ULTRARAM™: a low-energy, high-endurance, compound-semiconductor memory on silicon

Peter Hodgson
Lancaster University https://orcid.org/0000-0001-5813-8772
Dominic Lane
Lancaster University
Peter Carrington
Department of Engineering, Lancaster University, Lancaster, LA1 4YW, United Kingdom
https://orcid.org/0000-0003-2107-5602
Evangelia Delli
Lancaster University
Richard Beanland
University of Warwick https://orcid.org/0000-0003-1749-4134
Manus Hayne (m.hayne@lancaster.ac.uk)
Lancaster University https://orcid.org/0000-0001-5224-9156

Article

Keywords: Non-volatile Memory, Triple-barrier Resonant Tunnelling Heterostructure, Molecular Beam Epitaxy, Switching Energy

Posted Date: July 14th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-690726/v1

ULTRA**RAM™: a low-energy, high-endurance, compound-semiconductor memory on silicon**

Peter D. Hodgson,[1] Dominic Lane,[1] Peter J. Carrington,[2] Evangelia Delli,[2] Richard Beanland[3] and Manus Hayne[1*]

[1] Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom

[2] Department of Engineering, Lancaster University, Lancaster LA1 4YW, United Kingdom

[3] Department of Physics, University of Warwick, Coventry, CV4 7AL, United Kingdom

*Corresponding author: Manus Hayne m.hayne@lancaster.ac.uk

ULTRA**RAM™ is a non-volatile memory with the potential to achieve fast, ultra-low-energy electron storage in a floating gate accessed through a triple-barrier resonant tunnelling heterostructure. Here we report the implementation of** ULTRA**RAM™ on a Si substrate; a vital step towards cost-effective mass production. Sample growth was carried out using molecular beam epitaxy, by first depositing an AlSb nucleation layer to seed the growth of a GaSb buffer layer, followed by the III-V memory epilayers. Fabricated single-cell memories show clear 0/1 logic-state contrast after ≤10-ms duration program/erase pulses of ~2.5 V, a remarkably fast switching speed for 10- and 20-µm devices. Furthermore, the combination of low voltage and small device capacitance per unit area results in a switching energy that is orders of magnitude lower than dynamic random access memory and flash, for a given cell size. Extended testing of the devices revealed retention in excess of 1000 years and degradation-free endurance of over 10^7 program/erase cycles, exceeding very recent results for similar devices on GaAs substrates.**

A memory that is fast and non-volatile, with high endurance and low-energy logic-state switching, *i.e*. a so-called universal memory, has long been dismissed as unachievable due to the apparently contradictory physical properties such a device would require[1]. Conventionally, a fast, high-endurance memory seemingly necessitates a frail logic state that is easily lost, even requiring constant refreshing; for example, dynamic random access memory (DRAM). In contrast, a robust, non-volatile logic state ostensibly requires large amounts of energy to switch, which (gradually) damages the memory structure, reducing endurance; for example, flash. The detrimental aspects of these two conventional memories have stimulated research into a range of emerging memory technologies, such as resistive RAM[2], magnetoresitive RAM[3] and phase-change memory[4]. Significant progress has been made, with emerging memory products in small- or large-scale commercial production, but, as with conventional memories, the trade-off between logic state stability and switching energy remains. ULTRA**RAM**™ breaks this paradigm via the exploitation of InAs quantum wells and AlSb barriers to create a triple-barrier resonant-tunnelling (TBRT) structure. The 2.1-eV conduction band offset of AlSb with respect to the InAs that forms the floating gate (FG) and channel, provides a barrier to the passage of electrons that is comparable to the SiO_2 dielectric used in flash. However, inclusion of two InAs quantum wells (of different thicknesses) within the TBRT structure, as shown in Fig. 1, allows it to become transparent to electrons when a low voltage (~2.5 V) is applied, due to resonant tunnelling. By using the TBRT heterostructure as the barrier between FG and channel, rather than the usual monolithic material, a charge-based memory with extraordinary properties can be achieved.

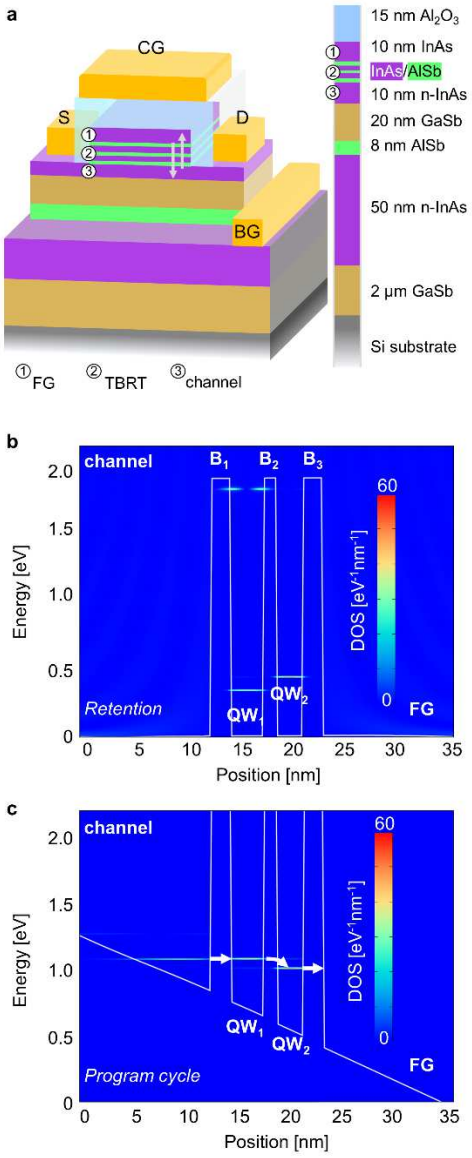
Incorporation of ULTRA**RAM**™ onto Si substrates is a vital step towards realising low-cost, high-volume production. Si substrates offer several advantages over III-Vs, including mechanical strength and large wafer sizes, thereby allowing fabrication of more devices in parallel and reducing production cost. Moreover, Si is the preferred material for digital logic and has a highly mature fabrication route. Integration of the complementary technologies of III-Vs and Si has historically been carried out by wafer bonding techniques, but such approaches have drawbacks such as added complexity of fabrication and differences of wafer size, often resulting in inefficient use of substrate material[5]. More

recently, heterogeneous epitaxy has offered a promising route towards the efficient co-integration of III-Vs and Si. However, direct epitaxial deposition is challenging due to the polar (III-Vs) to non-polar (Si) interface, lattice mismatch and differences in thermal expansion coefficients[6-9]. In order to prevent defect formation and realise practical, high-performance, integrated, III-V devices on Si substrates, careful consideration must be given to mitigate these challenges.

