ADVANCED ELECTRONIC MATERIALS
www.advelectronicmat.de

ULTRARAM: A Low-Energy, High-Endurance, Compound-Semiconductor Memory on Silicon

*Peter D. Hodgson, Dominic Lane, Peter J. Carrington, Evangelia Delli, Richard Beanland, and Manus Hayne**

ULTRARAM is a nonvolatile memory with the potential to achieve fast, ultralow-energy electron storage in a floating gate accessed through a triple-barrier resonant tunneling heterostructure. Here its implementation is reported on a Si substrate; a vital step toward cost-effective mass production. Sample growth using molecular beam epitaxy commences with deposition of 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 devices reveals retention in excess of 1000 years and degradation-free endurance of over 10^7 program/erase cycles, surpassing very recent results for similar devices on GaAs substrates.

1. Introduction

A memory that is fast and nonvolatile, 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, nonvolatile 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] magnetoresistive 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. ULTRARAM breaks this paradigm via the exploitation of InAs quantum wells (QWs) and AlSb barriers to create a triple-barrier resonant-tunneling (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 **Figure 1**a, allows it to become transparent to electrons when a low voltage (≈2.5 V) is applied, due to resonant tunneling. 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 ULTRARAM onto Si substrates is a vital step toward realizing 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. In contrast, III–V substrates are fragile, expensive, and generally only available in much smaller wafer sizes, making them less suitable for high-volume production. But III–V semiconductors do provide advantages such as high electron mobilities, superior optoelectronic properties and a greater degree of bandgap engineering, making them the preferred material for LEDs, laser diodes, infrared detectors and for power, radio frequency and high electron mobility transistors.[5–8] Therefore, the integration of the complementary technologies of III–Vs and Si is highly desirable

P. D. Hodgson, D. Lane, M. Hayne
Department of Physics
Lancaster University
Lancaster LA1 4YB, UK
E-mail: m.hayne@lancaster.ac.uk
P. J. Carrington, E. Delli
Department of Engineering
Lancaster University
Lancaster LA1 4YW, UK
R. Beanland
Department of Physics
University of Warwick
Coventry CV4 7AL, UK

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/aelm.202101103.

