wider range of tasks, particularly because the computational power of a single dopant network is larger than that of a single perceptron (it can solve XNOR whereas a single perceptron cannot).

At the system level, we anticipate a number of necessary developments. First, the total evolution time of the filters, which scales linearly with their number, can be reduced (by a factor  $10^6$ ; Methods and Supplementary Notes 7, 8). Besides competitive evolutionary approaches<sup>32</sup>, we will also explore gradient-based methods<sup>33</sup>. Second, it will be highly advantageous to store the evolved control voltages locally, employing,

for example, memristors<sup>30</sup> (Supplementary Note 8). Third, memristive technology is also suitable for in materio implementation of the linear classification step in our scheme with energy efficiency comparable to our material-based nonlinear feature filters. Fourthly, processing analogue instead of binary signals would be more natural for our devices. To filter more complex, non-binary features, such as edge detection by the brain<sup>1</sup>, more electrodes per device are needed and/or multiple devices need to be interconnected, so that more input signals can be processed in parallel. This will also allow for more control voltages per filter (at present, three) to improve the signal-to-noise ratio. Lastly, for practical applications, room-temperature operation with long retention, low-voltage supplies and without a backgate is desired, which we deem possible by engineering the deactivation effect in a silicon-on-insulator-based system.

Our silicon-based system provides a powerful platform for carrying out machine learning tasks in hardware. By material learning, we harness the intrinsic nonlinearity and tunability of a nanomaterial system to efficiently realize functional tasks without the need to design circuitry for the underlying elementary operations. The small footprint and silicon-compatible fabrication process facilitates scaling up for massively parallel, high-throughput information processing platforms for complex computational tasks. Whereas the randomness and discreteness of dopants pose challenges on conventional silicon electronics, we have presented a computational paradigm that takes full advantage of these properties. When integrated with other technologies, complex classification problems can be solved fully in materio, potentially achieving ultrahigh computational density and energy efficiency<sup>14</sup>.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-019-1901-0>.

- Hubel, D. H. & Wiesel, T. N. Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.* **148**, 574–591 (1959).
- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- Haykin, S. *Neural Networks and Learning Machines* (Pearson Prentice Hall, 2008).
- Cover, T. M. Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Trans. Electron. Comput.* **EC-14**, 326–334 (1965).
- Torrejon, J. et al. Neuromorphic computing with nanoscale spintronic oscillators. *Nature* **547**, 428–431 (2017).
- Tanaka, G. et al. Recent advances in physical reservoir computing: a review. *Neural Netw.* **115**, 100–123 (2019).

- Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- Du, C. et al. Reservoir computing using dynamic memristors for temporal information processing. *Nat. Commun.* **8**, 2204 (2017).
- Hung, C. S. & Gliessman, J. R. Resistivity and Hall effect of germanium at low temperatures. *Phys. Rev.* **96**, 1226–1236 (1954).
- Mott, N. F. Conduction in glasses containing transition metal ions. *J. Non Cryst. Solids* **1**, 1–17 (1968).
- Gantmakher, V. F. *Electrons and Disorder in Solids* (Clarendon Press, 2005).
- Bose, S. K. et al. Evolution of a designless nanoparticle network into reconfigurable Boolean logic. *Nat. Nanotechnol.* **10**, 1048–1052 (2015).
- Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
- Xu, X. et al. Scaling for edge inference of deep neural networks. *Nat. Electron.* **1**, 216–222 (2018).
- Zabrodskii, A. G. & Zinov'eva, K. N. Low-temperature conductivity and metal-insulator transition in compensated n-Ge. *Sov. Phys. JETP* **59**, 425–433 (1984).
- Jenderka, M. et al. Mott variable-range hopping and weak antilocalization effect in heteroepitaxial Na<sub>x</sub>IrO<sub>3</sub> thin films. *Phys. Rev. B* **88**, 045111 (2013).
- Miller, J. F. & Downing, K. Evolution in materio: looking beyond the silicon box. In *Proc. 2002 NASA/DoD Conference on Evolvable Hardware* 167–176 (IEEE, 2002).
- Harding, S. & Miller, J. F. Evolution in materio: a tone discriminator in liquid crystal. In *Proc. 2004 Congress on Evolutionary Computation* 1800–1807 (IEEE, 2004).
- Mohid, M. & Miller, J. F. Evolving robot controllers using carbon nanotubes. In *Proc. 13th European Conference on Artificial Life* 106–113 (MIT Press, 2015).
- Wolfram, S. Approaches to complexity engineering. *Physica D* **22**, 385–399 (1986).
- Backus, J. Can programming be liberated from the von Neumann style? A functional style and its algebra of programs. *Commun. ACM* **21**, 613–641 (1978).
- Maass, W., Natschläger, T. & Markram, H. Real-time computing without stable states: a new framework for neural computation based on perturbations. *Neural Comput.* **14**, 2531–2560 (2002).
- Dale, M., Stepney, S., Miller, J. F. & Trefzer, M. Reservoir computing in materio: an evaluation of configuration through evolution. In *2016 IEEE Symposium Series on Computational Intelligence, SSC/1–8* (IEEE, 2016).
- Björk, M. T., Schmid, H., Knoch, J., Riel, H. & Riess, W. Donor deactivation in silicon nanostructures. *Nat. Nanotechnol.* **4**, 103–107 (2009).
- Pierre, M. et al. Single-donor ionization energies in a nanoscale CMOS channel. *Nat. Nanotechnol.* **5**, 133–137 (2010).
- Hartstein, A. & Fowler, A. B. High temperature ‘variable range hopping’ conductivity in silicon inversion layers. *J. Phys. C* **8**, L249–L253 (1975).
- Minsky, M. & Papert, S. *Perceptrons: An Introduction to Computational Geometry* (MIT Press, 1969).
- Chen, T. et al. DianNao: a small-footprint high-throughput accelerator for ubiquitous machine-learning. *ACM SIGPLAN Not.* **49**, 269–284 (2014).
- Lee, J. et al. UNPU: an energy-efficient deep neural network accelerator with fully variable weight bit precision. *IEEE J. Solid-State Circuits* **54**, 173–185 (2019).
- Li, C. et al. Analogue signal and image processing with large memristor crossbars. *Nat. Electron.* **1**, 52–59 (2018).
- Tapson, J. & van Schaik, A. Learning the pseudoinverse solution to network weights. *Neural Netw.* **45**, 94–100 (2013).
- Such, F. P. et al. Deep neuroevolution: genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning. Preprint at <http://arxiv.org/abs/1712.06567> (2017).
- Kingma, D. P. & Ba, L. J. Adam: a method for stochastic optimization. Preprint at <https://arxiv.org/abs/1412.6980> (2015).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

# Article

## Methods

### Samples

300 nm of thermal oxide was grown on an n-type silicon substrate (Extended Data Fig. 1a), in which  $26 \times 60 \mu\text{m}^2$  implantation windows were defined by photolithography and wet etching. Another 35 nm of oxide was thermally grown in the implantation window to serve as a stopping layer (Extended Data Fig. 1b). After boron implantation (9 keV equivalent,  $3.5 \times 10^{14} \text{ cm}^{-2}$ ), and activation via rapid thermal annealing ( $1,050^\circ\text{C}$ , 7 s; Extended Data Fig. 1c), the 35-nm stopping layer was removed by wet etching. The boron concentration near the silicon surface exceeds  $2 \times 10^{19} \text{ cm}^{-3}$  to ensure Ohmic contact with the electrodes, and decreases monotonically with depth (Extended Data Fig. 1h). After lift-off of the wire-bonding pads (1.5 nm Ti/40 nm Pd) defined by photolithography (Extended Data Fig. 1d), eight 1.5 nm Ti/40 nm Pd nanoelectrodes were patterned on top of the silicon by electron-beam lithography (Extended Data Fig. 1e). The devices were annealed at  $160^\circ\text{C}$  for 10 min to promote the metal/silicon contact quality. The silicon surface was further etched by reactive ion etching to reduce the boron concentration in the active gap area (Extended Data Fig. 1f, g; see also Supplementary Note 6). The surface was finally treated with mild oxygen plasma, followed by 1% HF etching to remove possible contaminants.

