

Three-dimensional integration of nanotechnologies for computing and data storage on a single chip

Max M. Shulaker^{1,2}, Gage Hills¹, Rebecca S. Park¹, Roger T. Howe¹, Krishna Saraswat¹, H.-S. Philip Wong¹ & Subhasish Mitra^{1,3}

The computing demands of future data-intensive applications will greatly exceed the capabilities of current electronics, and are unlikely to be met by isolated improvements in transistors, data storage technologies or integrated circuit architectures alone. Instead, transformative nanosystems, which use new nanotechnologies to simultaneously realize improved devices and new integrated circuit architectures, are required. Here we present a prototype of such a transformative nanosystem. It consists of more than one million resistive random-access memory cells and more than two million carbon-nanotube field-effect transistors—promising new nanotechnologies for use in energy-efficient digital logic circuits^{1–3} and for dense data storage⁴—fabricated on vertically stacked layers in a single chip. Unlike conventional integrated circuit architectures, the layered fabrication realizes a three-dimensional integrated circuit architecture with fine-grained and dense vertical connectivity between layers of computing, data storage, and input and output (in this instance, sensing). As a result, our nanosystem can capture massive amounts of data every second, store it directly on-chip, perform *in situ* processing of the captured data, and produce ‘highly processed’ information. As a working prototype, our nanosystem senses and classifies ambient gases. Furthermore, because the layers are fabricated on top of silicon logic circuitry, our nanosystem is compatible with existing infrastructure for silicon-based technologies. Such complex nano-electronic systems will be essential for future high-performance and highly energy-efficient electronic systems⁵.

Physical device scaling of traditional silicon metal–oxide–semiconductor field-effect transistors (MOSFETs) has driven progress in computing for decades⁶; however, continued scaling is becoming increasingly difficult⁷. Consequently, there is a need for beyond-silicon nanotechnologies. Carbon-nanotube field-effect transistors (CNFETs) represent an emerging transistor technology that can scale beyond the limits of silicon MOSFETs, and promise an order-of-magnitude improvement in the energy–delay product (a metric of energy efficiency) of digital circuits^{1–3}, enabling energy-efficient computation. However, experimental demonstrations of CNFETs have been small-scale, limited to integrating only tens or hundreds of devices^{8–10}. Resistive random-access memory (RRAM) represents an emerging memory technology that promises high-capacity, non-volatile data storage, with improved speed, energy efficiency and density compared to dynamic random-access memory (DRAM)^{4,11}. Yet simply continuing to improve existing devices is insufficient, because data-intensive applications increasingly rely on being able to analyse massive volumes of data at very high rates¹². As the volume of data to be processed increases, the finite rate at which data can be transferred in a system, such as between off-chip memory and on-chip computing logic, results in a communication bottleneck^{13,14}. Therefore, new nanotechnologies must be leveraged to realize device improvements as well as new integrated circuit architectures that also address this communication bottleneck.

We present an experimental prototype of a new computing system, which integrates multiple new nanotechnologies to realize a 3D integrated circuit architecture. It contains RRAM arrays, silicon and CNFET computation units and memory access circuitry, and more than one million CNFET-based gas sensors for inputs, all fabricated on overlapping vertical layers. The key advances of our nanosystem compared to current technologies are that it uses: (1) CNFETs instead of traditional silicon-based integrated circuits; (2) on-chip non-volatile RRAM for data storage instead of DRAM-based off-chip memory; (3) monolithic 3D integration—a new computer architecture whereby layers of CNFET digital logic and RRAM are built vertically on top of each other on the same starting substrate¹⁵—instead of a single-layer of transistors; and (4) dense back-end-of-line metal wire vias to connect vertical layers of computing and data storage within the monolithic 3D integrated circuit (see Methods) instead of packaging techniques (such as interposers) or traditional 3D integration (which uses chip-stacking with through-silicon vias to connect vertical layers)¹⁶.

Monolithic 3D integration is naturally enabled by CNFETs and RRAM, owing to their low-temperature fabrication (see Methods). We fabricate CNFETs and RRAM at temperatures of at most 200 °C, whereas commercial silicon complementary metal–oxide–semiconductors (CMOSs) require temperatures of at least 1,000 °C (temperatures of more than 400 °C can damage bottom-layer transistors and wires¹⁷). The dense wire vias that are used for monolithic 3D integration can enable vertical connectivity that is 1,000 times more dense than conventional packaging and chip-stacking solutions allow¹⁸; this greatly improves the data communication bandwidth between vertically stacked functional layers, thus addressing the communication bottleneck⁵.

Owing to the fine-grained and dense connectivity between vertical layers of computing, data storage and sensing, 3D integrated circuit architectures, such as demonstrated in our nanosystem, can capture and directly store massive amounts of data from the outside world (see Methods), and perform *in situ* processing of the data captured to produce ‘highly processed’ results. As a demonstration of this, we use our nanosystem to sense ambient gases, to store the data and to perform *in situ* classification of the data—all on the same chip. The nanosystem we present here is the most complex nano-electronic system to be demonstrated so far, using prominent emerging nanotechnologies for energy-efficient digital logic and dense data storage in a 3D architecture.

Images of a fabricated chip are shown in Fig. 1, with a schematic shown in Fig. 2. All design and fabrication steps are wafer-scale and compatible with existing infrastructure for silicon-based technologies. The power, clock and control signals (such as memory addresses and read versus write voltages) are generated off-chip and applied through input/output pins. All of the sensors, data storage, logic and interconnects are pre-determined during design, and no post-fabrication selection, configuration or fine-tuning of functional components is

¹Department of Electrical Engineering, Stanford University, Stanford, California, USA. ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. ³Department of Computer Science, Stanford University, Stanford, California, USA.

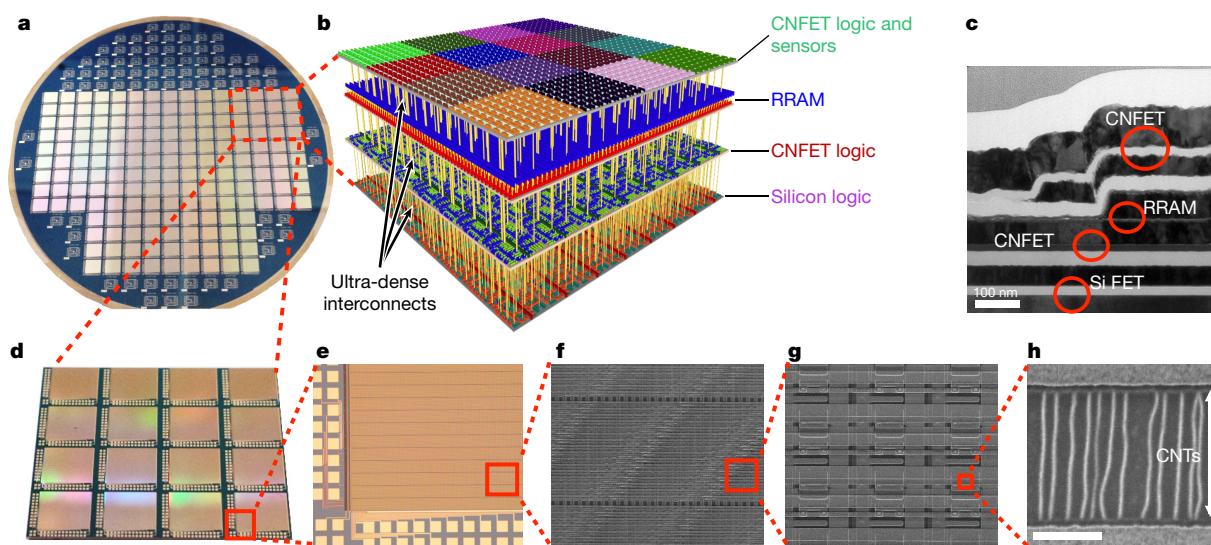


Figure 1 | Our 3D nanosystem. **a**, A 100-mm-wide wafer on which the integrated circuits are fabricated; one chip is outlined in red. **b**, Illustration of our nanosystem. It consists of four monolithically integrated vertical layers, connected through dense vertical inter-connects: fourth (top) layer, CNFET sensors and logic (including more than one million CNFET inverters, which operate as gas sensors); third layer, RRAM (1 Mbit); second layer, CNFET logic (the CNFET row decoders and CNFET classification accelerator); first (bottom) layer, silicon FET logic. **c**, Cross-sectional transmission electron microscopy (TEM) image of the four-layer chip, highlighting each layer. The brighter sections of the TEM image are

cross-sections of wires and the darker sections are different oxides (such as gate dielectrics or inter-layer dielectrics). Scale bar, 100 nm. **d–h**, Progressively magnified top views of our nanosystem, showing: the full chip (**d**; dimensions, 1.7 cm × 2.2 cm; scale bar, 500 nm); the input and output pins on the periphery, which are used to input and output signals (such as power, clock and control signals) to and from the chip (**e**); and carbon nanotubes (CNTs) bridging the source and drain contacts of a CNFET (**h**; scale bar, 500 nm). The images in **d** and **e** are optical microscopy images; those in **f–h** are scanning electron microscopy (SEM) images of sections of the chip.

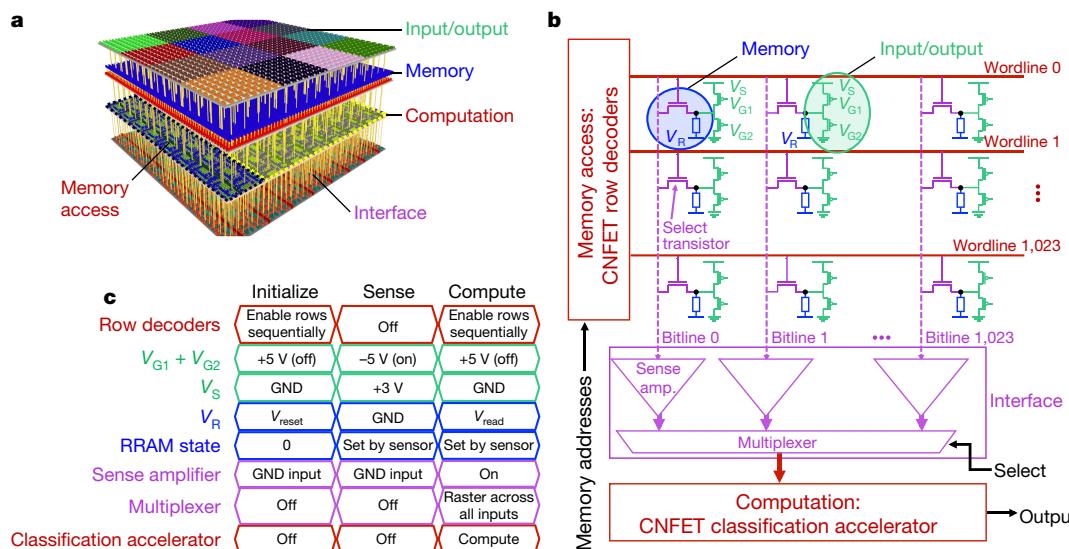


Figure 2 | Illustration, schematic and operation of our nanosystem.