Here we report on the epitaxial incorporation of ULTRA**RAM**™ heterostructures onto Si substrates, their fabrication into devices and testing. Results show non-volatile, high-endurance operation that exceeds recently reported results on GaAs substrates[10].

Memory concept

ULTRA**RAM**™ is a charge-based memory where the logic state is determined by the presence or absence of electrons in a FG. As can be seen in Fig. 1, the FG is electrically isolated from the control gate (CG) by Al_2O_3 dielectric, and from the underlying channel by the InAs/AlSb TBRT heterostructure. The presence of electrons in the FG (defining a logic 0 state) depletes carriers in the underlying n-type InAs channel, reducing its conductance. Thus, the charge state of the FG and, therefore, the logic state of the memory is read non-destructively by measuring the current through the channel when a voltage is applied between the source (S) and drain (D) contacts. The final component of the memory is the InAs back-gate (BG), which allows voltages to be applied vertically across the gate stack for various operations.



68

69 **Fig. 1 |** ULTRA**RAM™ device concept. a**, Schematic cross-section of an ULTRA**RAM™**

70 device with corresponding material layers. The floating gate (FG), triple-barrier

71 resonant-tunnelling structure (TBRT) and readout channel are highlighted. Arrows

72 indicate the direction of electron flow during program/erase operations. **b,** and **c,** Non-

73 equilibrium Green's functions (NEGF) calculations of density of states alongside

74 conduction band diagrams for no applied bias (*i.e*. retention) and program-cycle bias

75 respectively. B_1, B_2 and B_3 are the AlSb barrier layers. QW_1 and QW_2 are the InAs

76 quantum wells in the TBRT.

The novel component of this memory is the TBRT structure[11], which, unlike single layer barriers, can be switched from a highly electrically-resistive state to a highly conductive state by the application of just ±2.5 V. This is achieved by careful design of the thicknesses of the AlSb barriers and InAs quantum well (QW) layers[12]. When the memory is in the retention state (Fig. 1**b**), *i.e*. when no voltage is applied to the device, the electron ground states in the TBRT QWs are misaligned with each other and are energetically well above the 300-K electron populations of the InAs FG and channel layers. Indeed, the non-volatility is strengthened by the QW ground states residing at an unusually high energy for a resonant-tunnelling structure. This is due to a combination of the ultrathin QWs and the extraordinarily low electron effective mass in InAs[13]. In this state, the TBRT provides a large barrier that prevents electron transfer into or out of the FG. However, the application of a suitable bias across the device tilts the conduction band such that the TBRT QW ground states align with occupied electron states in the channel (during the program operation, Fig. 1**c**) or the FG (during the erase operation). This allows electrons to move rapidly across the TBRT region in the intended direction by the inherently-fast quantum-mechanical process of resonant tunnelling. Due to the low voltages required and the extremely low capacitance per unit area of the device, ultra-low logical state switching energies of 10^{-17} J are predicted for 20-nm feature size ULTRA**RAM**™ memories[12,14], which is two and three orders of magnitude lower than DRAM and flash respectively[12,15]. ULTRA**RAM**™ prototype devices grown on GaAs substrates have previously exhibited non-volatile retention of 10^5 s and an endurance of 10^6 program-erase cycles.[10] Thus, the demonstration of devices grown on Si substrates with similar or improved performance would be a major step towards commercialisation.

Epitaxy of III-Vs on Si and sample characterisation

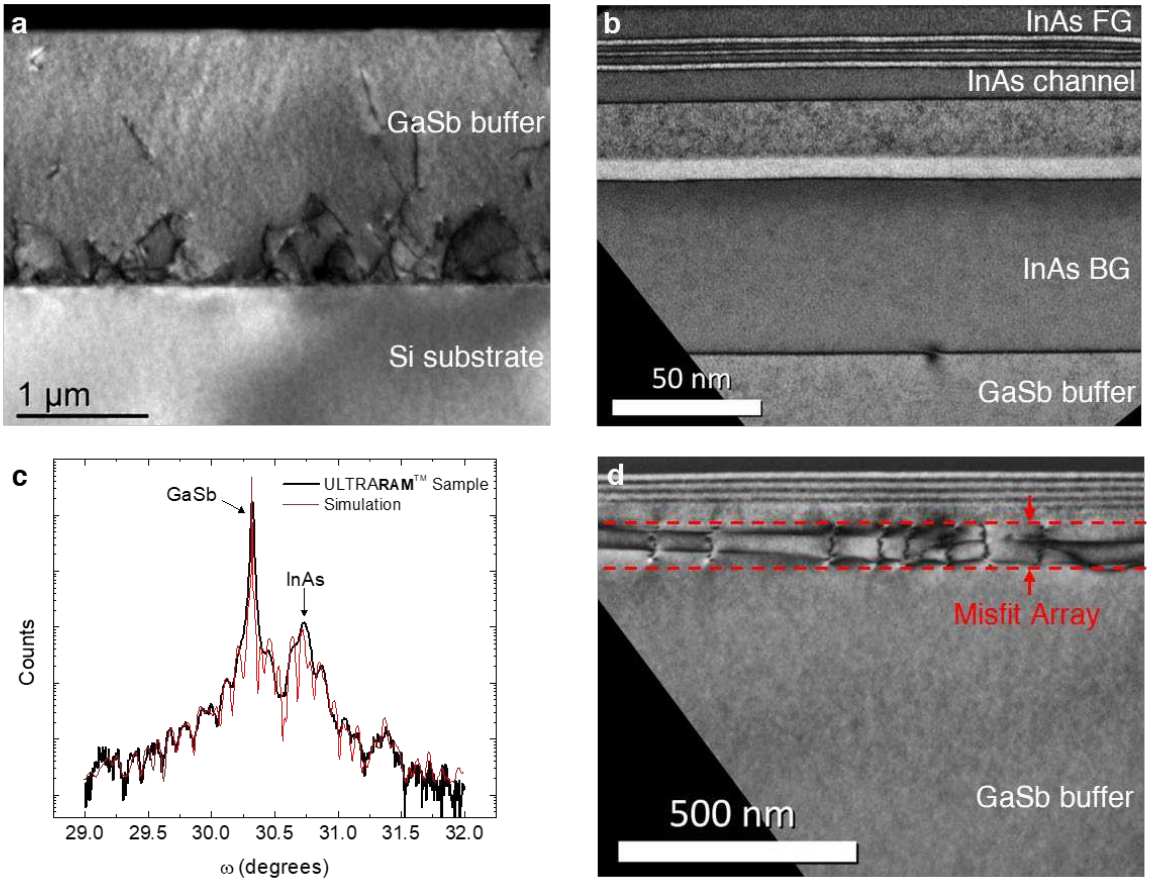
Sample growth on 3” n-type Si substrates was carried out by molecular beam epitaxy (MBE) on a Veeco GENxplor system. The substrates have a 4˚ offcut towards the [0-11] crystal direction, creating diatomic steps on the non-polar Si surface, which inhibits the formation of anti-phase domains (APDs)