### Charge transport

Following Zabrodskii et al.<sup>15</sup>, we introduce the logarithmic derivative  $w = d(\log G)/d(\log T)$ . From equation (1), we see that if the hopping term  $G_h e^{-(T_h/T)^p}$  dominates,  $\log w \approx \log p + p(\log T_h - \log T)$ , and  $p$  can be derived from the slope of the  $\log w - \log T$  curve (Fig. 2e), thus allowing us to identify the exact hopping conduction model. For  $T < 70 \text{ K}$ , the measurement noise level prevents unambiguous identification of the charge transport mechanism (Fig. 2d), but probably VRH continues<sup>34</sup>. The charge transport behaviour described in the main text has been observed in the two devices we characterized.

### Measurements

We conducted the charge transport measurements and evolution of logic gates at different temperatures in a customized flow cryostat. The cryostat is equipped with 12 coaxial cables to reduce capacitive cross-talk. We use a battery-powered electronics rack (IVVI rack and matrix rack; <http://qwork.tudelft.nl>) composed of digital-to-analogue converters (DAs) and I/V converters for low-noise measurements (Extended Data Fig. 6). The output range of the DAs is from  $-2 \text{ V}$  to  $2 \text{ V}$ . The I/V converter has four amplification settings,  $1 \text{ G}\Omega$ ,  $100 \text{ M}\Omega$ ,  $10 \text{ M}\Omega$  and  $1 \text{ M}\Omega$ , each corresponding to a different measurement range. For measurements at cryogenic temperatures,  $1 \text{ G}\Omega$  amplification is chosen as default, by which currents from  $-3.4 \text{ nA}$  to  $3.4 \text{ nA}$  can be measured. The output of the I/V converter is sampled by a multimeter (Keithley 2000) or digitizer (ADwin-Gold II). For room-temperature evolution, the I/V converter amplification is set to  $10 \text{ M}\Omega$ , resulting in a current measurement range from  $-340 \text{ nA}$  to  $340 \text{ nA}$ . The measurements are automated with either LabVIEW or Python scripts. For fixed-temperature measurements, the devices were inserted into a liquid-helium ( $4.2 \text{ K}$ ) or liquid-nitrogen ( $77 \text{ K}$ ) dewar with a customized dipstick.

### Readout speed

In our system, the relaxation time of hopping conduction is less than  $10 \text{ ns}$  at  $77 \text{ K}$  and even smaller at higher temperatures (Supplementary Note 8), so it is not the dominant timescale in our present devices. Like in all measurements on resistive devices<sup>35</sup>, the readout speed of our dopant networks is constrained by a large capacitive load (Extended Data Fig. 6b). The long, twisted pairs (about  $3 \text{ m}$ ) as well as the filters of the matrix rack amount to a large load capacitance  $C_L$  (about  $4 \text{ nF}$ ) that limits the signal speed. With the existing setup, we have a bandwidth

(cutoff frequency of the resistor–capacitor (RC) circuit in Extended Data Fig. 6b)

$$\text{BW} = \frac{1}{2\pi C_L(R_{\text{out}}||R_{\text{IV}})} \approx 40 \text{ Hz}$$

where  $R_{\text{IV}} = 1 \text{ M}\Omega$  is the input resistance of the I/V converter at  $1 \text{ G}\Omega$  amplification, and the dopant network output resistance  $R_{\text{out}}$  is typically hundreds of  $\text{M}\Omega$  (Extended Data Fig. 7d).

By monolithically integrating a transistor-based readout circuitry close to the dopant network<sup>35</sup> (Extended Data Fig. 6c), we can reduce the capacitive load for fast readout, and also enable interconnection with other devices. With existing CMOS technology, the load capacitance can be easily reduced to below  $1 \text{ fF}$ , and the RC-related bandwidth can reach  $160 \text{ MHz}$ , or even more, by reducing  $R_{\text{IV}}$ .

Given a signal intensity (the difference between high and low output current levels; see ‘Fitness functions’ below), the signal-to-noise ratio (SNR) is predominantly set by the Johnson–Nyquist noise from  $R_{\text{IV}}$ , because its noise power is proportional to the bandwidth. Therefore, for a required SNR (computation precision), the bandwidth and the subsequent energy efficiency, are determined by the signal intensity (Supplementary Note 8). The signal intensity of our devices ranges from the order of  $0.1 \text{ nA}$  to the order of  $1 \text{ nA}$  (Supplementary Note 6), thus allowing over  $100 \text{ MHz}$  bandwidth (Extended Data Fig. 7a).

### Fitness functions

For Boolean logic gate evolution, the input sequences, representing the four input entries of truth tables (Fig. 3b), were fed to the input electrodes (Fig. 3a) after the control voltages were set. We monitored the output current waveform  $\bar{Y}$  and fitted it with  $\bar{Y} = m\bar{X} + C$ , where  $\bar{X}$  is the expected output waveform of a logic gate (logic high and low taking numerical values of 1 and 0, respectively).  $m$  is the proportionality factor and its value thus equals the separation of the high and low levels (signal intensity).  $C$  represents the offset. For each set of control

voltages, a fitness is evaluated by  $F = m / (\sqrt{r_{ss}} + kC)$ , with  $r_{ss}$  being the fitting residual<sup>12</sup> and  $k$  an empirical constant. A larger  $k$  puts more emphasis on minimizing the offset  $C$  in the evolution process. For the evolution of logic gates, we found that there is minimal offset for  $k=0.2$  (Fig. 3c). In the random search of logic gates at different temperatures,  $k$  has been set to 0.01 to give the waveform shape more weight than the offset. Then, a fitness value of  $F=1$  implies that the signal intensity (related to  $m$ ) almost equals the noise intensity (related to  $\sqrt{r_{ss}}$ ), and a fitness value of  $F=2$  corresponds to more robust logic gates. Based on the fitness, we define the abundance of each gate. For the 10,000 output waveforms from a random search, we assessed the fitness of each output waveform for six major logic gates. In this way, each logic gate is associated with 10,000 fitness values. The abundance of a gate  $A_i$  (where  $i$  is AND, OR, NAND, NOR, XOR, XNOR) is defined as the number of fitness values larger than a threshold, divided by 10,000. The total abundance is then defined as  $A = 1/\sum_i (1/A_i)$ . The fitness function for the feature filter evolution was defined as  $F = |I_{\text{out},i}| / [\text{avg}(|I_{\text{out},j=i}|) + \text{std}(|I_{\text{out},j=i}|)]$ , where  $I_{\text{out},i}$  is the output current corresponding to feature  $f_i$ , and avg and std stand for the average and standard deviation, respectively, of all the other feature outputs  $I_{\text{out},j \neq i}$ . Here,  $i$  runs from 1 to 16.