a, Illustration of the subsystems within our nanosystem. The five subsystems (see text) are labelled, colour-coded to match **b** and **c**. **b**, Schematic of our nanosystem. The colour of each component corresponds to the layer of the 3D integrated circuit on which it is fabricated (see Fig. 1). The inputs to the chip are the memory addresses, the control signals (the ‘select’ signal for the multiplexer in the interface subsystem V_S , the control voltages to the sensing circuits V_{G1} and V_{G2} , and the reset or read voltage applied to the RRAM V_R), and the power and clock signals; these inputs are generated off-chip and routed on-chip through input/output pins. Signals with the same label (V_R , V_S , V_{G1} or V_{G2}) are connected on-chip to the same input/output pin. Wordlines (horizontal wires in the array, labelled 0 to 1,023) and bitlines (vertical wires in the array, labelled 0 to 1,023) are shown as red solid lines and purple dashed

lines, respectively. The sense amplifier (‘sense amp.’) is the circuitry that reads the value of the RRAM memory cell, and the select signal controls the multiplexer to select which sensor amplifier is connected to the computation subsystem. **c**, Our nanosystem operates in three phases: initialization, sensing and computation. In the initialization phase, all RRAM cells in the memory subsystem are reset and initialized to 0; voltage applied to reset the RRAM $V_{reset} = -2.75$ V. In the sensing phase, all CNFET sensors write either a 1 or a 0 (depending on how each sensor reacts with the ambient gas) into the RRAM cell underneath directly and in parallel. In the computation phase, the CNFET row decoders and silicon interface logic select individual RRAM cells sequentially (by the memory addresses and control signals generated off-chip), enabling the CNFET-based classification accelerator to perform classification; voltage applied to read the RRAM $V_{read} = 1.25$ V. GND is ground (0 V). See text and Methods for more details.

performed. Our nanosystem comprises four monolithically integrated vertical layers and five integrated subsystems (see Methods, Fig. 2a), as outlined in the following.

The first subsystem performs computation. The second vertical layer of the monolithic 3D integrated circuit implements a simple CNFET-based classification accelerator, which implements a basic support-vector-machine classifier¹⁹. This classifier computes on the raw data that is captured and stored from the CNFET gas sensors, producing a simple output (such as distinguishing between ambient vapours).

The second subsystem is the memory or data storage. The first and third vertical layers form a monolithically integrated 3D memory system. The memory array uses a cross-point architecture, with each cell containing one transistor and one RRAM (referred to as 1-transistor 1-RRAM, 1T1R, cells) (Fig. 2)¹¹. The RRAM cells (with a total array size of 1 Mbit) are fabricated on the third layer, vertically above the silicon select transistors on the first (bottom) layer (Fig. 1b–d).

The third subsystem provides access to the memory subsystem. The second layer has CNFET row decoders to drive the wordlines (wires) of the memory array, enabling the cells in that row to be read, set (programmed to have low resistance) or reset (programmed to have high resistance).

The fourth subsystem acts as the input/output. On the fourth layer are more than one million (2^{20}) CNFET inverters, functionalized to operate as simple chemical vapour sensors (hereafter referred to as the CNFET sensing circuits; Fig. 2b, c, Methods)^{20,21}. Each sensor is fabricated vertically above a memory cell in the memory subsystem. The fine-grained connectivity that is afforded by monolithic 3D integration enables each sensor to connect directly to its respective underlying memory cell with an inter-layer via. This in turns enables the sensors to write their data in parallel directly into memory, thus realizing massive sensing-to-memory bandwidth. Importantly, the on-chip sensing is used only for demonstration purposes; the top layer could be replaced with additional computation or data storage subsystems, or with other forms of input/output.

The fifth subsystem acts as interface circuitry. On the first layer are additional silicon-based interface circuits, including sense amplifiers to read the memory array and multiplexers to route the memory to the CNFET computation subsystem.

The silicon transistors are fabricated on the bottom layer, because the temperature at which they are processed (more than 1,000 °C) precludes fabrication on the upper layers. The ordering of the other layers within the 3D integrated circuit reflect the data flow. The CNFET gas sensors are placed on the fourth (top) layer so that they can write directly into the RRAM on the third layer. The CNFET memory-access subsystem and CNFET-based classification accelerator are placed beneath the RRAM, on the second layer.

To demonstrate the functionality of our nanosystem, we perform environmental monitoring, sampling the air and classifying ambient vapours. There are three phases of operation (Fig. 2c): initialization, sensing and computation.

First, the memory is initialized by resetting all RRAM cells to 0 (high-resistance state). Using the CNFET memory-access circuitry, rows (that is, wordlines) of the memory array are enabled sequentially, and silicon select transistors (on the first layer) reset the RRAM. The CNFET sensing circuits (fourth layer) are powered down during this phase (their supply voltage is disconnected and all CNFET gates in the CNFET sensing circuits are biased off).

Second, during a 10-μs pulse, all of the CNFET sensing circuits are turned on simultaneously and generate an output voltage that depends on the response of each functionalized sensor to the ambient gases. Each output voltage is written into the respective RRAM cell (which lies vertically underneath the relevant sensing circuit) directly, through a connecting inter-layer via. If the output voltage is sufficiently high (above the required voltage to set the RRAM in a low-resistance state), then the RRAM cell is set to 1; otherwise, the RRAM remains in the initialized reset state of 0. This process occurs in parallel across all CNFET

sensing circuits into the entire 1-Mbit memory array, allowing for massive sensing-to-memory bandwidth.

Third, to classify the recorded sensor data, the CNFET memory access circuitry turns on each wordline of the memory array sequentially. A pattern of ‘1’s and ‘0’s is read from the memory array, corresponding to the response of each CNFET sensing circuit during the sense phase. The CNFET classification accelerator classifies the ambient gas by counting the number of matches between the pattern that is read out to previously learned patterns, making a positive identification when the count exceeds a pre-set threshold (see Methods).

In Fig. 3 we show the electrical characterization of the technologies that are integrated in our nanosystem. A fundamental building block of our nanosystem is a monolithic 3D cell (Fig. 3c), which contains one silicon select transistor (on the first layer), one RRAM cell (third layer) and one CNFET inverter (the CNFET sensing circuit; fourth layer). Our nanosystem integrates more than one million such monolithic 3D cells. Figure 3d illustrates the operation of one monolithic 3D cell: the RRAM can be programmed both by the silicon select transistor underneath and by the CNFET inverter above (the silicon select transistors reset the RRAM during the initialization phase, whereas the CNFET inverters set the RRAM during the sense phase). To demonstrate that all design and fabrication steps are robust and wafer-scale, we measure 100,000 individual monolithic 3D cells across a 100-mm wafer as a control. As shown in Fig. 3e, the yield (defined as the combined yield of all components within the 3D cell as well as correct cell operation over five repeated set-reset cycles) is 98.82%. The primary causes of yield loss (see Methods) are attributed to performing all fabrication in an academic facility. The functionality of our nanosystem is robust to such inadvertently faulty monolithic 3D cells, because many classification algorithms (including the basic support vector machine that is implemented in our nanosystem) naturally exclude errors (see Methods).

To test our nanosystem, we loaded it into a vacuum test chamber, which is filled with different vapours (see Methods). As a demonstration, our nanosystem is trained to distinguish between common gases and vapours: pure nitrogen, and the vapours of lemon juice, white vinegar, rubbing alcohol, vodka, wine and beer. In Fig. 4a we show samples of the raw sensing data that are generated under each ambient gas. Each pixel corresponds to the value that a CNFET sensing circuit (fourth layer) writes into a RRAM cell (third layer). The on-chip CNFET classification accelerator identifies the ambient gas by processing the raw sensor data and comparing the response of the sensor array to previously learned expected responses to each of the gases. The results upon exposure to lemon juice vapours and to rubbing alcohol vapours are shown in Figs 4b, c and 3d. The output of the CNFET-based classification accelerator (Fig. 4b, c) demonstrates correct classification of both gases. It exceeds the pre-determined classification threshold when comparing the response to lemon juice vapours to the previously learned response to lemon juice vapour, and when comparing the response to rubbing alcohol vapours to the previously learned response to rubbing alcohol vapour. Although our nanosystem uses a basic support vector machine with a single feature for classification, alternative classification and pattern-recognition techniques could be used¹⁹. For instance, in Fig. 4d we show the classification of the seven gases that we tested (see Fig. 4a) using an off-chip classifier that computes two different features as criteria for classification. Principal-component analysis of these two features, computed by the off-chip classifier, confirms that our nanosystem generates a unique response to each of the gases that we tested, and can therefore distinguish between them (Fig. 4d).

Although the specific 3D nanosystem implementation presented here performs vapour classification, it highlights an example of how new nanotechnologies can be implemented to enable a new generation of nanoelectronic computing systems⁵. As an integrated nanosystem,