during growth of polar III-V materials[5]. Prior to the growth of a complete structure with the ULTRA**RAM**™ memory layers, a GaSb buffer growth procedure was developed to provide a high-quality III-V surface. To achieve this, *in-situ* thermal desorption of the Si wafer's native oxide was followed by deposition a 17-monolayer AlSb nucleation layer, which forms into three-dimensional islands[16]. These islands reduce the diffusion length of Ga atoms during the initial growth of the subsequent 2-µm GaSb buffer, helping to promote two-dimensional epitaxy and preventing the formation of planar twinning defects[8,9,17]. The large lattice mismatch between the Si substrate and the GaSb buffer of 12.3% is relieved by a periodic array of 90˚ interface misfit dislocations propagating laterally along the Si-III/V interface[18]. A two-temperature-step GaSb growth method[19] was used to reduce the density of vertically propagating threading dislocations (see Methods section for more details). Using this approach, a GaSb/Si buffer layer was obtained with a surface defect density of $(2.5 \pm 0.1)\times10^{8}$ cm^{-2}, as measured by electron channelling contrast imaging (ECCI), and a root-mean-square surface roughness of 1.9 ± 0.2 nm, as measured by atomic force microscopy (AFM) by taking the average of three 100 μm^2 scans. High-resolution X-ray diffraction (XRD) measurements further confirmed the high quality of the buffer layer. A rocking curve measurement of the GaSb (004) plane revealed an extremely narrow peak with a full-width at half maximum of 172 arcsec. In addition, XRD pole-figure measurements of the GaSb (111) plane indicate that the buffer is either free from, or has an extremely low density of, planar twins[17]. Cross-sectional transmission electron microscope (TEM) images, such as Fig. 2**a**, show that the buffer layer is APD free, and that the majority of threading dislocations are confined to the first 500 nm of GaSb.

These results demonstrate that the GaSb buffer layer on Si is of high quality, and suitable for deposition of the ULTRA**RAM**™ memory layers. The first epilayers grown on such a buffer are the 50-nm InAs n-type BG and 8-nm AlSb barrier, which separates the BG from the active parts of the device, as shown in Fig. 1**a** and Fig. 2**b**. Next, a 20-nm GaSb spacer was deposited, followed by a 10-nm InAs n-type layer (channel) and the InAs/AlSb TBRT. The nominal TBRT layer thicknesses are 1.8 nm of AlSb,

3.0 nm of InAs, 1.2 nm of AlSb, 2.4 nm of InAs and 1.8 nm of AlSb. Finally, a 10-nm InAs FG completes the structure. More information on the MBE growth of the sample is given in the Methods section.

The ECCI-measured surface defect density in the ULTRA**RAM**™ sample was found to be $(2.1 \pm 0.1)\times10^{8}$ cm^{-2}, which is slightly lower than that observed for the buffer. This is not surprising, since interfaces with misfit strain, such as the GaSb/InAs/AlSb layers used here, are known to encourage the recombination and termination of threading dislocations and are routinely used as dislocation filters[20-22]. A high-resolution XRD ω-2θ scan of the memory sample, which displays clear satellite peaks from the memory epilayers, is shown in Fig. 2**c**. Optimal fitting to the data is achieved assuming 15% misfit strain relaxation in the InAs BG. Further investigation by TEM revealed the presence of a misfit dislocation array at the interface between the GaSb buffer and InAs BG layer (Fig. 2**d**), which are responsible for this relaxation. These misfit dislocations do not propagate into the GaSb layer above, and act as a threading dislocation filter. Remnant threading dislocations in the BG are not expected to significantly influence device performance as this layer is only used as a ground contact when biasing the CG.



141

142 **Fig. 2 | III-V on Si Material characterisation. a**, Dark-field **g** = 220 transmission electron
143 microscope (TEM) image of a GaSb/Si buffer layer. **b,** Dark-field **g** = 002 TEM image of
144 the ULTRA**RAM**™ sample. Only the memory layers and the top of the GaSb buffer are
145 visible in this image. A single 60° misfit dislocation is visible in the InAs/GaSb buffer
146 interface. **c,** Experimental and simulated ω-2θ high-resolution X-ray diffraction scan of
147 the ULTRA**RAM**™ sample. The Si substrate's diffraction peak occurs at a larger angle and
148 is not shown in this scan. **d,** Dark-field **g** = 220 TEM image of the upper part of the
149 structure. The specimen tilt of ~20°allows the misfit dislocation array in the InAs back
150 gate (BG) / buffer layer interface to be seen. The lower and upper dashed lines indicate
151 the intersection of this interface with the specimen surfaces.