### Genetic algorithm

The genetic algorithm mimics natural evolution. An initial generation of 20 genomes, with the length of each genome equal to the number of control electrodes, is first randomly generated and mapped to control voltages. The fitnesses of the 20 genomes are evaluated and ranked. Then the off-spring generation of 20 genomes is produced in the following way: (1) inheriting the five elite genomes (with highest fitnesses) from the previous generation; (2) cross-breeding of the elite genomes to produce five off-spring genomes; (3) mutation of the five elite genomes

by a probability of 0.1, then cross-breeding with the five elite genomes to generate five other genomes; (4) cross-breeding of the five elite genomes with five random genomes to generate five other genomes. The genetic algorithm keeps iterating until it reaches a satisfactory fitness value (Extended Data Fig. 4a; see also Supplementary Note 7). A more detailed description of the evolution procedure is given in our previous work<sup>12</sup>.

### Power consumption

To estimate the power consumption and energy efficiency of our device, we measured the static power consumption of the six major Boolean logic gates for four different input voltage combinations, so in total 24 configurations. To measure the current of the  $i$ th ( $i$  running from 1 to 8) electrode, the voltage  $V_i$  (current  $I_i$ ) is set (measured) by a source meter (Keithley 2401), while the voltages on the other electrodes are set by either the DACs (control voltages and input voltages) or an I/V converter (output electrode). For each of the 24 configurations, the

total power  $P$  is calculated as  $P = \sum_{i=1}^8 V_i I_i$ . The average power of the 24 configurations is found to be about 1  $\mu\text{W}$ . Under operational conditions, the voltage changes on the electrodes are accompanied by charging and discharging of wire capacitances. As mentioned above ('Readout speed' section), the capacitances can be reduced to below 1 fF, making the dynamical power consumption negligible compared with the static power consumption. The static power consumption could be substantially reduced by using electrostatic electrodes (see also Supplementary Note 8).

### Weight matrix training and test

In the digit classification task, each  $28 \times 28$  pixel digit is divided into  $27 \times 27$  receptive fields of  $2 \times 2$  pixels, overlapping by one row/column of pixels. The pixels of each receptive field are mapped to the 4 inputs of 16 filters (with their experimentally determined response), each of which filters 1 of the 16 distinctive  $2 \times 2$  pixel features shown in the inset of Fig. 4a. For the  $d$ th digit in the  $N_d = 60,000$  MNIST training database, we stack the  $N_f = 27 \times 27 \times 16 = 11,664$  outputs of the filters in a feature vector  $\mathbf{O}_d = (O_{d,1}, \dots, O_{d,N_f})$ . Combining the vectors  $\mathbf{O}_d$  of 60,000 training digits together, we obtain an  $N_d \times N_f$  output matrix  $\mathbf{O} = (\mathbf{O}_1, \dots, \mathbf{O}_{N_d})^\top$ .

The true label of each digit is represented by a ten-dimensional label vector  $\mathbf{L}_d$ , whose elements are all zeros except for the  $(l+1)$ th entry being 1, where  $l \in \{0, \dots, 9\}$  is the true label of the  $d$ th MNIST digit. Ideally, the weight matrix  $M_W$  converts the feature vector of a digit to

its corresponding label vector  $\mathbf{O}_d M_W = \mathbf{L}_d$ . So, in matrix form,  $\mathbf{O} M_W = \mathbf{L}$ , where  $\mathbf{L} = (\mathbf{L}_1, \dots, \mathbf{L}_{N_d})^\top$ . The weight matrix  $M_W$  has a dimension of  $N_f \times 10$ , and is simply obtained by  $M_W = O^+ \mathbf{L}$ , where  $O^+$  is the pseudoinverse of matrix  $O$ . Once the weight matrix is trained, we test it with the  $N_t = 10,000$  MNIST test data. The feature vector of each test digit  $\mathbf{O}_t$ , ( $t = 1, \dots, N_t$ ), is multiplied by the weight matrix to acquire the predicted label vector  $\mathbf{P}_t$ ,  $\mathbf{O}_t M_W = \mathbf{P}_t$ .

The index of the maximal element of  $\mathbf{P}_t$  minus one gives the predicted label. The accuracy is calculated as the ratio of the total counts of the correctly classified digits, that is, the sum of diagonal entries in Fig. 4c, to the total number of test digits  $N_t$ .

### Data availability

Data are available from the corresponding author upon reasonable request.

34. Aharonov, A., Zhang, Y. & Sarachik, M. P. Universal crossover in variable range hopping with Coulomb interactions. *Phys. Rev. Lett.* **68**, 3900–3903 (1992).
35. Pettersson, J. et al. Extending the high-frequency limit of a single-electron transistor by on-chip impedance transformation. *Phys. Rev. B* **53**, R13272–R13274 (1996).
36. The Green500 TOP500.org <https://www.top500.org/green500/> (2019).
37. Hu, M. et al. Memristor-based analog computation and neural network classification with a dot product engine. *Adv. Mater.* **30**, 1705914 (2018).

**Acknowledgements** We thank T. Bolhuis, M. H. Siekman and J. G. M. Sanderink for technical support. We thank C. P. Lawrence, B. J. Geurts, M. Nass, A. J. Annema, M. Dale and J. Dewhurst for discussions. We thank W. M. Elferink, R. Horri, J. Wildeboer and T. Dukker for help with measurements. We acknowledge financial support from the MESA+ Institute for Nanotechnology, and the Netherlands Organisation for Scientific Research (NWO): NWA Startimpuls grant number 680-91-114 and Natuurkunde Projectruimte grant number 400-17-607.

**Author contributions** T.C. and W.G.v.d.W. designed the experiments. J.v.G., T.C., B.v.d.V. and S.V.A. fabricated the samples. T.C., J.v.G. and B.v.d.V. performed the measurements and simulations. T.C. analysed the data with input from all authors. H.B., H.C.R.E. and P.A.B. provided theoretical inputs. B.d.W. and H.C.R.E. contributed to measurement script. T.C. and W.G.v.d.W. wrote the manuscript and all the authors contributed to revisions. W.G.v.d.W. and F.A.Z. conceived the project. W.G.v.d.W. supervised the project. F.A.Z. co-supervised the sample fabrication.