DOI: 10.1002/aelm.202101103

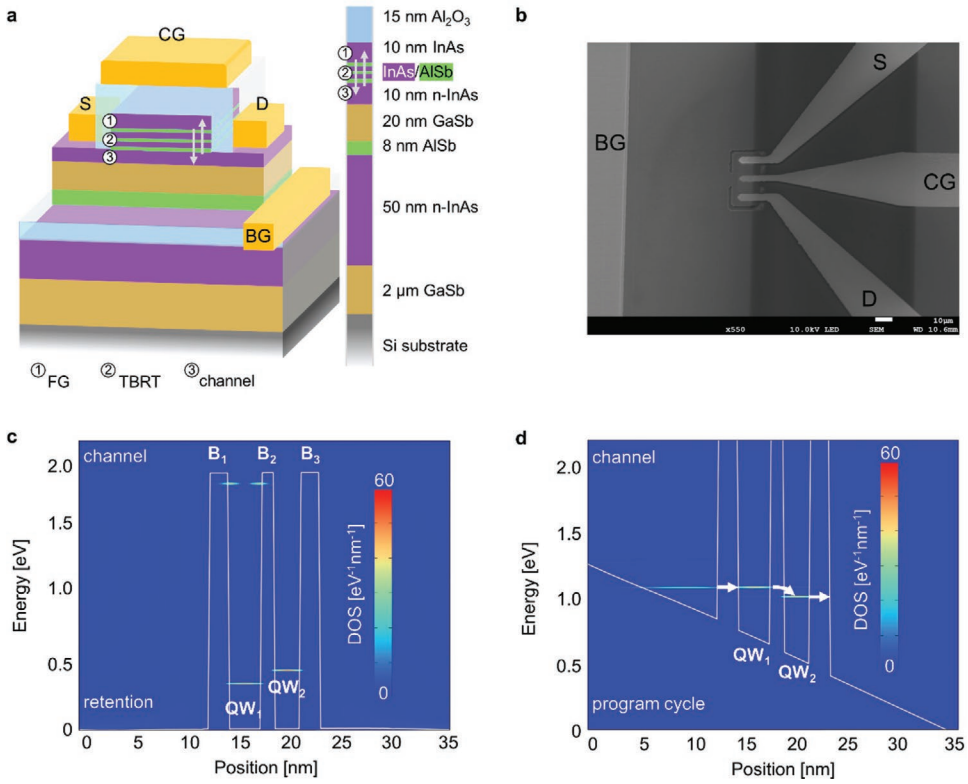

Figure 1. ULTRARAM device concept. a) Schematic cross-section of a device with corresponding material layers. The floating gate (FG), triple-barrier resonant-tunneling structure (TBRT), and readout channel are highlighted. Arrows indicate the direction of electron flow during program/erase operations. b) Scanning electron micrograph of a fabricated device of 10 μm gate length. c,d) Nonequilibrium Green's functions (NEGF) calculations of density of states alongside conduction band diagrams for no applied bias (i.e., retention) and program-cycle bias respectively. B_1, B_2, and B_3 are the AlSb barrier layers. QW_1 and QW_2 are the InAs quantum wells in the TBRT.

for many applications.[9] This 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.[10] More recently, heterogeneous epitaxy has offered a promising route toward the efficient cointegration of III–Vs and Si. However, direct epitaxial deposition is challenging due to the polar (III–Vs) to nonpolar (Si) interface, lattice mismatch and differences in thermal expansion coefficients.[11–14] In order to prevent defect formation and realize 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 ULTRARAM heterostructures onto Si substrates, their fabrication into devices and testing. Results show nonvolatile, high-endurance operation that even surpasses recently reported results on GaAs substrates.[15]

2. Memory Concept

ULTRARAM is a charge-based memory where the logic state is determined by the presence or absence of electrons in an FG. As can be seen in Figure 1a, 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 nondestructively 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.

The novelty underpinning the memory is the TBRT structure,[16] 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 QW layers.[17] When the memory is in the retention state (Figure 1c), 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, nonvolatility is strengthened by the QW ground states residing at an unusually high energy for a resonant-tunneling structure. This is due to a combination of the ultrathin QWs and the extraordinarily low electron effective mass in InAs.[18] 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, Figure 1d) 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 tunneling. Due to the low voltages required and the low capacitance per unit area of the device compared to DRAM, ultralow logic state switching energies of 10^{-17} J are predicted for 20 nm feature size ULTRARAM memories,[17,19] which is two and three orders of magnitude lower than DRAM and flash respectively.[17,20] However, before this ultralow switching energy can be realized by fabricating nm-scale devices, the fundamental properties of µm-scale devices must first be understood and optimized. ULTRARAM prototype devices grown on GaAs substrates have previously exhibited experiment-limited, not device-limited, nonvolatile retention of 10^5 s and an endurance of 10^6 program-erase cycles.[15] Thus, the demonstration of devices grown on Si substrates with similar or improved performance would be a major step toward commercialization.

3. Epitaxy of III–Vs on Si and Sample Characterization

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 toward the [0–11] crystal direction, creating diatomic steps on the nonpolar Si surface, which suppresses the formation of antiphase domains (APDs) during growth of polar III–V materials.[10] Prior to the growth of a complete structure with the 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 of a 17 monolayer (ML) AlSb nucleation layer, which forms into 3D islands.[21] These islands reduce the diffusion length of Ga atoms during the initial growth of the subsequent 2 µm GaSb buffer, helping to promote 2D epitaxy and preventing the formation of planar twinning defects.[13,14,22] 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.[23] A two-temperature-step GaSb growth method[24] was used to reduce the density of vertically propagating threading dislocations (see Experimental Section for more details).