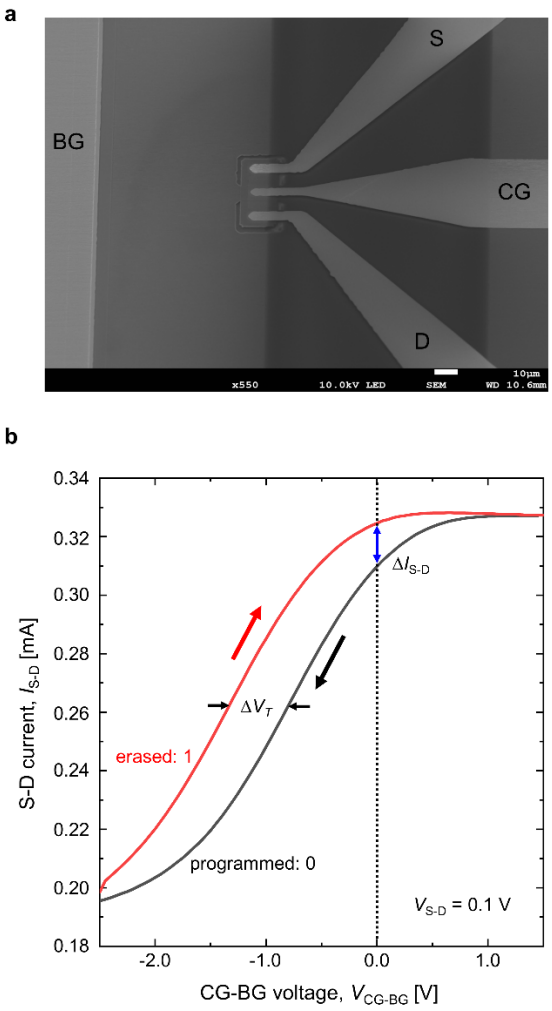
152

153

Basic memory operations

Single bit ULTRA**RAM**™ memory cells on Si substrates were fabricated at gate lengths of 10- and 20-µm. Fig. 1**a** shows a schematic cross-section of a fabricated memory cell. A scanning electron microscope image of a fabricated 10-µm gate-length device is shown in Fig. 3**a** with device terminals labelled. Here, we define a charged FG as logic 0, and the absence of charge as logic 1. Program and erase cycles, to charge and discharge the FG respectively, use voltage pulses of ≤ ±2.55 V on the CG.

The memory's readout logic is demonstrated in Fig. 3**b**. A 4-terminal measurement is used, where the CG to BG voltage ($V_{CG\text{-}BG}$) is varied whilst measuring the current from S to D under a constant 0.1 V S-D bias ($V_{S\text{-}D}$). A negative $V_{CG\text{-}BG}$ bias produces carrier depletion of the channel layer, similar to that observed in an n-type metal-oxide field-effect transistor, resulting in an observable increase in resistance with increasing negative bias. When the device is in a programmed state (0), the negative charge on the FG enhances the CG bias such that depletion of the channel occurs at a lower $V_{CG\text{-}BG}$. The difference between the two states allows a 0.5 V threshold-shift (ΔV_T) to be observed, centred at a $V_{CG\text{-}BG}$ of approximately -1.25 V, as shown in Fig. 3**b**. By increasing $V_{CG\text{-}BG}$ to ±2.5 V the memory can be programmed/erased, creating the hysteresis seen in the data. Note that the joining of the states at >0.8 V positive CG bias does not indicate program and erase operations: the memory window cannot be observed under high positive gate bias due to saturation of the carrier population in the channel. These results demonstrate that the FG logic state can be read non-destructively. Furthermore, the program/erase voltage is around 10 times lower than that of flash memories[15,23], corresponding to a switching energy per unit area that is lower than DRAM and flash by factors of 100 and 1000 respectively[12,15].



175

176 **Fig. 3 |Memory device and basic operation. a**, Scanning electron microscope image of a
177 fabricated ULTRA**RAM™** device of 10-µm gate length. **b**, S-D current measurement at
178 constant S-D voltage, V_{S-D} = 0.1 V, during a voltage sweep across the CG-BG (V_{CG-BG}) from
179 +2.5 V to -2.5 V (black line) and -2.5 V to +2.5 V (red line), representing the program and
180 erase cycles respectively. The hysteresis indicates a threshold voltage shift for the
181 memory (ΔV_T) of 0.5 V.

182 In this work, an n-type InAs, normally-on, channel layer was chosen for simplicity, as the primary
183 objective was to demonstrate single-cell ULTRA**RAM™** operation on Si substrates. Whilst this channel
184 design is sufficient for that purpose, it results in a low current contrast between logic states, labelled
185 as ΔI_{S-D} in Fig. 3**b**. It is important to emphasise that this is a consequence of the simplicity of the channel
186 and not a result of insufficient contrast in the programmed and erased state of the FG, as proven by

the 0.5 V threshold window in Fig. 3**b**. Indeed, the implementation of a normally-off channel should allow a similar readout mechanism to flash, in which the application of a reference gate voltage only induces conduction in the channel in the absence of charge in the FG (erased state), thereby greatly increasing 0/1 current contrast for the same ΔV_T memory window[12,24]. Such an improvement of the 0/1 contrast through careful modification of the channel design will allow memory arrays to be built with a novel high-density RAM architecture[10,12]. Furthermore, the inclusion of the gate dielectric provides flexibility to significantly increase ΔV_T by increasing the gate dielectric capacitance: this can be achieved by thinning the layer or by using higher-k dielectric materials.

Retention, endurance and speed testing

Retention testing of the memory state was carried out at room temperature on a 20-µm-gate-length device by repeatedly measuring $I_{S\text{-}D}$ under a $V_{S\text{-}D}$ bias of 0.2 V, but in the absence of $V_{CG\text{-}BG}$ bias: a simple readout scheme made possible by the n-doped InAs, normally-on channel. Memory retention was confirmed for >24 hours for both program and erase states using $>10^6$ readout operations, limited only by the length of the experiment (Fig. 4**a**). There is an initial decay in the $I_{S\text{-}D}$ contrast between the two logic states (*i.e.* the memory window, $\Delta I_{S\text{-}D}$), before it plateaus at around 22 µA after roughly 14 hours. To investigate the memory's retention further, $\Delta I_{S\text{-}D}$ was plotted on a log-scale and different fittings were made to the data, as shown in Fig. 4**b**. By extrapolating these fitted lines to the point at which $\Delta I_{S\text{-}D} = 0$, *i.e*. when the memory window closes, the retention time of the memory can be estimated. The plateauing of the memory window after 14 hours makes determination of the retention time difficult, as shown by the dashed line in Fig. 4**b**, which extends to infinity. Therefore, a second fitting is shown (solid line in Fig 4**b**), which follows the decay of the memory states, prior to the final 10 hours of data. This provides an extremely conservative lower limit for the memory's retention of at least 10^7 hours, which is more than 1000 years.

A similar, but more prominent, initial state decay was previously observed in the first prototype ULTRA**RAM**™ devices on GaAs substrates[14], and was subsequently eliminated with improved material

quality.[10] The return of the partial state decay for these first devices on Si substrates corroborates this correlation with material quality, where the most likely candidate is charge trapping at defect sites on the heterojunction interfaces. Thus, we predict a similar elimination of the state decay on Si substrates with continued development of the material epitaxy. Nevertheless, the 0/1 contrast throughout the 24-hour test and the extremely long predicted retention times clearly demonstrate the non-volatility of the logic states.