**Competing interests** The authors declare no competing interests.

#### Additional information

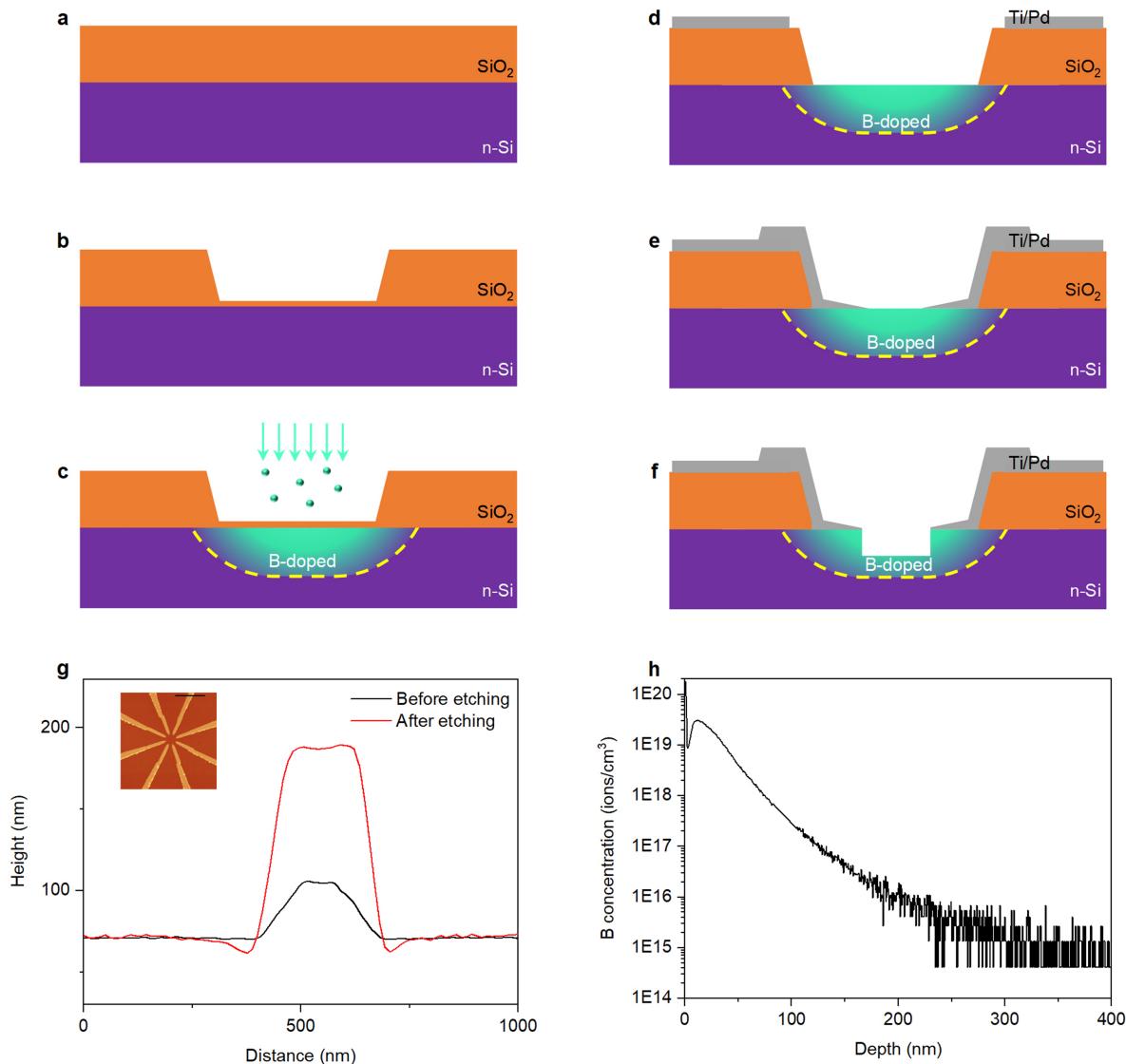
**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41586-019-1901-0>.

**Correspondence and requests for materials** should be addressed to W.G.v.W.

**Peer review information** *Nature* thanks Cyrus Hirjibehedin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

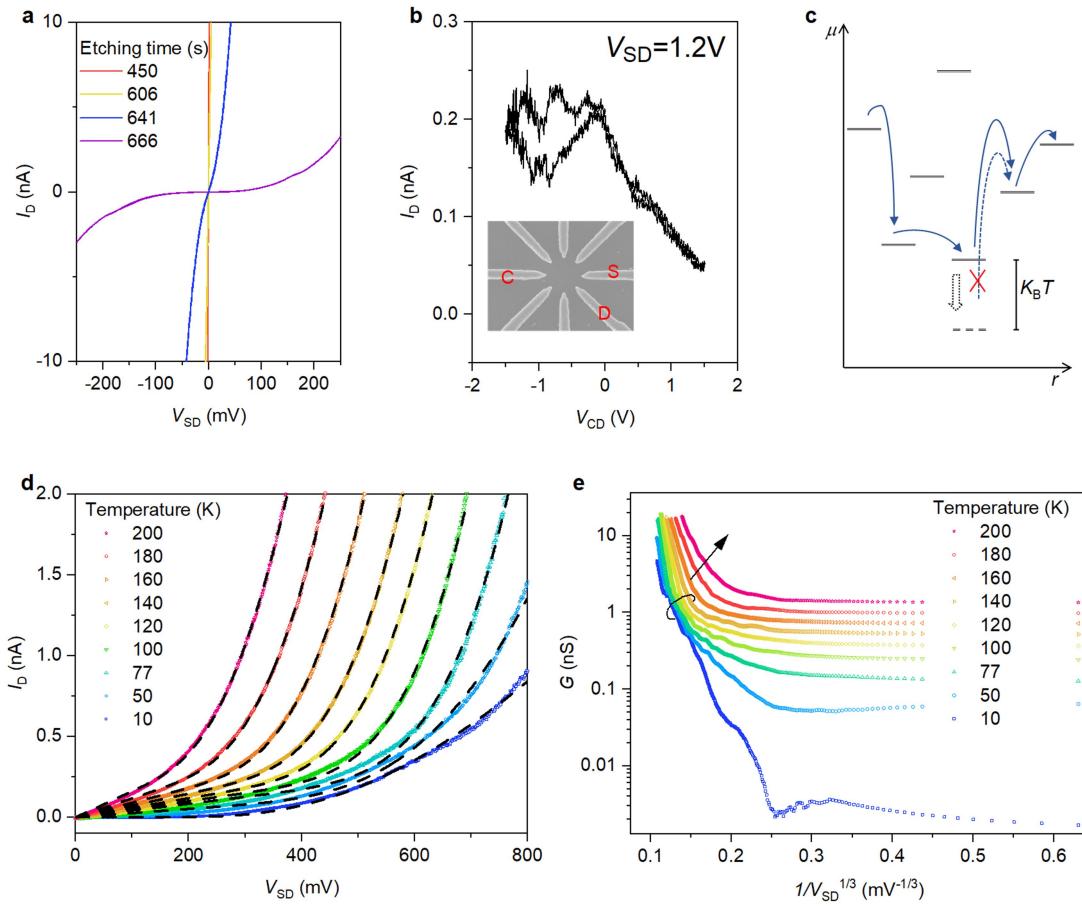
**Reprints and permissions information** is available at <http://www.nature.com/reprints>.