With this approach, an optimized GaSb/Si buffer layer was obtained with a surface defect density of $(2.5 \pm 0.1) \times 10^8$ cm^{-2}, as measured by electron channeling contrast imaging (ECCI, see Figure S1, Supporting Information), and a root-mean-square surface roughness of 1.9 ± 0.2 nm, as measured by atomic force microscopy (see Figure S2, Supporting Information) 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 (see Figure S3, Supporting Information). Cross-sectional transmission electron microscope (TEM) images, such as **Figure 2**a, show that the

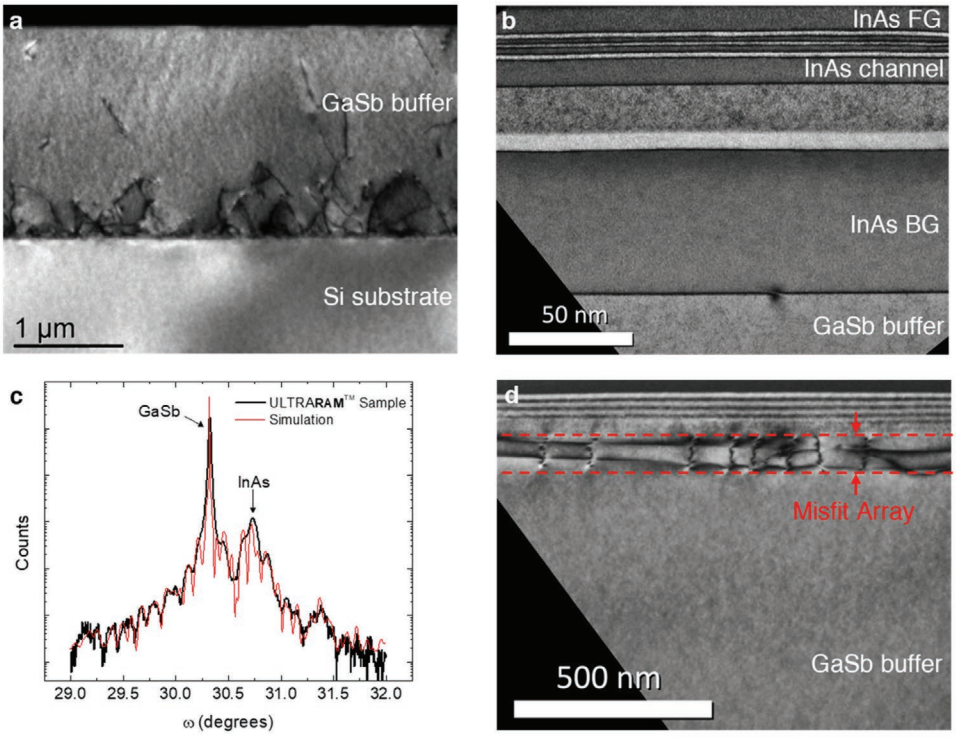

Figure 2. III–V on Si material characterization. a) Dark-field $g = 220$ transmission electron microscope (TEM) image of a GaSb/Si buffer layer. b) Dark-field $g = 002$ TEM image of the ULTRARAM sample. Only the memory layers and the top of the GaSb buffer are visible in this image. A single 60° misfit dislocation is visible in the InAs/GaSb buffer interface. c) Experimental and simulated ω-2θ high-resolution X-ray diffraction scan of the ULTRARAM sample. The Si substrate's diffraction peak occurs at a larger angle and is not shown (the full scan is displayed in Figure S4, Supporting Information). d) Dark-field $g = 220$ TEM image of the upper part of the structure. The specimen tilt of ≈20° allows the misfit dislocation array in the InAs back gate (BG)/buffer layer interface to be seen. The lower and upper dashed lines indicate the intersection of this interface with the specimen surfaces.