Endurance testing was carried out at room temperature by program-read-erase-read cycling on a second 20-µm device using 5-ms-duration $V_{CG\text{-}D}$ pulses of +2.1 V and -2.55 V respectively, with $I_{S\text{-}D}$ readout measurements at $V_{S\text{-}D}$ = 0.2 V in the absence of gate bias. The memory cell successfully underwent 10^6 program-read-erase-read cycles with a stable memory window and without degradation as shown in Fig. 4**c**. Moreover, the cell had zero cycle failures and <50 partial switches during the 10^6 cycles. Of particular note is the reproducible nature of the I_{SD} values for programmed and erased states, with neither the pronounced drift observed in the first prototypes[14], nor the large fluctuations in I_{SD} seen in more recent devices[10]. The latter was attributed to non-optimal etching of the channel during device fabrication, and has been almost completely eliminated due to an improved process (see Methods section for details). In Ref. 14 the drift in I_{SD} was attributed to an asymmetry in the program/erase process. This was later shown by quantum transport simulations to be the result of the asymmetry of the TBRT structure, allowing resonant tunnelling to occur at a lower voltage for the program cycle than for the erase cycle[12]. As a result, the use of symmetric voltages, such as ±2.5 V, results in over-programming, such that more electrons are added in a program cycle than are removed in by an erase, producing a current drift with each cycle. Fortunately, the state will quickly stabilise with repeated cycling once the correct voltages are identified. The remaining small drift of the I_{SD} window observed in Fig. 4**c** over many thousands of cycles is likely to become inconsequential with the implementation of the normally-off channel readout scheme.

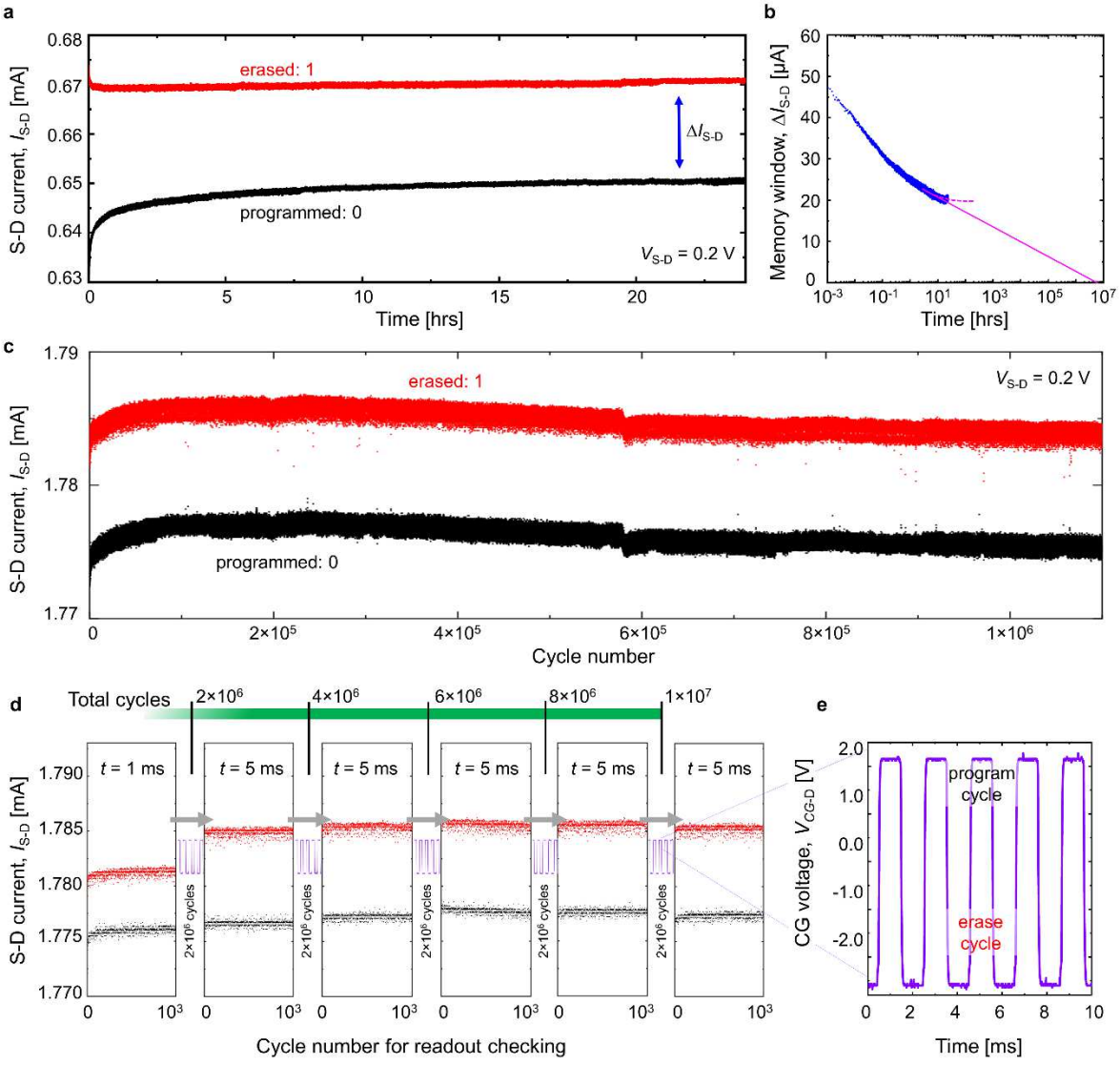


Fig. 4 | Retention and endurance characteristics. a, Retention data for a 20-µm-gate-length cell. Program and erase cycles consisted of 10-ms duration pulses at +2.5 V and -2.5 V respectively. Readout is performed at an S-D bias of 0.2 V, and in the absence of gate bias. **b**, S-D current difference (ΔI_{S-D}) for the >24 hour retention plotted on a log scale. The continuous magenta line indicates the least generous interpretation of the decay data, whereas the dashed magenta line follows the stabilisation of the memory window. **c**, Endurance data for continuous program-read-erase-read cycling (5-ms pulses) on a second 20-µm-gate-length cell demonstrating a clear 0/1 contrast exceeding 10^6 cycles. **d**, Extended endurance to >10^7 cycles using continuous pulse trains

of 2×10^6 program-erase cycles repeated five times separated by 1000 program-read-erase-read cycles to confirm memory operation persists. The program and erase pulse duration, *t*, was 1 ms in the absence of read, and 5 ms with the read, except for the first set of 1000 program-read-erase-read cycles, where it was also 1 ms. The data is presented chronologically from left to right with the total cycles counted by the green bar. **e**, Oscilloscope trace showing the applied gate bias for a section of the pulse train during *t* = 1 ms program-erase endurance cycling.