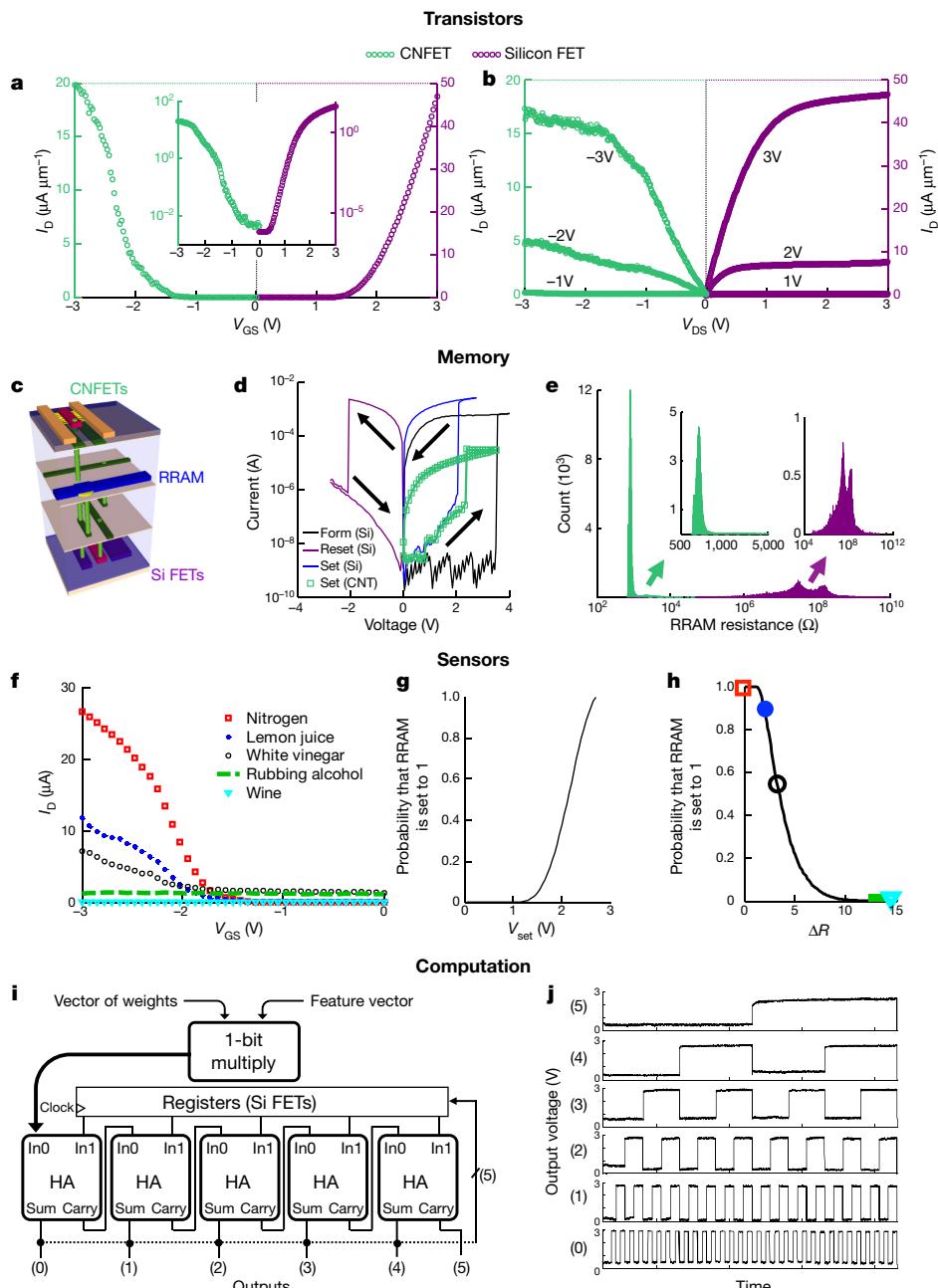


Figure 3 | Characterization of the components of our nanosystem. **a, b**, Plots of drain current I_D versus gate-source voltage V_{GS} for a drain-source voltage $|V_{DS}| = 3$ V (**a**) and I_D versus V_{DS} for different values (labelled) of V_{GS} (**b**) for a typical CNFET (second layer; red) and silicon FET (first layer; purple). The inset in **a** shows the same data on a logarithmic scale. **c**, Illustration of a monolithic 3D cell, used as the building block for our nanosystem (more than one million are used). **d**, Functionality of the monolithic 3D cell in **c**. Typical current–voltage curves showing how the RRAM is set and reset by the silicon FET, and then set again by the CNFET inverter. The black arrows show the direction of the voltage sweep. The ‘Form Si’ curve shows the initial formation of the RRAM filament, which occurs before setting and resetting. **e**, Results from 100,000 monolithic 3D cells, showing the distribution of the resistance of the RRAM in the ‘on’ state (set by CNFETs; orange) and the ‘off’ state (reset by silicon FET; purple). The insets show close-ups of the regions of interest. **f**, Plots of I_D versus V_{GS} for a functionalized CNFET gas sensor (see Methods). When the CNFET is exposed to various gases, its effective resistance changes. **g**, The probability (cumulative distribution function) that an RRAM cell is set to 1 as a function of the output voltage V_{set} supplied by a CNFET inverter (measured from 30,000 RRAM and CNFET inverters). The CNFET inverter (comprising a CNFET

gas sensor) converts the change in the resistance of the CNFET sensor to a change in output voltage, which acts as the programming voltage of the RRAM. The probability that the RRAM is set to 1 (low-resistance state) is determined by the set voltage V_{set} . **h**, As the resistance of the CNFET gas sensor increases, the probability of setting the RRAM to 1 decreases. The change in the resistance of the sensor ΔR is defined as the resistance measured upon exposure to gas divided by the baseline resistance in vacuum. The percentage of monolithic 3D cells that set their RRAM to 1 corresponds to the probability of each individual sensing cell setting their RRAM to 1. The symbols highlight the probability that an RRAM is set to 1 when the sensors are exposed to the gases shown in **f**, based on the ΔR associated with the sensor when exposed to that gas. **i, j**, Characterization of the computation subsystem. The combinational logic is implemented with CNFETs on the second layer and the sequential logic (latches) is implemented with silicon FETs on the first layer. The vector of weights are learned and stored off-chip. The feature vector is obtained from the memory subsystem. **i**, Schematic of the classification accelerator (HA, half-adder; sum, the sum output from the HA; carry, the carry output from the HA; In0 and In1, the two inputs to the HA). **j**, Measured output waveforms, demonstrating correct, functional operation by testing every input combination (test frequency, 2 kHz).

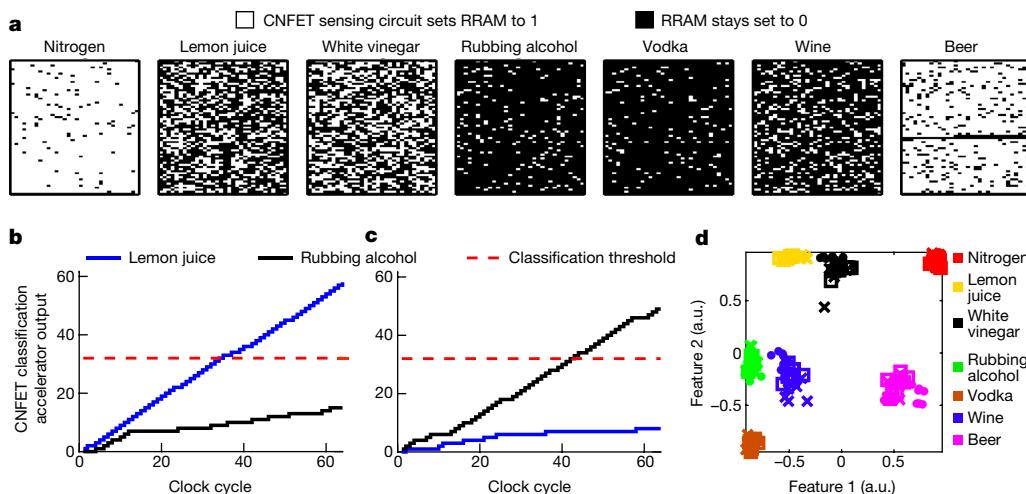


Figure 4 | Results from our nanosystem. **a**, Sensor data (generated from the fourth layer) written into the RRAM (on the third layer) from a sample of 2,048 monolithic 3D cells, measured under seven different ambient gases. A white pixel corresponds to the RRAM in that monolithic 3D cell setting to 1 during the sensing phase; a black pixel corresponds it staying set to 0. **b**, **c**, Measured output from a CNFET-based classification accelerator upon exposure to vapours of lemon juice (**b**) and rubbing alcohol (**c**). Our nanosystem compares the measured output (from 128

monolithic 3D cells) to previously learned vectors of weights (learned and stored off-chip). Each gas is correctly classified: in both cases, the only output that exceeds the positive classification threshold corresponds to the gas that was present during sensing. **d**, Principal-component analysis (performed off-chip) of the sensor data shown in **a** demonstrates the ability to correctly classify nitrogen and six vapours, illustrating that our nanosystem functions correctly. a.u., arbitrary units.

fabricated using beyond-silicon nanotechnologies and consisting of a monolithic 3D architecture with vertically interleaved layers of computing and data storage with fine-grained and dense connectivity, our nanosystem represents a considerable and important advance in the field of computing.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 August 2016; accepted 2 May 2017.

- Chang, L. Short course. In *IEEE International Electron Devices Meeting (IEDM)* (2012).
- Franklin, A. et al. Sub-10 nm carbon nanotube transistor. *Nano Lett.* **12**, 758–762 (2012).
- Wei, L., Frank, D. J., Chang, L. & Wong, H. S. P. A non-iterative compact model for carbon nanotube FETs incorporating source exhaustion effects. In *IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2009).
- Wong, H. S. P. & Salahuddin, S. Memory leads the way to better computing. *Nat. Nanotechnol.* **10**, 191–194 (2015).
- Aly, M. M. S. et al. Energy-efficient abundant-data computing: the N3XT 1,000x. *Computer* **48**, 24–33 (2015).
- Dennard, R. H., Gaenslen, F. H., Rideout, V. L., Bassous, E. & LeBlanc, A. R. Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE J. Solid-State Circuits* **9**, 256–268 (1974).
- Frank, D. J. et al. Device scaling limits of Si MOSFETs and their application dependencies. *Proc. IEEE* **89**, 259–288 (2001).
- Cao, Q. et al. Medium-scale carbon nanotube thin-film integrated circuits on flexible plastic substrates. *Nature* **454**, 495–500 (2008).
- Shulaker, M. M. et al. Carbon nanotube computer. *Nature* **501**, 526–530 (2013).
- Shulaker, M. M. et al. Sensor-to-digital interface built entirely with carbon nanotube FETs. *IEEE J. Solid-State Circuits* **49**, 190–201 (2014).
- Wong, H. S. P. et al. Metal-oxide RRAM. *Proc. IEEE* **100**, 1951–1970 (2012).
- Mayer-Schönberger, V. & Cukier, K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2013).
- Rogers, B. M. et al. Scaling the bandwidth wall: challenges in and avenues for CMP scaling. In *ACM SIGARCH Computer Architecture News Vol. 37*, 371–382 (ACM, 2009).
- Villa, O. et al. Scaling the power wall: a path to exascale. In *Proc. International Conference for High Performance Computing, Networking, Storage and Analysis (SC14)* 830–841 (IEEE, 2014).
- Shulaker, M. M. et al. Monolithic 3D integration of logic and memory: carbon nanotube FETs, resistive RAM, and silicon FETs. In *IEEE International Electron Devices Meeting (IEDM)* 27–34 (IEEE, 2014).
- Leduc, P. et al. Enabling technologies for 3D chip stacking. In *International Symposium on VLSI Technology, Systems and Applications (VLSI-TSA)* 76–78 (IEEE, 2008).
- Vinet, M. et al. 3D monolithic integration: technological challenges and electrical results. *Microelectron. Eng.* **88**, 331–335 (2011).
- Batude, P. et al. Advances, challenges and opportunities in 3D CMOS sequential integration. *IEEE Int. Electron Devices Meeting (IEDM)* 7–13 (IEEE, 2011).
- Steinwart, I. & Christmann, A. *Support Vector Machines* (Springer, 2008).
- Liu, S. F., Moh, L. C. & Swager, T. M. Single-walled carbon nanotube–metalloporphyrin chemiresistive gas sensor arrays for volatile organic compounds. *Chem. Mater.* **27**, 3560–3563 (2015).
- Kong, J. et al. Nanotube molecular wires as chemical sensors. *Science* **287**, 622–625 (2000).