# Article



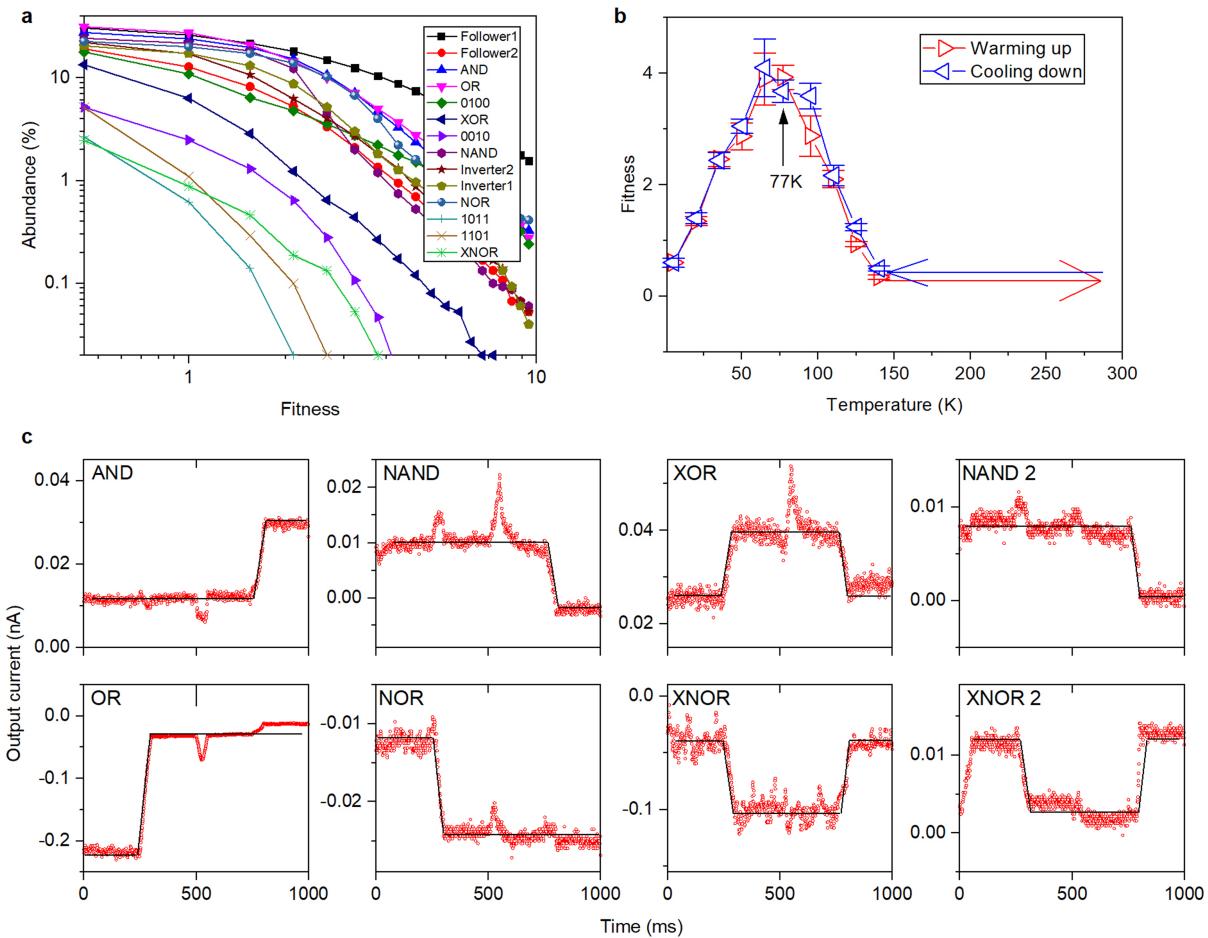
**Extended Data Fig. 1 | Fabrication steps and dopant concentration.**  
**a**, Thermal oxidation. **b**, Implantation window definition and growth of 35 nm oxide. **c**, Ion implantation. **d**, Photolithography and contact pads lift-off. **e**, Electron-beam lithography and nanoelectrodes lift-off. **f**, Reactive ion etching (RIE) of silicon. **g**, Height profile of the metal electrodes with respect to silicon before (black) and after (red) RIE etching. The etch depth of silicon is estimated by measuring the height change of the metal electrodes with respect

to the silicon surface (indicated by the black line on the atomic force microscopy image in the inset, not to scale). Assuming that the metal is not etched by RIE, the etch depth of silicon is around 83 nm. **h**, Secondary ion mass spectroscopy of the boron dopant depth profile after implantation. On the basis of the etch depth, the boron concentration near the recessed silicon surface is of the order of  $5 \times 10^{17} \text{ cm}^{-3}$ .



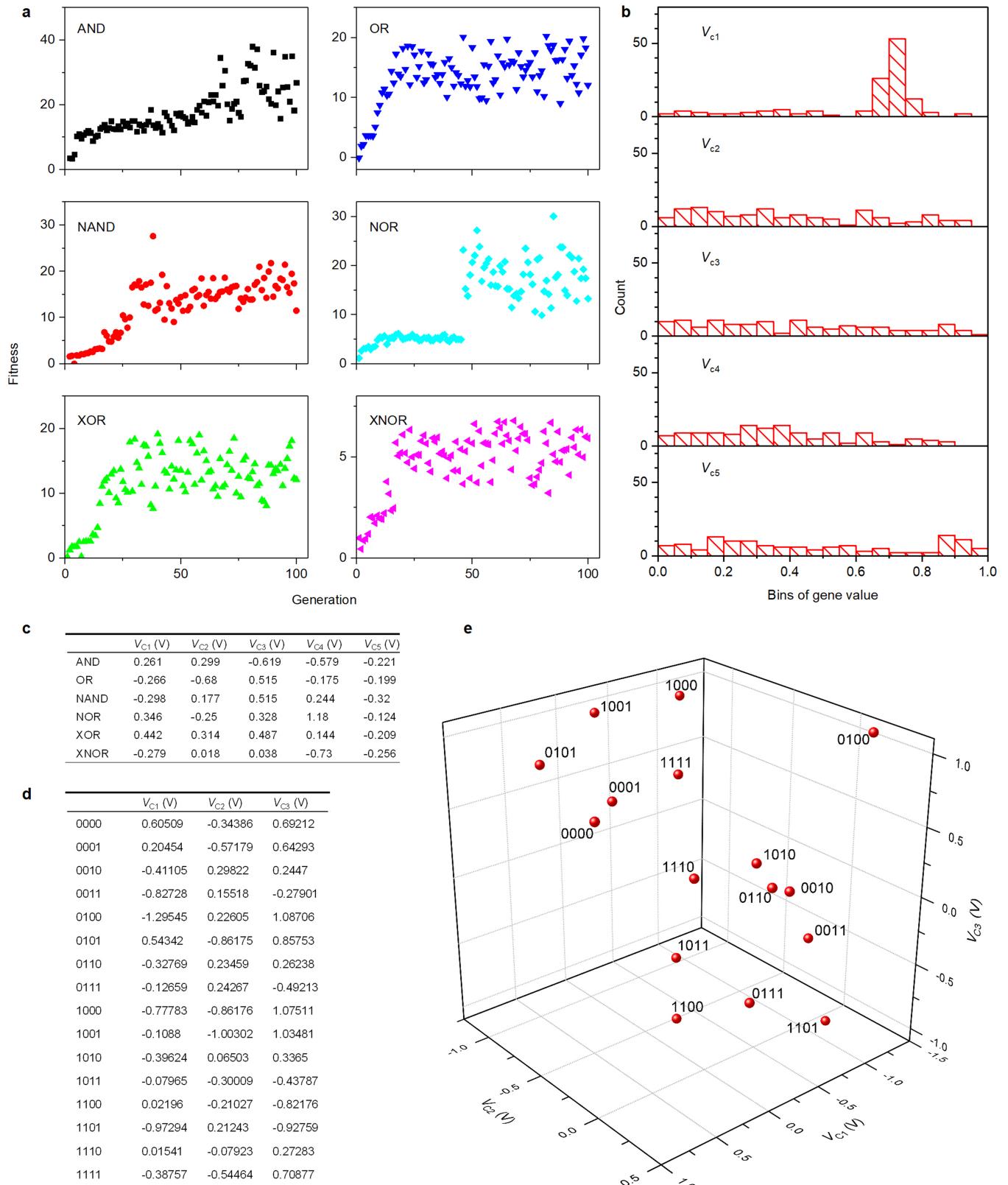
**Extended Data Fig. 2 | Nonlinear and tunable hopping conduction.** **a**,  $I_V$  characteristics at 4.2 K with different total etching time by RIE. As the total etching time increases, the nonlinearity becomes increasingly prominent, signalling the dominance of hopping conduction. **b**, Drain current versus control voltage for constant source–drain voltage  $V_{SD} = 1.2$  V at 4.2 K. The source (S), drain (D) and control (C) electrodes are shown in the inset. The hysteresis for negative gate voltage is probably due to charging of the other five floating electrodes. **c**, Schematic plot of electrochemical potential  $\mu$

versus position  $r$ , illustrating the tunability. The solid lines represent impurity states and the arrows represent hopping of carriers among states. See Supplementary Note 3 for detailed discussion. **d**, Fitting the temperature-dependent  $I_V$  curves with the model described by equation (2) in Supplementary Note 2. Black dashed lines represent the fitted curves. **e**, Conductance versus the reciprocal of the cube root of the source–drain voltage at different temperatures. The black circle groups data at temperatures below 140 K. See also Supplementary Note 2 for more discussion.