The endurance testing on this device was then extended by a further order of magnitude using a slightly modified methodology to increase the cycling speed by substantially reducing the number of read operations; these take significantly longer than program and erase due to testing equipment limitations. Firstly, an initial set of 1000 program-read-erase-read cycles was applied to the device to confirm correct operation, followed by a pulse train of 2×10^6 program-erase cycles without any read, in both cases with program/erase pulse durations of 1 ms. The 2×10^6, 1-ms program-erase pulses were then repeated multiple times, an example section of which is shown in Fig. 4**e**. In each case, these were followed by 1000 program-read-erase-read cycles, but with program/erase pulses of 5 ms duration to increase the width of the $\Delta I_{S\text{-}D}$ memory window, making it easier to detect any potential degradation. The process was repeated five times, resulting in a little over 10^7 program/erase cycles applied to the device. As can be clearly seen in Fig. 4**d**, there is no degradation of the $\Delta I_{S\text{-}D}$ window throughout these tests, meaning that the endurance is at least 10^7. This therefore represents an endurance capability that is at least two to three orders of magnitude better than flash[15,23].

In all of the above tests, the program and erase states were set using between 1- and 10-ms voltage pulses, two times longer than the switching times used in our recent report of ULTRA**RAM**™ on GaAs substrates[10]. In both cases, the devices operate at a remarkably high speed for their large (20 μm) feature size. Assuming ideal capacitive scaling[25] down to state-of-the-art feature sizes, the switching performance would be faster than DRAM, although testing on smaller feature size devices is required

to confirm this. The small loss in performance when switching from GaAs to Si substrate may be a result of charge trapping at defect sites contributing to the RC-delay and program/erase voltage screening. However, note that the switching speed for the present devices on Si is 1000 times faster than the first devices on GaAs substrates, suggesting only a modest change in material quality on Si compared to the more mature GaAs substrate growth method[10,14].

Finally, regarding reproducibility, the majority of devices with 10- and 20-µm gate length on the chip exhibited memory characteristics. However, there was significant device-to-device variation in absolute channel current, memory window size and state decay during retention. The difference in S-D current ($I_{S\text{-}D}$) between the device data presented in Fig. 4**a** and **c** shows this clearly, and is the most extreme example to be found on the chip. It is believed to be caused by S-D to BG current leakage due to etch pitting of defects during fabrication, which creates an alternative conducting path from S to D. Whilst the TEM image in Fig. 2**d** shows the movement of threading defects to produce a misfit dislocation array at the bottom of the BG, threading defects are not eliminated and etch pitting suggests that a small number may continue into the overlying channel. Work is ongoing to reduce this device-to-device variation with further development of material epitaxy and fabrication processes.

Conclusions

We have demonstrated ULTRA**RAM**™ compound-semiconductor, floating-gate memory cells on Si substrates. A high-quality GaSb buffer on Si was developed, onto which the memory layers were successfully implemented. TEM and XRD scans revealed the presence of threading defects in the InAs BG, which likely slightly impinge on device performance due to varying levels of S-D to BG current leakage. However, the remaining memory epilayers exhibit excellent quality with abrupt material interfaces and a low surface defect density of $(2.1 \pm 0.1)\times10^8$ cm^{-2}. Testing of the fabricated single cell memory devices shows strong potential, with devices demonstrating a clear memory window during ≤10-ms program/erase operations, which is remarkably fast for 10- and 20-µm gate-length devices. The ~2.5 V program/erase voltage and low device-areal-capacitance results in a switching energy per

unit area that is 100 and 1000 times lower than DRAM and flash respectively. Extrapolated retention times in excess of 1000 years and degradation-free endurance tests of over 10^7 program-erase cycles prove that these memories are non-volatile and have high endurance. Further work to improve epitaxial quality, fine-tune the fabrication process, implement a normally-off channel design and scale the devices is ongoing.

Methods

Non-equilibrium Green's function calculations. The quantum transport of electrons through the resonant tunnelling structure was calculated with the nextnano.MSB software package, which uses the non-equilibrium Green's function framework. This method also generalizes the so-called Büttiker probe model and takes into account all relevant individual scattering mechanisms to accurately model the tunnelling structure. Further details of the method can be found elsewhere[12,26,27].

Molecular beam epitaxy. Samples were grown using a solid source Veeco GENxplor MBE system equipped with As and Sb valved cracker cells. Epilayers were deposited on 3" Si n-type (100) substrates with a 4˚ offcut towards [0-11]. The substrate's native oxide was thermally desorbed in the MBE growth chamber at a temperature of approximately 1000 ˚C. The substrate was then cooled to 490 ˚C, and was exposed to a Sb flux for 5 minutes. A 17-monolayer (ML) AlSb nucleation layer was then deposited at a growth rate of 0.36 ML/s. Next, a 2-µm-thick GaSb layer was deposited at a growth rate of 0.66 ML/s, using a two-step temperature technique[19] where the substrate temperature was increased to 515 ˚C after 1.5 µm of GaSb had been deposited. The substrate was then cooled to 435 ˚C and the 50-nm InAs BG was deposited at a growth rate of 0.4 ML/s, followed by a 8-nm-thick AlSb barrier at a growth rate of 0.1 ML/s. Next, the substrate was heated to 515 ˚C for the deposition of a 20-nm GaSb layer at 0.66 ML/s. The substrate was then cooled back to 435 ˚C for the remainder of the growth, starting with a 10-nm InAs channel, with n-type doping by Si to a nominal carrier density of 1×10^{18} cm^{-3}. This was followed by the TBRT layers, which consist of 1.8 nm of AlSb, 3.0 nm of InAs, 1.2 nm of AlSb, 2.4 nm of InAs and 1.8 nm of AlSb. Finally, a 10-nm thick InAs

FG layer completed the structure. The growth rates of the InAs and AlSb in the TBRT and FG were 0.2 ML/s and 0.1 ML/s respectively.