Acknowledgements We acknowledge the support of NSF (CNS-1059020), DARPA (W9009MY-16-1-0001), STARnet SONIC, member companies of the Stanford SystemX Alliance, and the Hertz Fellowship and Stanford Graduate Fellowship for M.M.S. We are grateful to C. Gupta for discussions.

Author Contributions M.M.S. led and was involved in all aspects of the project, and performed all of the design, layout, fabrication and testing. G.H. contributed to the design and testing. R.S.P., R.T.H. and K.S. contributed to the design of the silicon transistors. H.-S.P.W. and S.M. were in charge and advised on all parts of the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.M.S. (shulaker@mit.edu).

METHODS

Fabrication. The fabrication flow is shown in Extended Data Fig. 1, with details below.

First, a conventional starting silicon substrate is fabricated, using depletion-load negative-channel MOS (NMOS) logic. The silicon logic is fabricated using a conventional silicon FET fabrication flow. The technology node is defined by the lithographic limitations of an academic fabrication facility, with a minimum feature of about 500 nm. The silicon FETs consist of a 25-nm SiO₂ gate dielectric, titanium/platinum contacts and a titanium/platinum metal gate (all deposited using electron-beam evaporation and defined using a photolithographic lift-off process). Owing to the approximately 1-μm technology node, the silicon circuits operate at a 3-V supply. However, with more advanced lithography and chip fabrication capabilities, this bottom layer of silicon could be implemented at a scaled technology node, without affecting the subsequent monolithic 3D integrated circuit (IC) processing. The silicon logic is used only to demonstrate that all subsequent monolithic 3D processing is compatible with silicon CMOS logic.

Second, following completion of each vertical layer, a thin inter-layer dielectric (ILD) is deposited, followed by etching and filling in with metal inter-layer vias (ILVs). The ILD is deposited through a low-temperature (180 °C) plasma-enhanced chemical vapour deposition (PECVD) oxide (nominally 50 nm after subsequent processing steps). The metal ILVs are then patterned lithographically, etched using a directional plasma-reactive ion etch, and then filled with metal (1-nm adhesion layer of titanium, followed by platinum). The tightest pitch ILVs in our monolithic 3D IC is 2 μm (1 μm ILV diameter, 1 μm ILV spacing), which is set by the lithographic resolution of our academic stepper. These ILVs can have the same pitch and dimensions as tight-pitched metal layer vias (about 100 nm), and can be fabricated using a standard damascene process. Therefore, existing ILD and ILV technologies can be used within monolithic 3D ICs.

Third, the CNFET logic, which makes up the second layer, is fabricated directly over the deposited ILD. A full description of the CNFET fabrication flow is given in Methods section ‘CNFETs’. To begin the CNFET fabrication, highly aligned carbon nanotubes (CNTs) are first grown on a crystalline quartz substrate (ST-cut quartz, 100-mm substrate, at about 850 °C with methane gas and with iron metal nanoparticles as the catalyst^{22,23}). The CNTs grow horizontally on the substrate, along the crystalline plane of the quartz. Following growth of the CNTs, the quartz wafer is coated with a 150-nm-thick layer of gold (or copper). A thermal release tape²² is applied over the metal and is peeled from the substrate, removing the metal and all of the CNTs (embedded within the metal). This tape is then placed on the target substrate (the monolithic 3D IC, directly over the ILD and ILVs, above the silicon FETs). This layer-transfer process is similar to processes used to fabricate silicon-on-insulator wafers, although the requirements on the target substrate are relaxed for the CNT layer-transfer process, because the CNTs are able to conform to topology on the wafer surface. The only alignment necessary is rotational alignment—the CNTs are transferred so that they are perpendicular to the source and drain contacts that will be patterned (rotational alignment is more simple and can routinely be performed to with microradian accuracy using bonding equipment)—translational alignment is not required. The CNTs are transferred everywhere on the wafer substrate, and the subsequent CNFET fabrication is performed without any knowledge of the exact locations of the CNTs (only their general directional alignment). In contrast, if the entire CNFET circuitry was transferred onto the substrate, alignment inaccuracy would limit ILV density. After the layer transfer, the substrate is heated to approximately 130 °C to remove the thermal release tape, and the transfer metal is removed with an etchant that does not damage the CNTs (dilute potassium iodide solution), leaving the highly aligned CNTs on the target substrate. The CNFET source and drain contacts are then patterned using traditional photolithography. The lithography uses the original alignment marks on the wafer substrate—the same alignment as used for the silicon FETs. Therefore, the only inaccuracy in the lithography is due to overlay inaccuracy, which is not affected by the CNT layer transfer. Following the CNFET source and drain contact definition, the remaining CNFET circuit fabrication is performed (for complete details of the CNFET fabrication, see Methods section ‘CNFETs’). The CNFETs are fabricated at the same technology node as the silicon FETs (approximately 1-μm technology node), and operate at the same 3-V supply voltage (owing to the 1-μm node implementation). CNFETs have been experimentally demonstrated to scale to sub-10-nm technology nodes²; in Methods section ‘CNFETs’, we describe how all of the CNFET processing techniques that we use can be applied (and have been experimentally demonstrated) to achieve highly scaled CNFET circuits.

Fourth, an ILD is again deposited and ILVs are patterned between every layer of the 3D IC. This step is identical to the second step, although the locations of the ILVs can vary, as dictated by the design.

Fifth, the RRAM, which makes up the third layer of the 3D IC, is a bipolar stack, with a platinum bottom electrode, a 5-nm-thick hafnium oxide (HfO_x) dielectric

deposited at 200 °C using thermal atomic-layer deposition (ALD), and a titanium nitride/titanium/platinum stack top electrode. The maximum processing temperature that is required for the RRAM fabrication is less than 200 °C. The RRAM characteristics are shown in Fig. 3 (for a sample size of more than 30,000 RRAM cells) and has a form voltage of about 3 V, a set voltage of about 2 V and a reset voltage of about −2 V. This RRAM stack was optimized to achieve these voltages, which match the voltage range of all of the logic and sensors (described later) used in the monolithic 3D IC (3-V supply voltage, so the RRAM set voltage is within this range). A 1-Mbit RRAM array is fabricated using a cross-point architecture, with one-transistor-one-RRAM (1T1R) structures, with the select transistor being a silicon FET transistor. The 1T1R structure is required to avoid leakage paths during the read operations. The size of the RRAM cell is approximately 500 nm², as determined by our lithographic resolution.

Sixth, an ILD is again deposited and ILVs are patterned between every layer of the 3D IC. This step is identical to the second step, although the locations of the ILVs can vary, as dictated by the design.

Seventh, and finally, the CNFET logic and CNFET sensors (which make up the fourth layer) are fabricated using the processes described previously. In addition to the CNFET fabrication, some of the CNFETs on the top layer are functionalized to turn into CNFET gas and vapour sensors. This process is described in detail in Methods section ‘CNFETs’.

This monolithic 3D IC fabrication is a baseline process. For future monolithic 3D ICs, additional layers can be integrated into the stack. For instance, future 3D ICs might require additional temperature management during operation to maintain low skin temperature for mobile and wearable systems. Several techniques, including electrothermal co-design, integrated micro- or nanoscale heat convection and conduction solutions, and embedded cooling technologies, are currently being pursued⁵.

All CNFET logic is depletion-load positive-channel MOS (PMOS) logic, whereas all silicon logic is depletion-load NMOS logic, owing to ease of fabrication; CMOS logic using both CNFETs and silicon FETs has previously been demonstrated experimentally.

For chip packaging, the chips are diced with an automated dicing saw, and are then mounted into standard dual in-line packages. The input/output pins from the chip are wire-bonded to the package for testing, using standard packaging techniques. The total height of the chip is about 555 μm (the wafer substrate itself is a standard approximately 550-μm-thick thin wafer; the height of the actual circuitry fabricated in the monolithic 3D IC is less than 5 μm).

Design decisions and analysis. Our 3D nanosystem was designed to illustrate a new generation of computing systems that can overcome major challenges facing computing systems today. To overcoming the challenges of energy efficiency and scaling, we integrate CNFETs for computing, which promise an order-of-magnitude improvement in energy–delay product compared to silicon CMOS^{2,3}, and RRAM for on-chip memory, which promises high-capacity non-volatile data storage with improved speed, energy efficiency and density compared to DRAM¹¹. To overcome the communication challenge, combinations of these emerging nanotechnologies of logic and memory (and sensing) are fabricated vertically above each other in a monolithic 3D fashion. Such a design provides more fine-grained vertical interconnects compared to conventional chip-stacking (detailed below), enabling high-bandwidth and low-latency communication between vertical circuit layers.

Although our nanosystem experimentally demonstrates all of these advances, it also faces practical limitations associated with fabrication in an academic nanofabrication facility. For instance, although CNFETs can be scaled to sub-10-nm nodes, an approximately 1-μm technology node is used for all layers of the monolithic 3D IC. This size is set by the minimum lithographic feature (approximately 500 nm) in our academic nanofabrication facility. Previous works (described in detail in Methods section ‘CNFETs’) have demonstrated that all of the techniques that we used for CNFET fabrication can be applied to highly scaled CNFET logic as well. The lagging technology also effects other metrics of the chip, such as area, speed and power. The area of the chip is 1.7 cm × 2.2 cm, although chip area decreases as technology node scales. The supply voltage is 3 V (a conventional supply voltage for a 1-μm technology node), whereas a highly scaled CNFET technology could operate at a scaled supply voltage of around 400 mV (ref. 24); this effects the power and speed of the 3D IC. The power consumption for an individual CNFET classification accelerator (shown in Fig. 3) is about 40 μW and the operation frequency is 2 KHz. The operation frequency is limited by both the lagging technology node and the fact that the CNFET circuits drive the output pins from the chip directly, which adds substantial capacitance load on the circuit. Moreover, we intentionally clock the chip at a low frequency to be conservative and to avoid timing errors; the motivation is to show feasibly and functionality. The power consumption is determined by the supply voltage and the implementation of the silicon FET and CNFET logic: both use depletion-load logic, which

has static power consumption (contrary to complementary logic). We implement depletion-load logic owing to the simplicity in the fabrication process flow (for both the silicon FETs and CNFETs), which is an important yield consideration in academic nanofabrication facilities. However, complementary logic for silicon FETs and CNFETs has been demonstrated and is not a fundamental obstacle for either technology.