**Extended Data Fig. 3 | Evolved logic gates at 77 K.** **a**, Abundance plot of 14 non-trivial truth tables at 77 K. From a search with 10,000 sets of randomly generated control voltages, we found all 16 possible truth tables that can be realized for a two-input–one-output configuration. **b**, Thermal stability of a NAND gate evolved at 77 K. Above 140 K, the output current clipped to compliance, and therefore the fitness was not extracted. The error bars represent the standard deviation of ten tests (see also Supplementary Note 4).

**c**, Boolean logic gates evolved at 77 K in a device other than the one in Fig. 3c. Red circles are experimental output currents, and black lines represent the normalized desired output currents. The left six panels show the six major logic gates evolved with input voltage levels 0 V and 0.5 V. The right two panels show a NAND and a XNOR gate evolved with input voltage levels of −0.25 V and 0.25 V, showing the adaptability of the dopant network to different voltage levels (see also Supplementary Note 6).

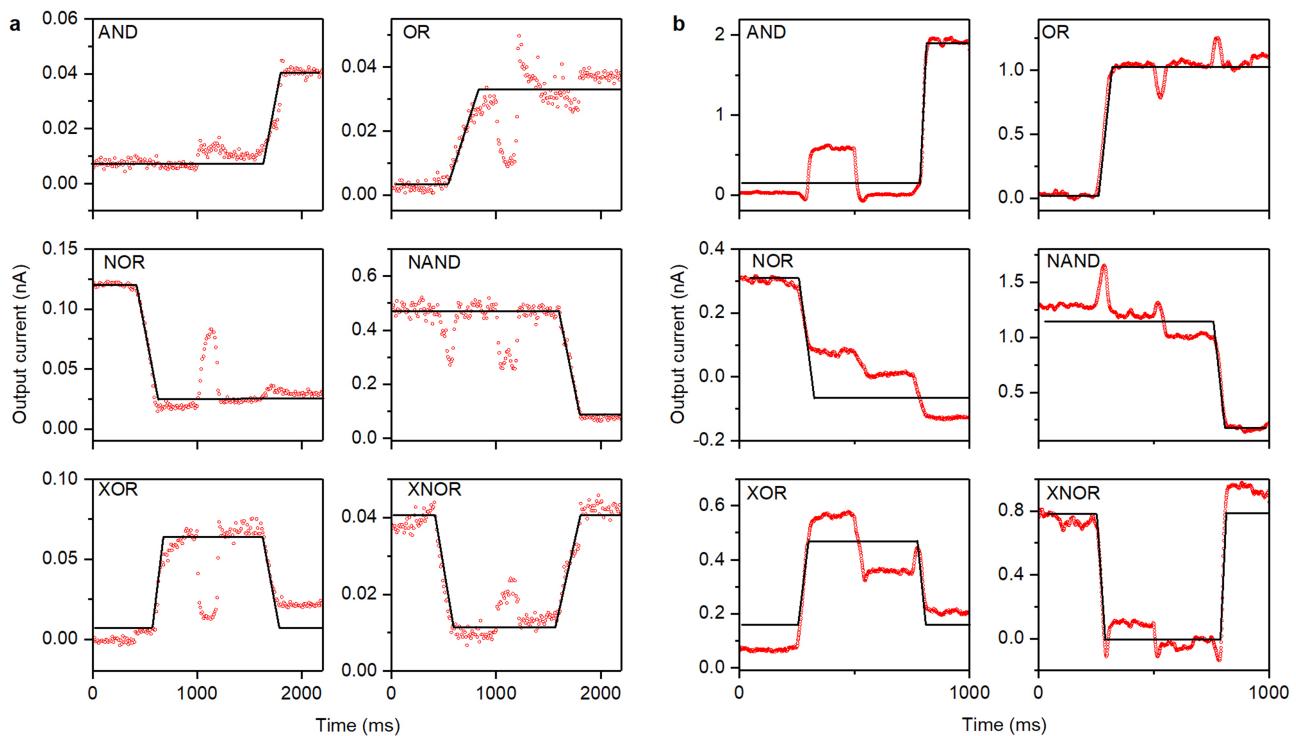


**Extended Data Fig. 4 | Convergence of genetic algorithm in the configuration space.** **a**, Genetic algorithm convergence for the six major Boolean logic gates at 77 K. The best fitness of the 20 genomes is plotted against generation. **b**, Histograms of the control voltages that configure the dopant network to the XNOR gates with fitness  $F$  larger than 1.5. The first control voltage is prominently concentrated in a small range, but the others do

not show a favourite range. The ranges of the five control voltages are  $(-600, 600)$ ,  $(-1,200, 1,200)$ ,  $(-1,200, 1,200)$ ,  $(-1,200, 1,200)$  and  $(-600, 600)$ .

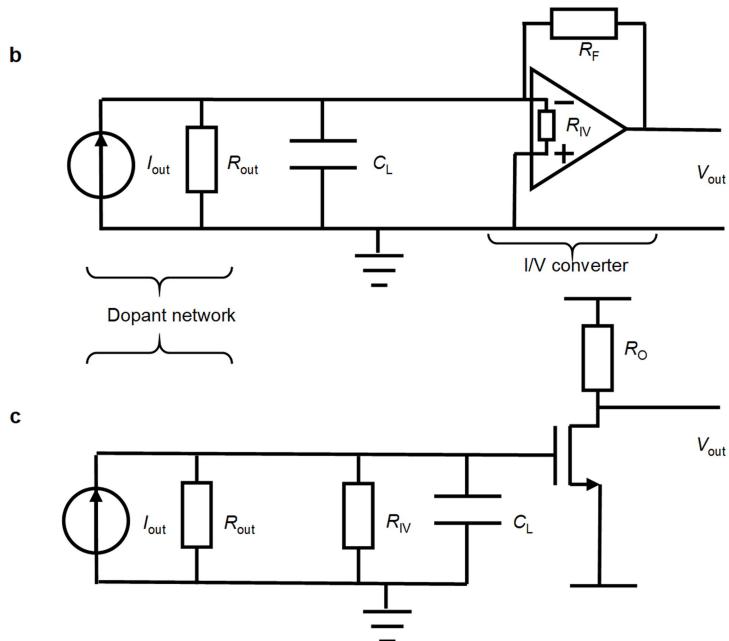
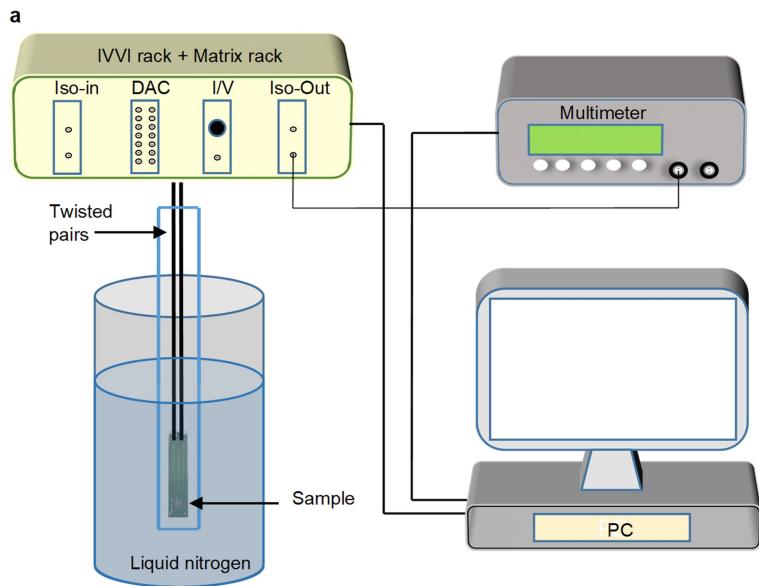
**c**, Control voltages for the six major logic gates. **d**, Control voltages for the 16 filters, which are visualized in **e**. The filters '0110' and '0010' have the smallest separation. See Supplementary Notes 3 and 7 for more discussion.