X-ray diffraction. High-resolution X-ray diffraction measurements were carried out on a Bruker D8 Discover system. The copper K-α X-ray beam was conditioned by a two-bounce Ge crystal and collimating optics. The diffracted signal was collected using a scintillation counter with a one-bounce Ge crystal at its entry slit. Fitting of ω-2θ data was carried out using RADS Mercury software.

Transmission electron microscopy. TEM specimens were prepared using standard techniques, *i.e.* grinding, polishing and ion milling to electron transparency using 6-kV Ar+ ions. A final low-energy (0.5 kV) ion mill was used to reduce surface damage. Samples were examined in a JEOL 2100 LaB_6 transmission electron microscope (TEM) operating at 200 kV.

Electron-channelling-contrast imaging. Electron-channelling-contrast images were collected using a Zeiss Gemini scanning electron microscope with a solid-state backscatter detector operating at 20 kV.

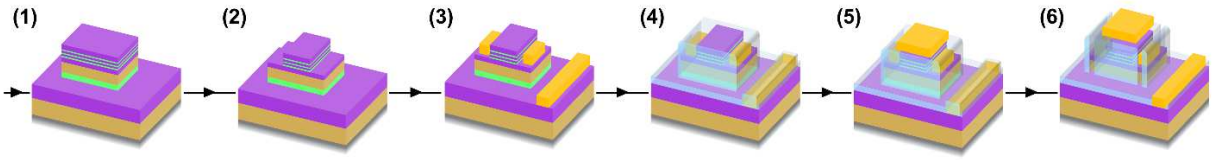


Fig. 5 | Process flow diagram for the fabrication of ULTRA**RAM™ memory cells.** The process progresses from left to right as labelled. **(1)** Inductively-coupled etch to the BG layer to form device mesas. **(2)** Alternating wet etch with citric and buffered oxide etchant to form S-D regions. **(3)** Ti-Au metallisation for S-D and BG. **(4)** Al_2O_3 gate dielectric deposition via atomic layer deposition. **(5)** Ti-Au metallisation for the CG terminal. **(6)** SiO_2 deposition for further passivation. In the final stage a buffered oxide etch is carried out in square photoresist windows within the S-D region: the slit structure here is presented to observe the etch within the schematic.

Device fabrication. A top-down processing procedure was employed to fabricate memory devices with gate lengths of 10 and 20 μm. Fig. 5 demonstrates the general process flow, in which a memory cell is formed via numerous steps. Separate ultraviolet (UV) lithography stages were used to pattern the BG, device mesa and source/drain areas. For the device mesa and BG, excess material was dry etched with an Oxford Instruments Plasma Lab 100 inductively-coupled plasma (ICP) machine. The etching process was carried out using a Cl_2/Ar (2.5/5 sccm) gas mixture and a chamber pressure of 10 mTorr, with an ICP power of 120 W and an RF power of 18 W. Removal of material to define the S and D areas was achieved by successive, alternating, selective wet-etching to remove the FG and TBRT layers of the heterostructure. A solution of citric acid, hydrogen peroxide and de-ionised water in 1:3:1 volumetric ratio was used to selectively etch InAs over AlSb, whilst buffered oxide etchant (BOE) at 10:1 was used to selectively etch AlSb over InAs. The use of BOE significantly improves etch uniformity compared to tetramethylammonium hydroxide-based etchant used previously[10], due to increased selectivity with the underlying GaSb layer, minimising damage by chemical contact through etch pits in the thin InAs channel layer. Thermally evaporated Ti-Au contact terminals were then added to the InAs channel layer (S and D), which was pre-treated by the HF from the prior etch step[28]. The Al_2O_3 gate dielectric is added via thermal atomic layer deposition using a Veeco Savannah S100 system. The deposition was carried out at 200 ˚C and consisted of 10 trimethylaluminum (TMA)-only pulses to ensure TMA self-cleaning (AsO_x removal) of the InAs FG[29], followed by 150 cycles of TMA and H_2O pulses, leading to an Al_2O_3 layer thickness of about 15 nm. Ti-Au CG contacts were deposited on top of the gate dielectric via thermal evaporation, followed by further passivation consisting of 120 nm of SiO_2 deposited using an Oxford Instruments plasma-enhanced chemical vapour deposition machine. Access to all device terminals was obtained by chemical etching of the oxide layers of SiO_2 and Al_2O_3 in the S, D and BG regions and SiO_2 and Ti in the CG region (using BOE, 10:1) through openings in the photoresist provided by UV lithography. A hard-baked, positive photoresist, lifting layer was then patterned on one edge of the devices by UV lithography, enabling

continuous contacts to access the device terminals. Lastly, final contacts for device probing and bonding were added by another Ti-Au thermal evaporation process.

Device Testing. Electrical characterisation was performed using a Keithley 2634B dual-channel source measure unit (SMU) controlled using dedicated LabVIEW programs. All tests were performed at room temperature and pressure. Remarkably, shielding the devices from varying ambient (room) light levels was found to be unnecessary. Device terminals were contacted using a Wentworth Laboratories probe station with triaxial probe connections. Endurance cycling (program-read-erase-read) was performed in pulsed-mode at around ~500 cycles per minute, where each of the program-read-erase-read pulses are separated by ~50 ms as a result of the pulse initiation delay of the SMU. Current measurements for the retention test were taken at 0.2 V S-D voltage at a rate of ~500 cycles per minute. The retention data was gathered under the same readout conditions as above in a single continuous run for the E state immediately after a 5-ms -2.5-V E-cycle (CG-D pulse), and similarly for P retention monitoring, which followed a +2.5-V pulse across the CG-D of the device of 5 ms duration.

Acknowledgements

This work was supported by the EPSRC, UK, via the 2017-2020 IAA funding allocation to Lancaster University under grant EP/R511560/1, a scholarship under grant EP/N509504/1 and the Future Compound Semiconductor Manufacturing Hub grant EP/P006973/, by the ATTRACT project funded by the EC under Grant Agreement 777222 and by the Joy Welch Educational Charitable Trust.

The underlying data in this paper is available from https://dx.doi.org/????? [URL will be added after peer review]

References

1 Wong, H. S. P. & Salahuddin, S. Memory leads the way to better computing. *Nature Nanotechnology* **10**, 191-194, doi:10.1038/nnano.2015.29 (2015).