Despite the fabrication limitations associated with an academic nanofabrication facility, the benefits of monolithic 3D ICs can still result in substantial improvements compared to the current state-of-the-art. The connectivity that is demonstrated experimentally in our nanosystem exceeds the capabilities of current chips, despite the fact that commercial fabrication facilities have much better fabrication capabilities (commercial facilities produce 14-nm node technologies compared to our 1- μm technology node, with orders of magnitudes improved lithography). Connectivity is a key metric used to describe the physical connectivity between different parts of a system—the higher the connectivity, the larger the number of concurrent accesses that can be made between them. The connectivity is determined by the pitch of the wires that connect parts of a system. For 2D chips packaged on board, connectivity is determined by the number of input and output pins on each chip. The number of such pins can reach only several thousand, and not all pins can be used for communication between two chips—pins have to be used as inputs and outputs to all forms of input/output as well as to deliver other signals to the chip (such as power). In contrast, our approach of monolithic 3D integration allows vertically overlapping parts of systems to be connected with dense ILVs. Using our lagging 1- μm technology node, there are more than one million ILVs connecting vertical layers in our nanosystem—an increase in connectivity of more than 1,000-fold. A scaled technology node (with approximately 100-nm ILV pitch) would result in an increase in connectivity of more than 10,000-fold. In comparison, conventional 3D chip stacking, which stacks and bonds separate chips vertically above each other, uses through-silicon vias (TSVs) and micro-bumps as vertical interconnects. However, conventional TSV pitch is limited to about 20–40 μm , and the best TSV pitch approaches about 10 μm (ref. 16). In our approach of monolithic 3D ICs, the ILVs, which can have the same pitch and dimensions as tight-pitched metal layer vias (around 100 nm), result in connectivity that is more than 1,000 times greater than in conventional chip stacking. Our nanosystem, with 2- μm tightest-pitch ILVs, thus achieves record connectivity with a vertical inter-connect density that is more than an order of magnitude greater than that achieved by state-of-the-art conventional 3D chip stacking, despite using a lagging technology; these benefits will increase to three orders of magnitude at an advanced technology node.

This increased connectivity in our nanosystem translates to an increase in bandwidth between vertical layers. In our experiments, the peak bandwidth that is achieved between vertical layers of the chip is determined by the connectivity between layers and the operating frequency. For instance, the bandwidth between the sensing layer and the memory layer is determined by the time that is required to transfer data between these layers, multiplied by the connectivity between them. In our experimental demonstration, all sensors write directly into their own RRAM cell in parallel. Using a 10- μs pulse time to write the RRAM and a 1-Mbit array of RRAM, the peak bandwidth of the system is 12.5 GB s⁻¹. However, a 10- μs pulse time is chosen conservatively to demonstrate functionality—a pulse time of less than 10 ns is sufficient for a writing operation to RRAM¹¹. If the pulse time used to write the RRAM was shortened to 100 ns (an order of magnitude higher than necessary), the sensor-to-memory bandwidth would be 10 Tbits s⁻¹. For comparison, the peak memory bandwidth (between memory chips and processing cores) for representative existing systems are 76.8 GB s⁻¹ for a desktop (Intel Core i7) and 29.8 GB s⁻¹ for mobile (Intel Core M).

Gas and vapour sensing was chosen purely as a demonstration platform for this technology, because it involves the integration of computation, memory and sensing within a single IC. For this demonstration, for the inputs/outputs for the chip, the sensors act as the inputs and the results from on-chip computations are the outputs. The performance of the CNFET sensors (including reproducibility, sensitivity, drift, response time, reversibility and selectivity) is described in detail in Methods subsection ‘Gas and vapour sensing’. Gas and vapour sensing is chosen simply as a prototype. The CNFET gas sensors used here are similar to CNFET gas sensors developed previously. Additionally, the individual responses of individual CNFET gas sensors is not as critical here. As shown in Fig. 4, regardless of the accuracy of an individual CNFET gas sensor, its output is digitized into only a single bit (logical ‘1’ or ‘0’), stored in the RRAM. Classification is performed by analysing sets of CNFET gas sensors, rather than the output of a single sensor.

CNFETs. Realizing VLSI digital systems. Despite the promise of CNFETs for a high-performance and energy-efficient very-large-scale integration (VLSI) digital logic technology, substantial inherent imperfections have prevented CNFET systems from being realized. These imperfections include the presence of mis-positioned CNTs (which result in stray conducting paths that lead to

incorrect logic functionality) and of metallic CNTs (which, owing to the diameter and chirality of CNTs, have little or no bandgap, resulting in high leakage currents and incorrect logic functionality). To overcome these imperfections, we use the ‘imperfection-immune paradigm’ (IIP)²⁵—a set of processing and design techniques that overcome these obstacles in VLSI-compatible manner. To overcome the presence of mis-positioned CNTs, we first perform wafer-scale aligned CNT growth²³. The CNTs are grown on a crystalline substrate (ST-cut quartz), resulting in greater than 99.5% CNT alignment along the crystalline plane of the quartz. These CNTs are then transferred from the growth substrate onto the target substrate where final circuit fabrication occurs (described previously in Methods section ‘Fabrication’). This transfer is a low-temperature transfer (at most 130 °C), and maintains both the alignment and density of the CNTs. This low-temperature transfer also decouples the high-temperature (more than 800 °C) CNT growth from the wafer where the final circuits are fabricated, allowing CNT circuits to be fabricated in a monolithic 3D fashion. Layer transfers are therefore key processing steps, and are similar to processes that are currently in use (for instance, in fabrication of silicon-on-insulator wafers). To overcome the remaining (at most 0.5%) mis-positioned CNTs, we use ‘mis-positioned CNT-immune design’²⁶. This design technique ensures that the resulting circuit is immune to any possible mis-positioned CNTs. It can also be applied to any arbitrary logic function, and does not require any die-specific customization. Moreover, it has a much smaller effect (in terms of area, power and speed) than traditional redundancy-based defect-tolerance techniques. To overcome the presence of metallic CNTs, we use a combined processing and circuit-design technique known as ‘VLSI-compatible metallic CNT removal’ (VMR)²⁷. VMR is a chip-scale electrical breakdown technique that removes metallic CNTs from an ensemble of metallic and semiconducting CNTs from the wafer. Electrical breakdown is performed by using the gates of the transistors to turn off semiconducting CNTs, after which a source-drain bias is applied. Although the semiconducting CNTs are turned off by the gate, the metallic CNTs are not (by virtue of being metallic), and therefore conduct current. With sufficient source-drain bias, the metallic CNTs flow enough current that they heat up, owing to a Joule self-heating process, and eventually breakdown (much like a fuse, at around 600 °C). Electrical breakdown has been shown to remove at least 99.99% of metallic CNTs, with at most 4% of semiconducting CNTs being inadvertently removed²⁸, which is sufficiently selective for VLSI systems. Recent work on modifying the electrical breakdown process has achieved even higher selectivity, with at least 99.99% of metallic CNTs and at most 1% of semiconducting CNTs being removed, across any arbitrarily scaled technology node²⁹. VMR allows this process to be applied at the chip-scale, across all CNFETs. It is also applicable to any arbitrary logic function, and follows VLSI design and processing flows. After applying the IIP, arbitrary CNFET digital logic systems can be fabricated.

Highly scaled CNFETs. As previously stated, the CNFETs are implemented at a 1- μm technology node, owing to the 500-nm minimum lithographic feature in academic nanofabrication facilities. However, this is not a fundamental challenge, and highly scaled CNFETs have previously been experimentally demonstrated^{2,28}. The entire IIP (described above) can be applied to highly scaled technology nodes, without any decrease in effectiveness. In addition, given that mis-positioned CNT-immune design is a layout technique, it is effective regardless of scaling. Metallic CNT removal has been demonstrated for CNFETs with sub-20-nm channel lengths and with 14-nm technology nodes, without affecting the properties of the process. Using these techniques, smaller-scale highly scaled CNFET circuits, such as an integrated sensor/sensor-interface circuit with 32-nm channel length CNFETs, and CNFETs with 14-nm technology nodes, have been fabricated and tested²⁹. In addition to these experimental demonstrations, projections based on experimentally calibrated CNFET models predict that CNFETs will achieve an order of magnitude benefit compared to silicon CMOS at highly scaled sub-10-nm technology nodes and that they will be able to scale beyond the limitations of silicon CMOS to sub-10-nm technology nodes.

Classification accelerator. The CNFET-based classification accelerator implements a support-vector-machine classifier¹⁹. The classifier computes the scalar product of a feature vector and a vector of weights. This scalar product is then compared to a threshold: if the product exceeds the threshold, positive classification is assigned. The vector of weights is predetermined to maximize the classification accuracy for separate training sets, and is learned and stored off-chip. A schematic of the CNFET-based classification accelerator is shown in Extended Data Fig. 2. The feature vector is formed by the values written in the memory array, accessed sequentially through the silicon interface circuitry. The CNFET logic then performs a sequential bit-wise multiplication between the feature vector and the vector of weights and adds the result to the running total through the accumulator; for example, if element i of the vector of weights has a value of 1 and element i of the feature vector also has a value of 1, then a 1 is added to the accumulator. After sequentially scanning through all elements in the vector, the

final output from the CNFET accumulator is a sum corresponding to the scalar product of the two vectors. An example of this process is shown in Extended Data Fig. 3. In Extended Data Fig. 4 we show the measured output from an isolated CNFET-based classification accelerator for which the feature vector and vector of weights contain only 1s, thereby operating the entire system as an incrementer and testing all possible outputs.

Gas and vapour sensing. As a demonstration of the enhanced functionality that can be achieved with emerging nanotechnologies, we operate more than one million CNFET inverters as gas sensors. A schematic of the CNFET gas sensors is shown in Extended Data Fig. 5. The liquids used in our tests were: lemon juice concentrate (ReaLemon), white vinegar (Heinz), rubbing alcohol (91% isopropyl alcohol; CVS), vodka (Tito's Handmade Vodka, 80 proof), wine (Yellow Tail sauvignon blanc) and beer (Stella Artois). The output of the sensor is taken as the output voltage from the inverter. The pull-down CNFET is covered with an evaporated oxide, which makes it insensitive to the ambient air. The pull-up CNFET is exposed to the ambient air, and thus its performance is affected by the gases and vapours therein. When a voltage is applied across the inverter, the two transistors act as two resistors in series, and thus a voltage divider. Therefore, as the resistance of the sensor (R_{sensor}) changes in response to different gases, the output voltage also changes. The gate biases of the two transistors (V_{G1} and V_{G2}) are chosen to bias the output voltage slightly above the V_{set} of the majority of the RRAM cells (about 3.3 V). Consequently, when the supply voltage is turned on during the sensing phase ($V_S = 4$ V), all of the RRAM cells in the array should be set to 1. However, if the gas or vapour introduced into the chamber increases R_{sensor} , the output voltage will decrease. If R_{sensor} increases by a sufficiently large amount that the output voltage decreases to below the required voltage to set any of the RRAM cells in the array, then all of the RRAM cells in the array will remain set to 0 during the sensing phase. If R_{sensor} increases by a lesser amount, such that the output voltage is within a range to set some but not all of the RRAM, then a proportion of the RRAM cells in the array will be set to 1 during the sensing phase, determined by the average value of the output voltage. Therefore, although each individual sensor encodes only a single bit value (1 or 0; or yes or no), by analysing the distribution and locations in the array of the RRAM cells that are set to 1, an accurate measurement and classification of the gases is possible.