# Article



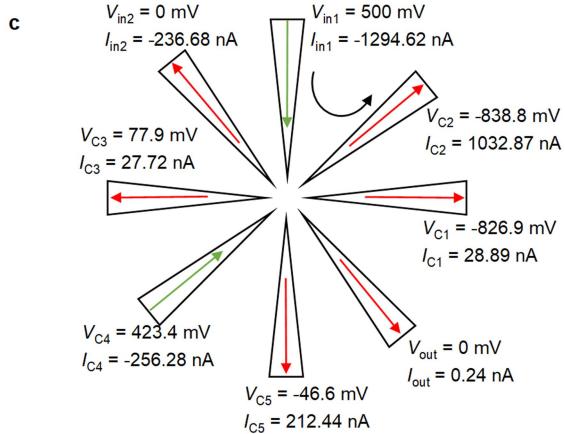
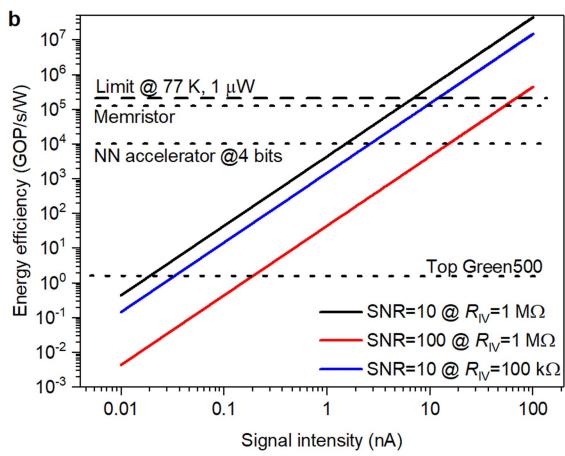
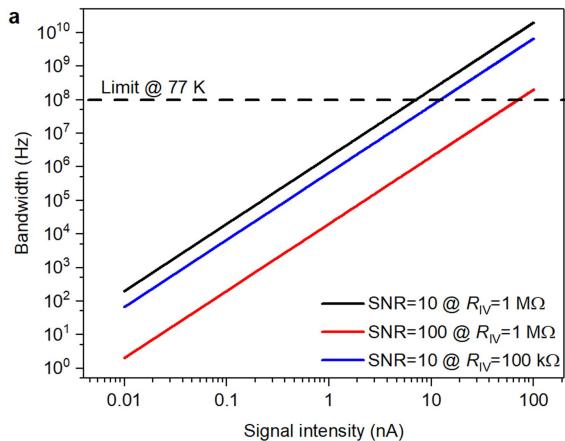
**Extended Data Fig. 5 | Evolution of logic gates at two ends of hopping conduction.** **a**, Evolved logic gates at 4.2 K, at which the charge transport mechanism is still VRH (Methods). **b**, Evolved logic gates at 140 K. Red circles

are experimental output currents and black lines represent the normalized desired output currents. See Supplementary Note 5 for a detailed discussion.



**Extended Data Fig. 6 | Measurement setup.** **a**, Schematics of the existing measurement setup. **b**, Equivalent circuit of the current measurement setup.  $I_{\text{out}}$  and  $R_{\text{out}}$  represent the output current and output resistance of the device.  $C_L$  is the parasitic capacitance of about 4 nF.  $R_{\text{IV}}$  and  $R_F$  are the input resistance and

feedback resistance of the I/V converter, respectively. **c**, Schematic of an integrated high-speed current reading circuit. Here,  $R_{\text{IV}}$  is a resistor to convert current to voltage.  $C_L$  is the parasitic capacitance that can be reduced to below 1fF.  $R_o$  is a resistor that sets the amplification.



**Extended Data Fig. 7 | Bandwidth and energy efficiency scaling.** **a**, The scaling of allowed bandwidth with signal intensity in a log–log plot. The back, blue and red solid lines represent three different indicated cases. Larger required SNR (red) and smaller  $R_{IV}$  (blue) lower the bandwidth. The horizontal black dashed line represents the limit set by the hopping relaxation time at 77 K, which increases with temperature. **b**, The scaling of equivalent energy efficiency with signal intensity in a log–log plot. Larger SNR (red) and smaller  $R_{IV}$  (blue) lowers the energy efficiency. The horizontal black dashed line represents the limit at 77 K and fixed power consumption. If the dopant network power consumption is lowered, then the limit and all three scaling trends shift upwards. The three black dotted lines mark three representative

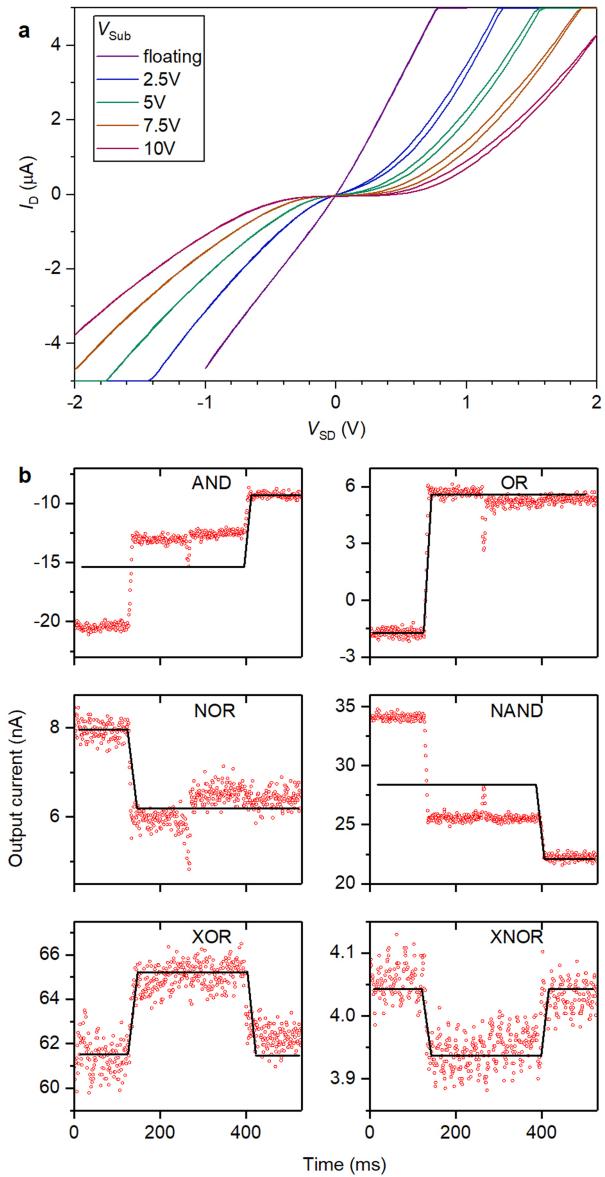
NAND00	V (mV)	I (nA)	$I_{std}$ (nA)	$R_{diff}$ (MΩ)
Input 1	0.00	-81.81	0.84	1.22
Input 2	0.00	14.74	0.08	8.49
Control 1	-826.90	10.79	0.12	21.75
Control 2	-838.80	119.42	1.16	1.99
Control 3	77.90	24.33	0.22	7.99
Control 4	423.40	-339.16	1.50	0.43
Control 5	-46.60	251.61	1.47	0.96
Output	0.00	0.28	0.01	48.19
Power		0.26 μW		