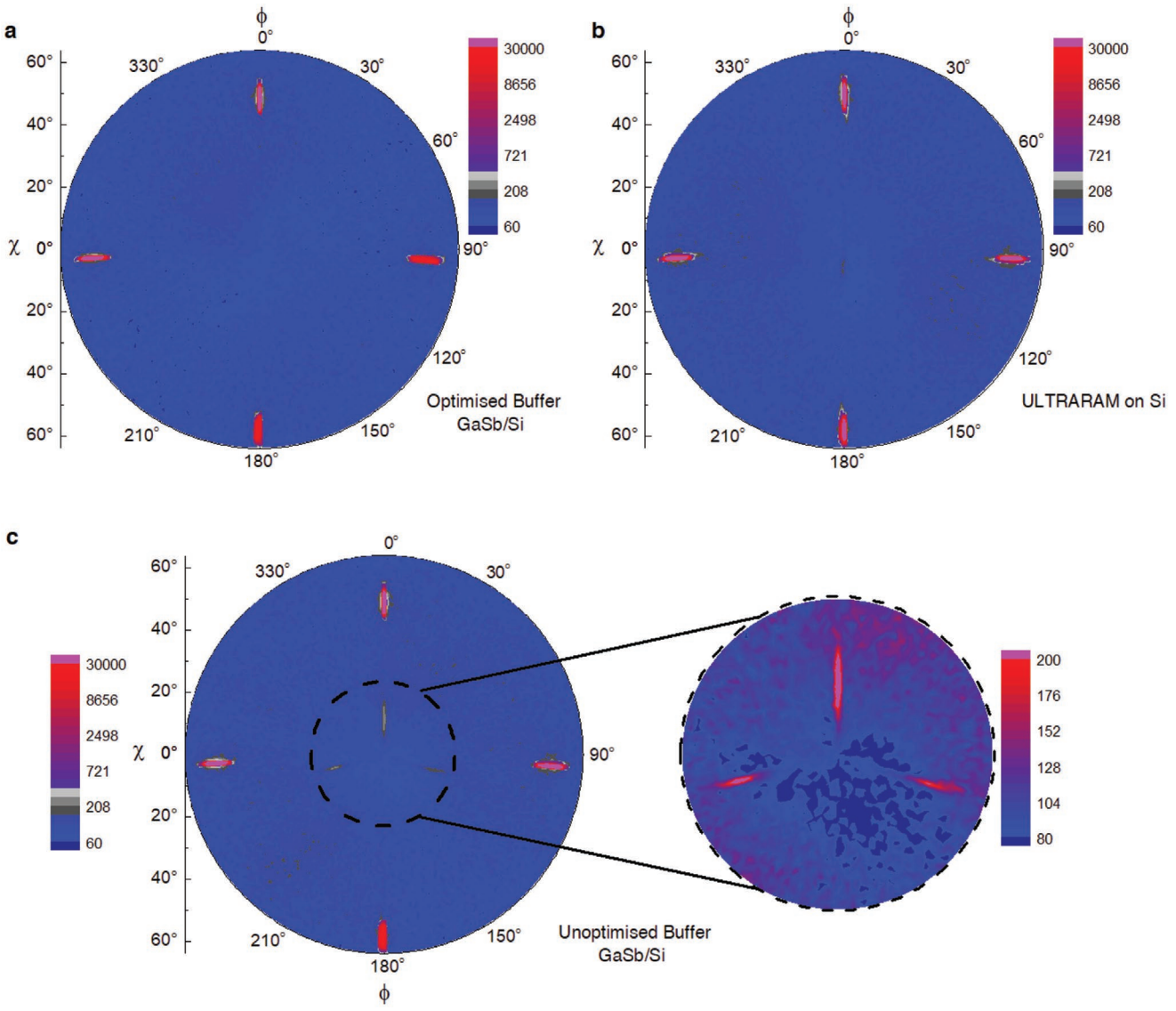

Figure 3. In-plane pole figures of the GaSb (111) diffraction peak from three different samples. The radial (χ) axis corresponds to the inclination of the sample in the scattering plane. The angular (φ) axis corresponds to the rotation of the sample about its surface normal. The data shown in (a) and (b) are for the optimized GaSb/Si buffer and ULTRARAM on Si respectively. Data from the unoptimized GaSb/Si sample is shown in (c), with a signal from planar twinning defects visible at low χ angles. The inset shows a magnified and rescaled view of this region to highlight these additional features.

buffer layer is APD free, and that the majority of threading dislocations are confined to the first 500 nm of GaSb.