To achieve greater range and specificity of the gas sensors, the pull-up CNFETs are functionalized in several different manners. We use non-covalent functionalization to avoid damaging the pull-up CNFET. If the pull-up CNFET's effective resistance changes substantially or uncontrollably, the voltage divider will no longer be able to be biased at the proper voltage to ensure that the modulation of its resistance modulates the output voltage on either side of the mean RRAM V_{SET} . Extended Data Fig. 6 shows the CNFET inverter response after non-covalent functionalization of the pull-up CNFET and oxide-deposition over pull-down CNFET, showing no degradation of the inverter operation or output swing.

There is a rich literature describing how to form resistive gas sensors from CNFETs. For our CNFET gas sensors, we use several different methods: DNA functionalization³⁰, metal porphyrins²⁰ and self-assembled monolayers (SAMs). The DNA is a custom oligonucleotide, the porphyrin is 5,10,15,20-tetraphenyl-21 H,23H-porphine copper(II), and the SAM is octanethiol (all from Sigma Aldrich). First, a 100 μM solution of a custom oligonucleotide in de-ionized water, a 1 mM solution of the SAM in ethanol, and a 1 mM solution of the metal porphyrin in toluene are prepared. Next, the wafer with the exposed pull-up CNFETs are incubated in the solution for about 15 min, after which they are blown dry with N₂. To select which set of CNFET sensors are functionalized, drops of the solution can be placed over each sub-array separately, or standard lithographic techniques can be used to expose the targeted CNFETs, if the solvent does not attack the photoresist. The functionalization occurs after chip fabrication but before packaging. To characterize the sensors, arrays of 30 sensors with each functional group are loaded into a custom-built vacuum probe station and exposed to each gas in our test separately. An example I_D – V_{GS} curve from one of the sensors tested under different ambient gas conditions is shown in Fig. 3f. The change in the conductance of the CNT is not constant across the entire range of V_{GS} . For instance, on comparing the I_D – V_{GS} curves measured for nitrogen and white vinegar, when the CNFET is biased 'on' with the gate (that is, $V_{GS} = -3$ V), the CNFET under white vinegar shows higher resistance; however, when the CNFET is biased 'off' with the gate (that is, $V_{GS} = 0$ V), the CNFET under white vinegar shows lower resistance. Therefore, although a single V_{GS} bias point results in a 1D sensor, the ability to sweep along multiple values of V_{GS} results in a higher-dimensional sensor, increasing the amount of information obtained from each single-sense circuit. In Extended Data Fig. 7 we show a characterization of the response of the three types of functionalized CNFET sensor (metal porphyrin, DNA and SAM) to the seven gases and vapours tested.

Additional sensor characterization is shown in Extended Data Fig. 8. Extended Data Fig. 8a shows that the sensors are reversible, returning to their steady-state response after approximately 45 s. Extended Data Fig. 8b shows that the CNFET sensors are repeatable, with 30 repeated measurements taken from the same sensor showing little drift over time. Extended Data Fig. 8c shows how using the IIP to achieve highly aligned and pure semiconducting CNTs to realize CNFET logic naturally leads to improved CNFET sensor results.

To overcome CNFET variability and ensure matching drive strengths between the pull-up and pull-down CNFETs, the CNFET circuits use aligned active layouts^{10,25}. Aligned active layouts (Extended Data Fig. 9) constrain the active regions of the CNFETs (the areas of the CNFET channel with CNTs) within the standard cells to be aligned along the direction of the aligned CNT growth. As shown in Extended Data Fig. 9, although the two inverters (A and B) comprised a different number of CNTs (referred to as the CNT count), the CNT count between CNFETs within the same inverter are essentially identical. Therefore, although the transistors between inverters have greater variability, through circuit layout we engineer correlations within the circuits to ensure that the pull-up and pull-down CNFETs match closely.

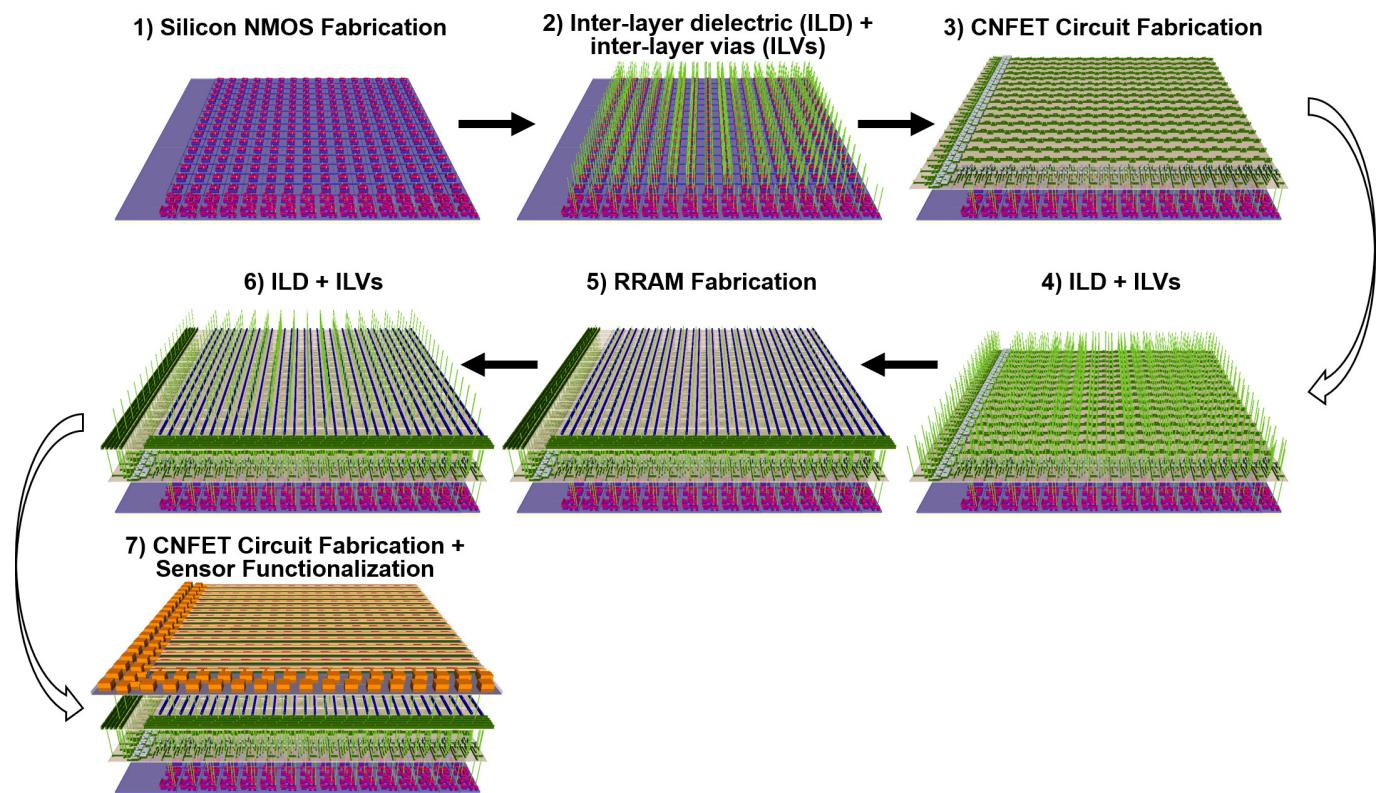
RRAM. The layers of memory in the monolithic 3D IC must be fabricated within the thermal budget of less than 400 °C, the same constraint as the logic. Many emerging memory technologies can be used within the monolithic 3D IC, including (but not limited to) RRAM, spin-transfer-torque RAM (STT-RAM) and conductive-bridge RAM (CBRAM). We use RRAM because it is non-volatile, simple to fabricate and promises high-capacity, non-volatile on-chip data storage, with better speed, energy efficiency and density than DRAM¹⁵. The 1 Mbit of RRAM is fabricated using a cross-point architecture, with 1T1R structures. Recent advances in RRAM technology have demonstrated fast-switching (about 300 ps), low-energy (about 0.1 pJ) operation, high endurance at an individual-cell level (10^{12} cycles) and high capacity (32 Gbit)¹¹. Details on the fabrication and classification of the RRAM are given in Methods section 'Fabrication' and shown in Fig. 3, respectively.

Silicon logic. The silicon logic is, along with the rest of the chip, fabricated at the Stanford Nanofabrication Facility. Consequently, our silicon is implemented at the same technology node as are the rest of the technologies of our chip (about 1 μm). Using an advanced silicon substrate from a commercial foundry would not affect the subsequent monolithic 3D fabrication of our nanosystem. In fact, a more complex starting silicon substrate would improve its performance and functionality. To simplify the processing involved, the silicon logic is a depletion-load NMOS logic, similarly to the CNFETs, which are implemented with a depletion-load PMOS logic. Whereas energy-efficient logic uses complementary PMOS and NMOS to form CMOS, the depletion-load logic involves a more simple fabrication and fewer lithography steps, which improves the yield in an academic cleanroom. Complementary silicon and CNFET logic could (and have) been implemented in academic and industrial cleanrooms. In our nanosystem, the silicon logic performs only three basic functions. First, it is used as the transistor in the 1T1R memory cells in the memory array. Second, it is used to fabricate the sense amplifiers on each bitline in the array. These sense amplifiers simply read the value of the selected (by the CNFET memory access subsystem) RRAM cell, and output a high or low voltage, which corresponds to the resistance of the RRAM cell. The sense amplifier is a simplified sense amplifier, using a transistor as a resistive load followed by cascaded inverters with high gain to output a final 1 or 0. Third, the silicon logic is used as the interface circuitry (implemented as multiplexers) between the sense amplifiers and the CNFET-based classification accelerator.