NAND01	V (mV)	I (nA)	$I_{std}$ (nA)	$R_{diff}$ (MΩ)
Input 1	0.00	77.73	0.88	0.98
Input 2	500.00	-276.23	2.11	0.42
Control 1	-826.90	9.76	0.13	28.85
Control 2	-838.80	147.69	1.21	1.91
Control 3	77.90	48.95	0.22	2.74
Control 4	423.40	-203.48	1.12	0.57
Control 5	-46.60	182.3	1.06	1.38
Output	0.00	0.16	0.02	138.59
Power		0.36 μW		

NAND10	V (mV)	I (nA)	$I_{std}$ (nA)	$R_{diff}$ (MΩ)
Input 1	500.00	-1294.62	4.09	0.23
Input 2	0.00	236.68	2.06	0.75
Control 1	-826.90	28.89	0.25	12.94
Control 2	-838.80	1032.87	5.09	0.63
Control 3	77.90	27.72	0.21	6.37
Control 4	423.40	-256.28	1.96	0.52
Control 5	-46.60	212.44	1.39	1.11
Output	0.00	0.24	0.02	105.88
Power		1.65 μW		

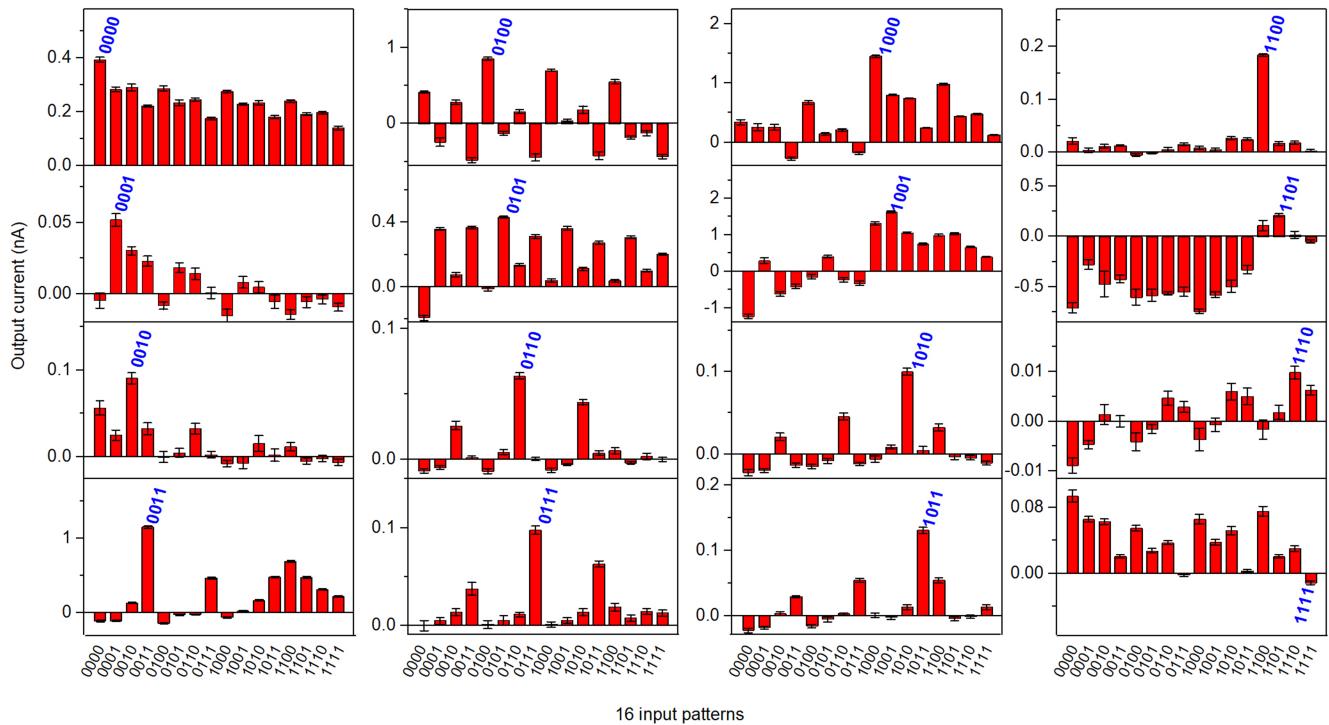
NAND11	V (mV)	I (nA)	$I_{std}$ (nA)	$R_{diff}$ (MΩ)
Input 1	500.00	-795.46	4.23	0.30
Input 2	500.00	-101.46	0.92	0.82
Control 1	-826.90	20.07	0.35	17.53
Control 2	-838.80	836.29	4.42	0.71
Control 3	77.90	48.01	0.35	2.70
Control 4	423.40	-173.57	1.19	0.64
Control 5	-46.60	161.28	1.10	1.20
Output	0.00	0.02	0.01	231.38
Power		1.24 μW		

computational technologies, the most energy efficient high-performance computer<sup>36</sup>, the neural network (NN) accelerator<sup>29</sup> and memristors<sup>37</sup> (Supplementary Note 8). **c**, Current flow pattern of a NAND gate (NAND10 in **d**) with inputs 500 and 0 mV. There is a large parasitic current flowing from input 1 to control electrode 2 (black curved arrow). This parasitic current limits the energy efficiency. This can be solved by using electrostatically coupled electrodes (Supplementary Note 8). **d**, Measured power consumption of a NAND gate for the four input combinations. The standard deviations in the current are calculated from ten measurements. The differential resistances  $R_{diff}$  are measured around the voltages in the second column.



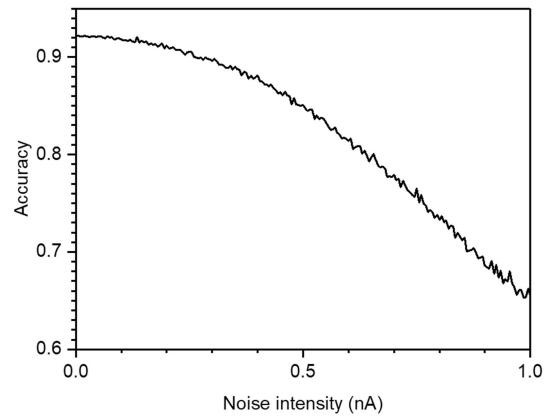
**Extended Data Fig. 8 | Backgate-induced nonlinearity and evolved logic gates at room temperature.** **a**, A positive voltage  $V_{Sub}$  with respect to the drain voltage is applied to the n-type substrate (Fig. 2b) to make the depletion region wider at the p–n junction, and to suppress the band conduction. **b**, Evolved gates at room temperature. Red circles are experimental outputs, and black lines represent the normalized desired outputs. The output current levels, and also the separation between these levels, are more than one order of magnitude larger than those of the logic gates evolved at 77 K, owing to the increased hopping conductance (Supplementary Note 3). The increased noise intensity is mainly due to the settings of the current measurement circuit (Methods).

# Article



**Extended Data Fig. 9 | Experimental response of the 16 filters.** Each of them is evolved to filter the feature given in blue. The output currents corresponding to features other than the desired one are not zero, but the output current of

the targeted feature is clearly separated from the other currents. Error bars represent the standard deviation obtained from ten tests.



**Extended Data Fig. 10 | Enhancing robustness of the linear classifier against noise.** Besides optimizing the SNR, the linear classifier's tolerance to noise can also be increased by taking noise into account during the training phase. The accuracy remains over 92% at 0.05 nA noise amplitude (see Supplementary Note 8 for a detailed discussion).