XRD pole-figure measurements of the GaSb (111) plane were used to further investigate the quality of the GaSb/Si buffer layer. Data from the optimized GaSb/Si buffer sample described above and an unoptimized GaSb/Si buffer sample are shown in **Figure 3**a,c respectively. The unoptimized sample used a higher Sb flux throughout the growth and the substrate temperature remained at 490 °C for deposition of the entire GaSb layer. In contrast, for the optimized buffer, a lower Sb flux was used and the substrate temperature was raised to 515 °C after 1.5 µm of GaSb had been deposited. In both samples, the four most intense peaks, at φ = 0°, 90°, 180°, and 270°, correspond to diffraction from the {111} planes and have the expected fourfold symmetry. The slight deviation of the peaks from φ = 90° and 270° is due to a combination of the sample offcut and the nonzero widths of the detector opening and X-ray beam. The substrate offcut is also responsible for shifting the χ angle of the peaks at φ = 0° and 180° by ±4°. The additional signal observed in the unoptimized sample at low χ angles (Figure 3c) indicates the presence of planar twinning defects,[22] which create (111)-type planes at shallower angles. These twinning planes also have fourfold symmetry. However, as has previously been detailed by Devenyi et al.,[25] the strong suppression (enhancement) of twin formation in the direction opposite (toward) the substrate offcut results in a reduction (increase) of the signal at φ = 180 (φ = 0). Consequently, only three peaks are seen, with the φ = 180° signal undetectable against background noise and the φ = 0° peak having the strongest signal. No such twinning defect signal was observed in the optimized buffer

(Figure 3a), indicating that it has an extremely low planar twinning defect density.[22]

These results demonstrate that the optimized GaSb buffer layer on Si is of high quality, and suitable for deposition of the memory layers. The first epilayers grown on such a buffer are the 50 nm InAs n-type BG and the 8 nm AlSb barrier, which separates the BG from the active parts of the device, as shown in Figures 1a and 2b. 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 Experimental Section.

The ECCI-measured surface defect density in the ULTRARAM sample was found to be $(2.1 \pm 0.1) \times 10^8$ 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.[26–29] A high-resolution XRD ω-2θ scan of the memory sample, which displays clear satellite peaks from the memory epilayers, is shown in Figure 2c. 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 (Figure 2d), which is 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. In-plane pole figure measurements (Figure 3b) indicate that the memory structure retains the extremely low planar twinning defect density observed for the GaSb/Si buffer.

4. Basic Memory Operations

Single bit memory cells on Si substrates were fabricated at gate lengths of 10 and 20 µm. Figure 1a shows a schematic cross-section of such a device, whilst a scanning electron microscope image of a 10 µm gate-length device is shown in Figure 1b, with terminals labeled. 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.

In this work, an n-type InAs, normally on, channel layer was chosen for simplicity, as the primary objective was to demonstrate single-cell ULTRARAM operation on Si substrates. Whilst this channel design is sufficient for that purpose, it results in a low current contrast, $\Delta I_{S\text{-}D}$, between programmed (0) and erased (1) logic states (**Figure 4**a). It is important to emphasize that this is a consequence of the simplicity of the channel and not a result of insufficient contrast in the programmed and erased state of the FG. Indeed, the implementation of a normally-off channel in future devices 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.[17,30] InAs channel transistors with submicrometer feature sizes and a subthreshold swing of <100 mV dec^{-1} have previously been demonstrated.[31] Consequently, due to the threshold voltage window of 350 mV in our devices (Figure S5, Supporting Information) one can expect the 0/1 current contrast of ULTRARAM to improve to three decades with the implementation of a normally-off channel. Such an improvement of the 0/1 contrast through careful modification of the channel will allow memory arrays to be built with a novel high-density RAM architecture.[15,17]