2 Zahoor, F., Azni Zulkifli, T. Z. & Khanday, F. A. Resistive Random Access Memory (RRAM): an Overview of Materials, Switching Mechanism, Performance, Multilevel Cell (mlc) Storage, Modeling, and Applications. *Nanoscale Research Letters* **15**, 90, doi:10.1186/s11671-020-03299-9 (2020).

3 Sverdlov, V. & Selberherr, S. Silicon spintronics: Progress and challenges. *Physics Reports* **585**, 1-40, doi:https://doi.org/10.1016/j.physrep.2015.05.002 (2015).

4 Ielmini, D. & Lacaita, A. L. Phase change materials in non-volatile storage. *Materials Today* **14**, 600-607, doi:https://doi.org/10.1016/S1369-7021(11)70301-7 (2011).

5 Bolkhovityanov, Y. B. & Pchelyakov, O. P. III-V Compounds-on-Si: Heterostructure Fabrication, Application and Prospects. *The Open Nanoscience Journal* **3**, 20, doi:10.2174/1874140100903010020 (2009).

6 Fang, S. F. *et al.* Gallium arsenide and other compound semiconductors on silicon. *J. Appl. Phys.* **68**, R31-R58, doi:10.1063/1.346284 (1990).

7 Kim, Y. H. *et al.* Growth mode and structural characterization of GaSb on Si (001) substrate: A transmission electron microscopy study. *Appl. Phys. Lett.* **88**, 241907, doi:10.1063/1.2209714 (2006).

8 Akahane, K., Yamamoto, N., Gozu, S.-i. & Ohtani, N. Heteroepitaxial growth of GaSb on Si(001) substrates. *Journal of Crystal Growth* **264**, 21-25, doi:https://doi.org/10.1016/j.jcrysgro.2003.12.041 (2004).

9 Rodriguez, J. B., Madiomanana, K., Cerutti, L., Castellano, A. & Tournié, E. X-ray diffraction study of GaSb grown by molecular beam epitaxy on silicon substrates. *Journal of Crystal Growth* **439**, 33-39, doi:https://doi.org/10.1016/j.jcrysgro.2016.01.005 (2016).

10 Lane, D., Hodgson, P. D., Potter, R. J., Beanland, R. & Hayne, M. ULTRARAM: Toward the Development of a III-V Semiconductor, Nonvolatile, Random Access Memory. *IEEE Transactions on Electron Devices*, 1-4, doi:10.1109/TED.2021.3064788 (2021).

11 Hayne, M. Electronic Memory Devices. United States patent US10243086B2 (2019).

12 Lane, D. & Hayne, M. Simulations of Ultralow-Power Nonvolatile Cells for Random-Access Memory. *IEEE Transactions on Electron Devices* **67**, 474-480, doi:10.1109/TED.2019.2957037 (2020).

13 Vurgaftman, I., Meyer, J. R. & Ram-Mohan, L. R. Band parameters for III-V compound semiconductors and their alloys. *J. Appl. Phys.* **89**, 5815-5875, doi:10.1063/1.1368156 (2001).

14 Tizno, O. *et al.* Room-temperature Operation of Low-voltage, Non-volatile, Compound-semiconductor Memory Cells. *Scientific Reports* **9**, 8950, doi:10.1038/s41598-019-45370-1 (2019).

15 Prall, K. in *2017 IEEE International Memory Workshop (IMW).* 1-5.

16 Vajargah, S. H., Ghanad-Tavakoli, S., Preston, J. S., Kleiman, R. N. & Botton, G. A. Growth mechanisms of GaSb heteroepitaxial films on Si with an AlSb buffer layer. *J. Appl. Phys.* **114**, 113101, doi:10.1063/1.4820255 (2013).

17 Rodriguez, J. B. *et al.* Characterization of antimonide based material grown by molecular epitaxy on vicinal silicon substrates via a low temperature AlSb nucleation layer. *Journal of Crystal Growth* **477**, 65-71, doi:https://doi.org/10.1016/j.jcrysgro.2017.04.003 (2017).

18 Serincan, U. & Arpapay, B. Structural and optical characterization of GaSb on Si (001) grown by Molecular Beam Epitaxy. *Semicond. Sci. Technol.* **34**, 035013, doi:10.1088/1361-6641/aafcbe (2019).

19 Delli, E. *et al.* Heteroepitaxial Integration of Mid-Infrared InAsSb Light Emitting Diodes on Silicon. *IEEE Photonics Journal* **11**, 1-8, doi:10.1109/JPHOT.2019.2911433 (2019).
20 Delli, E. *et al.* Mid-infrared type-II InAs/InAsSb quantum wells integrated on silicon. *Appl. Phys. Lett.* **117**, 131103, doi:10.1063/5.0022235 (2020).
21 Ward, T. *et al.* Design rules for dislocation filters. *J. Appl. Phys.* **116**, 063508, doi:10.1063/1.4892162 (2014).
22 George, I. *et al.* Dislocation filters in GaAs on Si. *Semicond. Sci. Technol.* **30**, 114004, doi:10.1088/0268-1242/30/11/114004 (2015).
23 Banerjee, W. Challenges and Applications of Emerging Nonvolatile Memory Devices. *Electronics* **9**, 1029 (2020).
24 Pavan, P., Bez, R., Olivo, P. & Zanoni, E. Flash memory cells-an overview. *Proceedings of the IEEE* **85**, 1248-1271, doi:10.1109/5.622505 (1997).
25 Dennard, R. H. *et al.* Design of ion-implanted MOSFET's with very small physical dimensions. *IEEE Journal of Solid-State Circuits* **9**, 256-268, doi:10.1109/JSSC.1974.1050511 (1974).
26 Birner, S. *nextnano: Software for semiconductor nanodevices*, <https://www.nextnano.de/> (2021).
27 Birner, S. *et al.* nextnano: General purpose 3-D simulations. *Ieee Transactions on Electron Devices* **54**, 2137-2142, doi:10.1109/ted.2007.902871 (2007).
28 Nishi, K. *et al.* Effects of buffered HF cleaning on metal–oxide–semiconductor interface properties of Al2O3/InAs/GaSb structures. *Applied Physics Express* **8**, 061203, doi:10.7567/apex.8.061203 (2015).
29 Milojevic, M. *et al.* Half-cycle atomic layer deposition reaction studies of Al2O3 on In0.2Ga0.8As (100) surfaces. *Appl. Phys. Lett.* **93**, 202902, doi:10.1063/1.3033404 (2008).