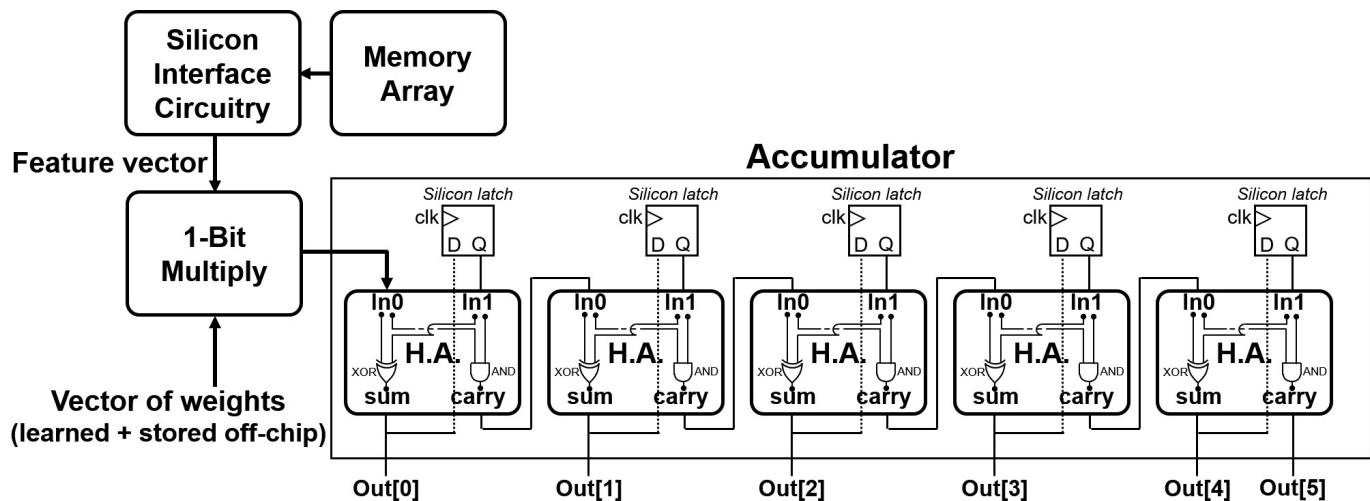
Test chamber. Our nanosystem was tested in a custom-built, environmentally controlled vacuum test chamber (Extended Data Fig. 10). The chip is wire-bonded to a package, which is then loaded into the test chamber. The chamber goes through 10 pump-purge cycles (of pumping to vacuum (less than 10^{-3} torr) and purging with N₂ to atmospheric pressure) before testing begins. Nitrogen and vapours from six different household liquids (lemon juice, white vinegar, rubbing alcohol, vodka, wine and beer) fill the test chamber through the bubbler system. To control the concentration, the bubbled gas is diluted with additional nitrogen gas to fill the chamber with the same concentration of each of the gases tested. Because most of the liquids are composed of many complex compounds, the exact concentration of each individual compound is unknown; this is in contrast to testing a single pure gas or compound, the concentration of which can be precisely controlled. Because the nanosystem presented here does not target a single gas compound, but can interact with its environment and subsequently learn from the resulting output from the millions of sensors, it is able to correctly classify these complex, real-world measurands.

Data availability. The data that support the findings of this study are shown in Figs 3 and 4 and Extended Data Figs 1–10, and are available from the corresponding author on reasonable request.

22. Shulaker, M. M. *et al.* Linear increases in carbon nanotube density through multiple transfer technique. *Nano Lett.* **11**, 1881–1886 (2011).
23. Patil, N., Lin, A., Myers, E. R., Wong, H. S. P. & Mitra, S. Integrated wafer-scale growth and transfer of directional carbon nanotubes and misaligned-carbon-nanotube-immune logic structures. In *Symposium on VLSI Technology* 205–206 (IEEE, 2008).
24. Shulaker, M. M. *et al.* High-performance carbon nanotube field-effect transistors. In *IEEE International Electron Devices Meeting (IEDM)* 33–36 (IEEE, 2014).
25. Zhang, J. *et al.* Carbon nanotube robust digital VLSI. *IEEE Trans. Computer-Aided Des.* **31**, 453–471 (2012).
26. Patil, N. *et al.* Scalable carbon nanotube computational and storage circuits immune to metallic and mispositioned carbon nanotubes. *IEEE Trans. Nanotechnol.* **10**, 744–750 (2011).
27. Patil, N. *et al.* VMR: VLSI-compatible metallic carbon nanotube removal for imperfection-immune cascaded multi-stage digital logic circuits using carbon nanotube FETs. *IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2009).
28. Shulaker, M. M. *et al.* Carbon nanotube circuit integration up to sub-20 nm channel lengths. *ACS Nano* **8**, 3434–3443 (2014).
29. Shulaker, M. M. *et al.* Efficient metallic carbon nanotube removal for highly-scaled technologies. *IEEE International Electron Devices Meeting (IEDM)* 32–34 (IEEE, 2015).
30. Staii, C., Johnson, A. T., Jr, Chen, M. & Gelperin, A. DNA-decorated carbon nanotubes for chemical sensing. *Nano Lett.* **5**, 1774–1778 (2005).



Extended Data Figure 1 | Fabrication flow for our nanosystem. See Methods section ‘Fabrication’ for details.



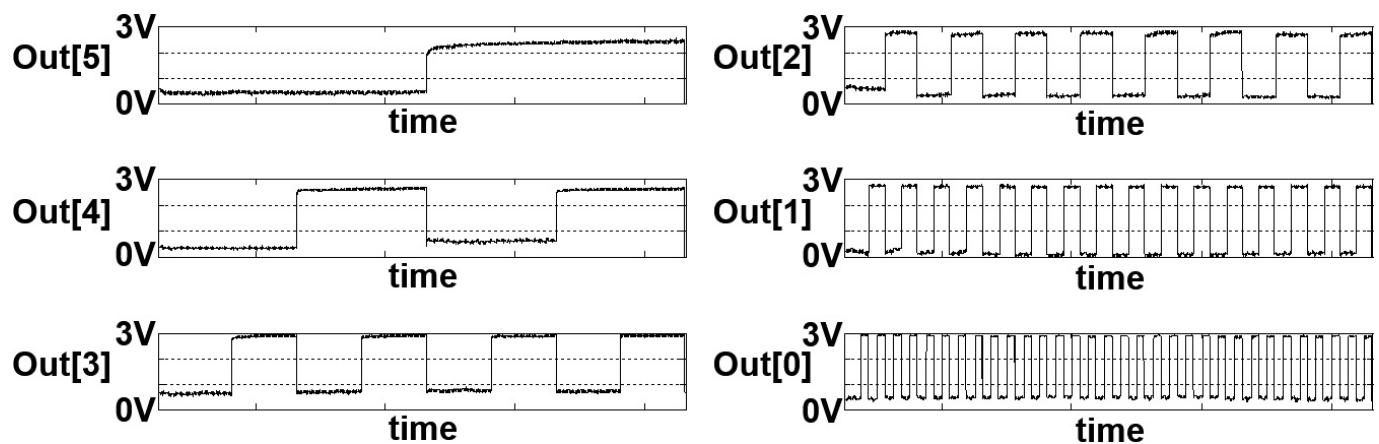
Extended Data Figure 2 | Schematic of the CNFET-based classification accelerator. The combinational logic is implemented with CNFETs (on the second layer), whereas the registers are implemented with silicon FETs (on the first layer). H.A., half-adder; clk, clock; D, latch input; Q, latch output.

Vector of Weights: Gas 1: [1 1 1 1 1 1 1 0 0 0 0 0 0 0]
 (predetermined to maximize classification accuracy)
 Gas 2: [0 0 0 0 0 0 0 1 1 1 1 1 1 1]

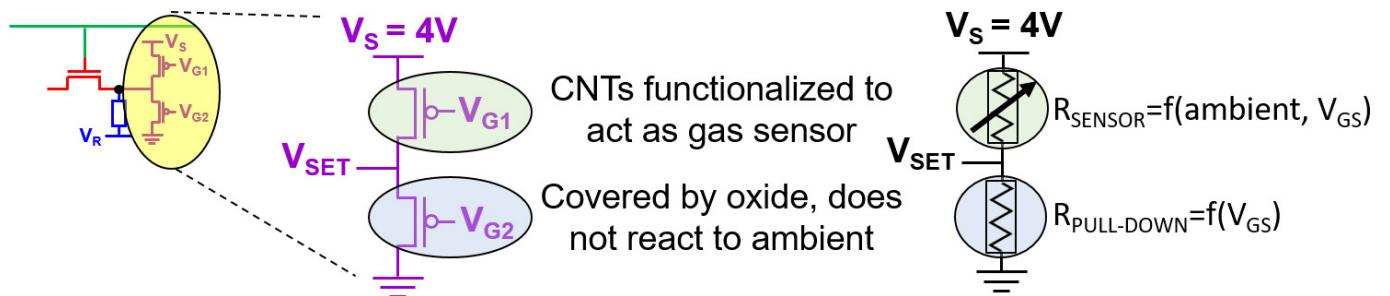
Threshold for classification: 5

	Memory Array Measured Under Gas 1	Memory Array Measured Under Gas 2																																
Memory Array (written by sensors):	<table border="1"> <tr><td>1</td><td>1</td><td>0</td><td>1</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>0</td><td>0</td></tr> </table>	1	1	0	1	1	1	0	0	1	1	1	0	1	1	0	0	<table border="1"> <tr><td>0</td><td>0</td><td>0</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>1</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>1</td><td>1</td></tr> </table>	0	0	0	1	0	0	1	1	0	0	1	0	1	0	1	1
1	1	0	1																															
1	1	0	0																															
1	1	1	0																															
1	1	0	0																															
0	0	0	1																															
0	0	1	1																															
0	0	1	0																															
1	0	1	1																															
Corresponding Feature Vector:	[1 1 1 1 1 1 1 0 0 1 0 1 0 0 0]	[0 0 0 1 0 0 0 0 0 1 1 1 1 0 1]																																
Test Gas 1? F.V. X Gas 1 Vector of Weights:	[1 1 1 1 1 1 1 1 0 0 0 0 0 0 0]	[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]																																
Accumulation:	8	1																																
Test Gas 2? F.V. X Gas 2 Vector of Weights:	[0 0 0 0 0 0 0 0 0 0 1 0 1 0 0]	[0 0 0 0 0 0 0 0 0 1 1 1 1 0 1]																																
Accumulation:	2	6																																
Classification:	Positive ID for Gas 1	Positive ID for Gas 2																																

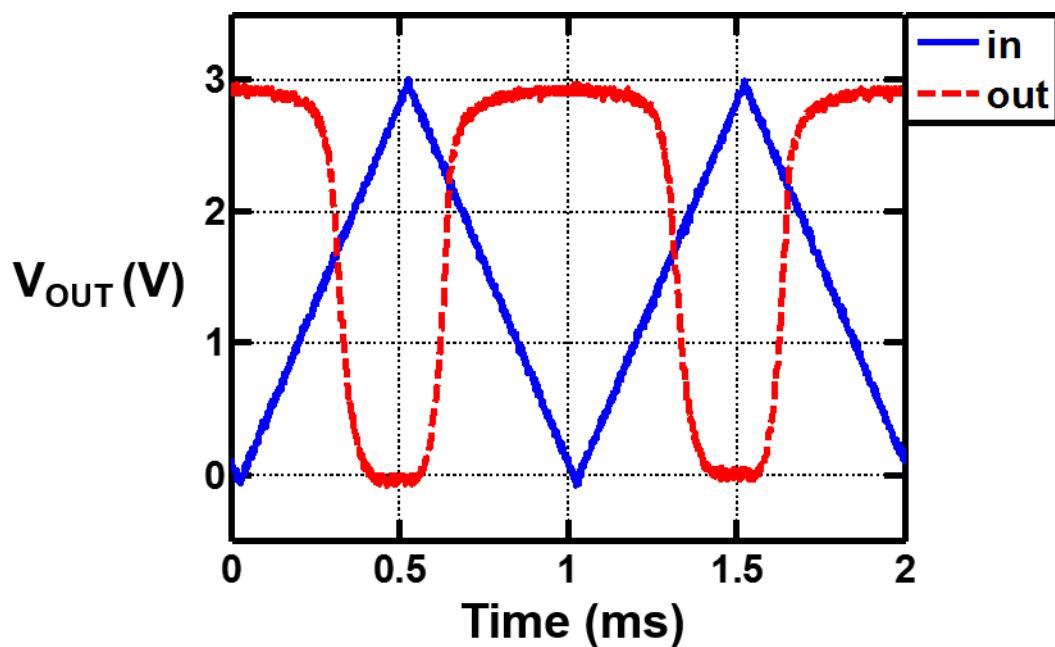
Extended Data Figure 3 | Small-scale example of how the CNFET-based classification accelerator performs classification. F.V., feature vector.