5. 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 h for both program and erase states using >10^6 readout operations, limited only by the length of the experiment (Figure 4a). 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 h. 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 Figure 4b. 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 h makes determination of the retention time difficult, as shown by the dashed line in Figure 4b, which extends to infinity. Therefore, a second fitting is shown (solid line in Figure 4b), which follows the decay of the memory states, prior to the final 10 h of data. This provides an extremely conservative lower limit for the memory's retention of at least 10^7 h, which is more than 1000 years.

A similar, but more prominent, initial state decay was previously observed in the first prototype ULTRARAM devices on GaAs substrates,[19] and was subsequently eliminated with improved material quality.[15] 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 h test and the extremely long predicted retention times clearly demonstrate the nonvolatility 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 Figure 4c. 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,

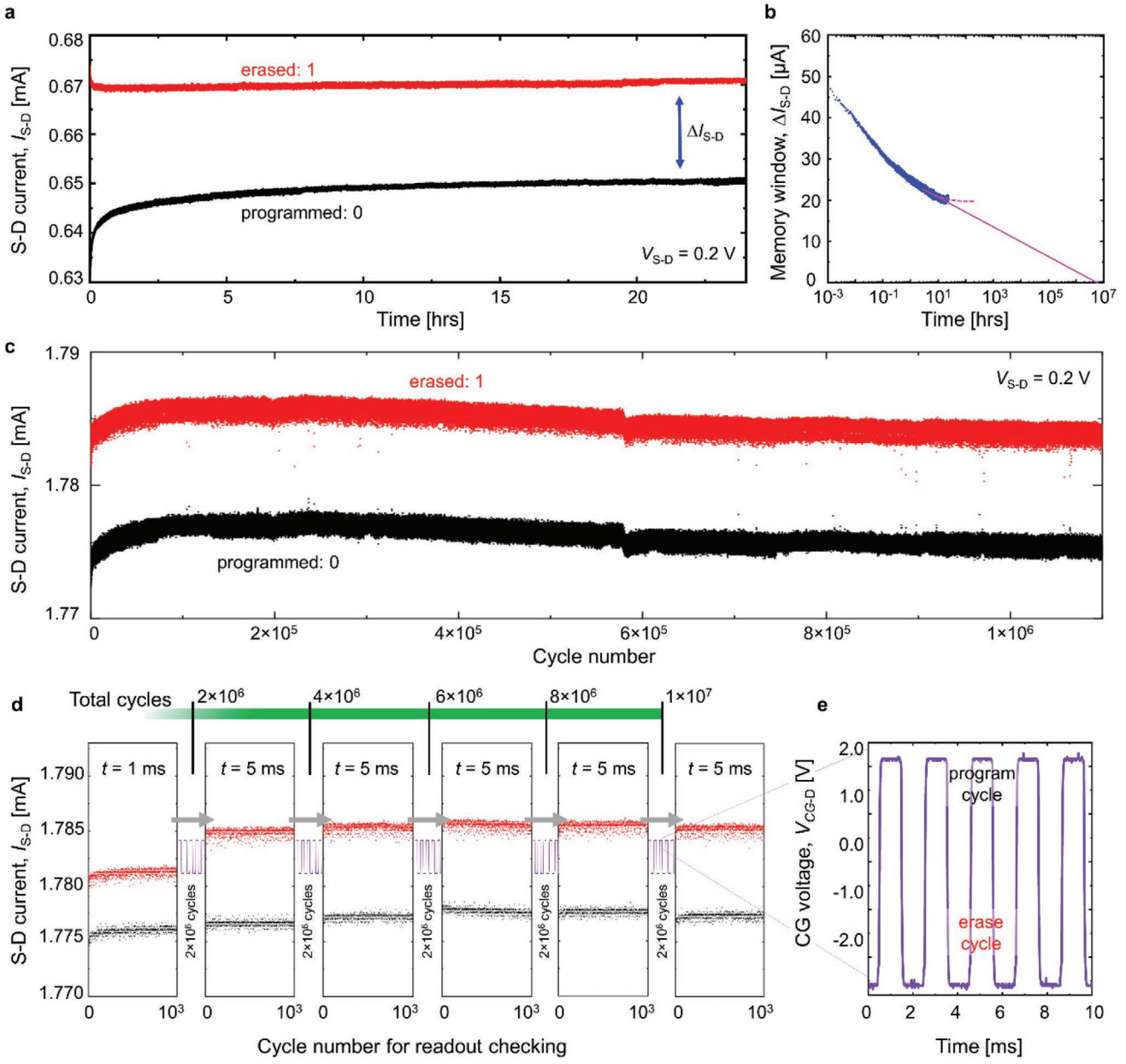

Figure 4. Retention and endurance characteristics. a) Retention data for a 20 μm gate-length cell. Program and erase cycles consisted of 10 ms duration pulses at +2.5 and −2.5 V, respectively. Readout is performed at an S-D bias of 0.2 V and in the absence of gate bias. b) S-D current difference ($\Delta I_{S\text{-}D}$) for the >24 h retention plotted on a log scale. The continuous magenta line indicates the most conservative interpretation of the decay data, whereas the dashed magenta line follows the stabilization 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 are 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.

with neither the pronounced drift observed in the first prototypes,[19] nor the large fluctuations in I_{SD} seen in more recent devices.[15] The latter was attributed to nonoptimal etching of the channel during device fabrication and has been almost completely eliminated due to an improved process (see Experimental Section for details). In ref. [19] 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 tunneling to occur at a lower voltage for the program cycle than for the erase cycle.[17] As a result, the use of symmetric voltages, such as ±2.5 V, results in overprogramming, 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 stabilize with repeated cycling once the correct voltages are identified. The remaining small drift of the I_{SD} window observed in Figure 4c over many thousands of cycles is likely to become inconsequential with the implementation of the normally-off channel readout scheme.

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. First, 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 Figure 4e. 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 Figure 4d, 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.[20,32]

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 ULTRARAM on GaAs substrates.[15] In both cases, the devices operate at a remarkably high speed for their large (20 µm) feature size. Assuming ideal capacitive scaling[33] 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 resistive-capacitive-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.[15,19]

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 Figure 4a,c shows this clearly and is the most extreme example to be found on the chip (for the full variation for all devices tested see Figure S6, Supporting Information). It is believed to be caused by S-D to BG current leakage due to etch pitting of defects during fabrication, or surface leakage, which creates an alternative conducting path from S to D. Whilst the TEM image in Figure 2d 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. All memory cells on the chip required similar switching voltages (around 2.5 V) to achieve a nonvolatile memory state. This result points toward good uniformity of the tunneling layers across the sample. Indeed, recent modeling work on the tunneling layers suggest that the layer thickness tolerance of the barrier structure could be as high as 50% before nonvolatile memory operation ceases, which is an encouraging result for future large-scale implementation on Si.[34]

6. Conclusions

We have demonstrated ULTRARAM compound-semiconductor, floating-gate memory cells on Si substrates. A high-quality GaSb buffer on Si, with undetectably low levels of planar twining defects, was developed, onto which the memory layers were successfully implemented. TEM images and XRD scans revealed the presence of threading defects in the InAs BG, which may 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 $(2.1 \pm 0.1) \times 10^8$ 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 nonvolatile 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.

7. Experimental Section

Nonequilibrium Green's Function Calculations: The quantum transport of electrons through the resonant tunneling structure was calculated with the nextnano.MSB software package, which uses the nonequilibrium 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 tunneling structure. Further details of the method can be found in refs. [17, 35, 36].