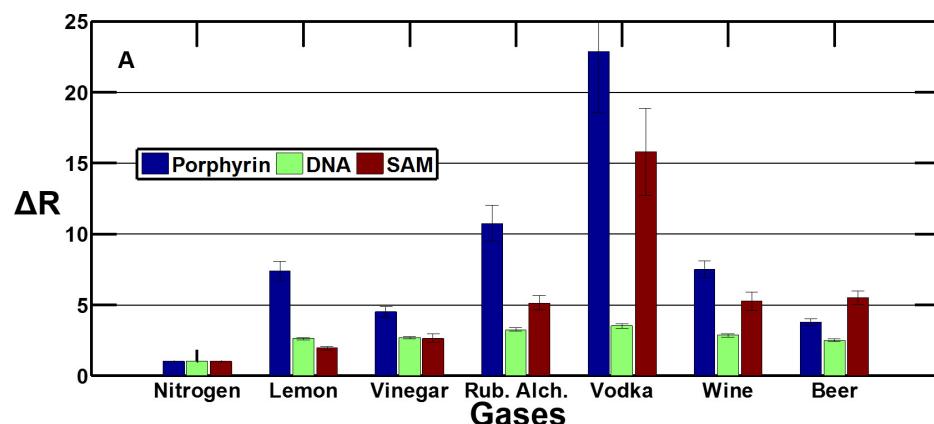
Extended Data Figure 4 | Measured waveform of the CNFET-based classification, testing all possible combinations.



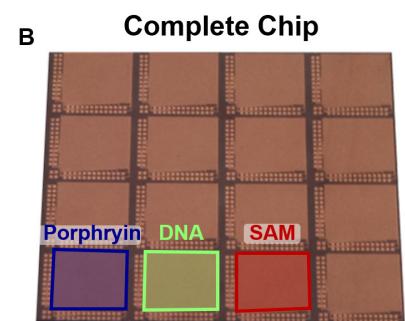
Extended Data Figure 5 | Implementation of the CNFET inverter operating as a gas sensor. The resistance of the sensor R_{sensor} depends on the ambient air and V_{GS} ($R_{\text{sensor}} = f(\text{ambient}, V_{GS})$), whereas the resistance of the pull-down CNFET $R_{\text{pull-down}}$ depends only of V_{GS} ($R_{\text{pull-down}} = f(V_{GS})$).



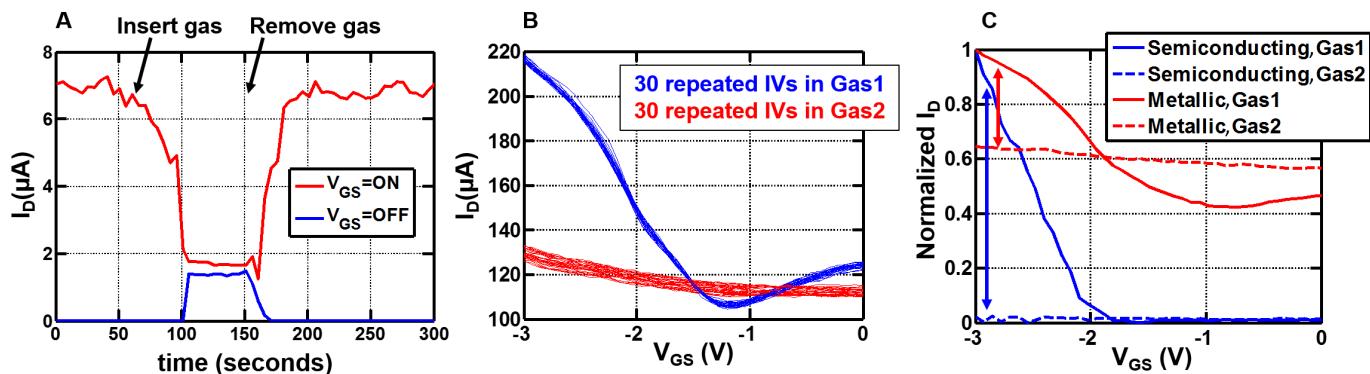
Extended Data Figure 6 | Input ('in') and output voltage V_{out} ('out') of the CNFET inverter (gas and vapour sensor) after non-covalent functionalization of the pull-up CNFET and oxide deposition over the pull-down CNFET.



Extended Data Figure 7 | CNFET gas sensors. **a**, Characterization of the CNFET gas sensors. Sample size is 90 (30 of each the three types of CNFET gas sensor). ΔR is defined as the resistance measured after exposure to the given gas divided by the baseline resistance in vacuum, with the resistance is both cases measured at $V_{GS} = -3$ V and $V_{DS} = -2$ V (error bars show

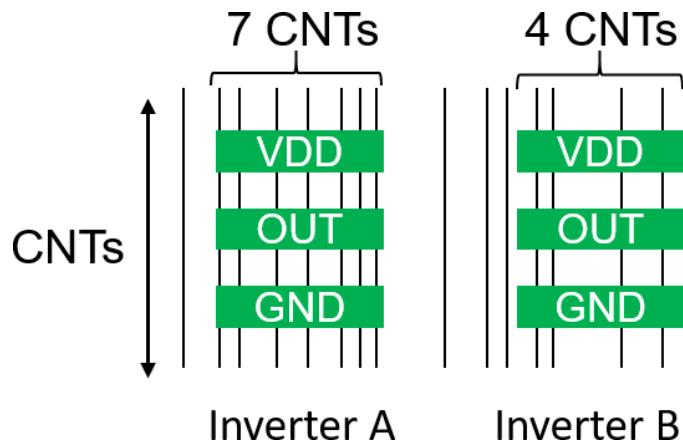


95% confidence intervals). **b**, Example layout showing how sub-arrays of the complete chip can be functionalized. By measuring the percentage of RRAM cells that are set to 1 in each sub-array during the sensing phase of operation, an average value for the CNFET sensing circuit can be calculated, which corresponds to R_{sensor} .

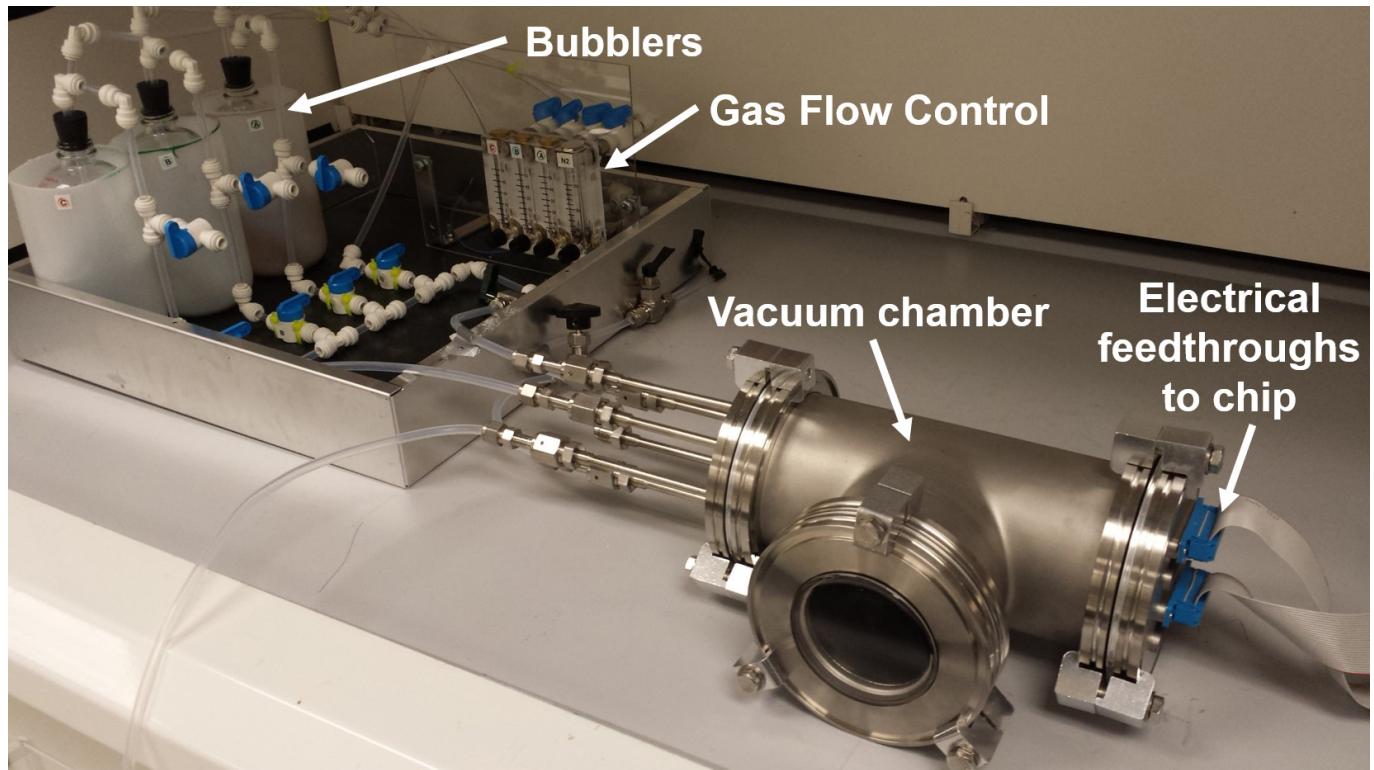

Extended Data Figure 8 | Characterization of the CNFET gas sensors.

a, Sensor response is reversible, responding and returning to steady-state within approximately 45 s. **b**, Sensor response is robust: 30 repeated measurements of the current–voltage curve ('IVs') from the same CNFET gas sensor yield similar responses. **c**, The techniques that we used to

realize VLSI-compatible CNFET logic simultaneously improve CNFET sensor performance. The CNFET with purely semiconducting CNTs ('Semicconducting') has a much larger sensitivity and change in its response than a CNFET with metallic CNTs ('Metallic'), as indicated by the arrows.



Extended Data Figure 9 | Aligned active layouts are used to overcome variability in CNTs. VDD, supply voltage; GND, ground; OUT, output node.



Extended Data Figure 10 | Test chamber for our nanosystem.