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 toward [0–11]. The substrate's native oxide was thermally desorbed in the MBE growth chamber at a temperature of ≈1000 °C. The substrate was then cooled to 490 °C and was exposed to a Sb flux for 5 min. A 17 ML AlSb nucleation layer was then deposited at a growth rate of 0.36 ML s^{-1}. Next, a 2 µm thick GaSb layer was deposited at a growth rate of 0.66 ML s^{-1}, using a two-step temperature technique[24] 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^{-1}, followed by a 8 nm thick AlSb barrier at a growth rate of 0.1 ML s^{-1}. Next, the substrate was heated to 515 °C for the deposition of a 20 nm GaSb layer at 0.66 ML s^{-1}. 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

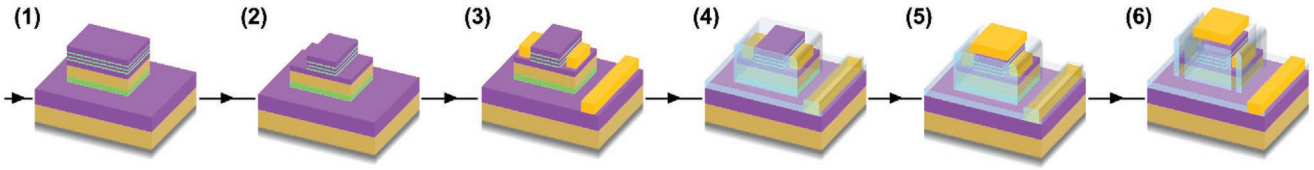

Figure 5. Process flow diagram for the fabrication of ULTRARAM memory cells. The process progresses from left to right as labeled: (1) Inductively coupled etch to the BG layer to form device mesas. (2) Alternating wet etch with citric and buffered oxide etchant to form S-D regions. (3) Ti-Au metallization for S-D and BG. (4) Al_2O_3 gate dielectric deposition via atomic layer deposition. (5) Ti-Au metallization 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.

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 and 0.1 ML s^{-1} 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. For ω and ω-2θ measurements, a one-bounce Ge crystal was used at the detector's entry slit. Fitting of ω-2θ data was carried out using Bede Rocking-curve Analysis by Dynamical Simulation Mercury software. Pole figure measurements used a variable slit instead of the one-bounce Ge crystal. The variable detector slit was set to 1 mm for sample alignment and 5 mm for pole figure measurement. To align the sample the ω, 2θ, and χ angles were optimized to the GaSb (111) diffraction peak at $\varphi = 0°$. The pole figure measurements were carried out using Bruker DIFFRAC. SUITE Wizard software with φ and χ resolutions of 1° and automatic thinning to minimize oversampling at low χ angles.

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 TEM operating at 200 kV.

Electron-Channeling-Contrast Imaging: Electron-channeling-contrast images were collected using a Zeiss Gemini scanning electron microscope with a solid-state backscatter detector operating at 20 kV.

Device Fabrication: A top-down processing procedure was employed to fabricate memory devices with gate lengths of 10 and 20 µm. **Figure 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 deionized 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,[15] due to increased selectivity with the underlying GaSb layer, minimizing 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 pretreated by the HF from the prior etch step.[37] 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,[38] 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 vapor 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 characterization 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 min^{-1}, 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 min^{-1}. The retention data were gathered under the same readout conditions as above in a single continuous run for the erase state immediately after a 5 ms −2.5 V erase cycle (CG-D pulse), and similarly for program retention monitoring, which followed a +2.5 V pulse across the CG-D of the device of 5 ms duration.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

P.D.H. and D.L. contributed equally to this work. This work was supported by the Engineering and Physical Sciences Research Council, UK, via the 2017–2020 Impact Acceleration Account funding allocation to Lancaster University under grant EP/R511560/1, a scholarship under grant EP/N509504/1, equipment funding under grant EP/T023260/1, and the Future Compound Semiconductor Manufacturing Hub grant EP/P006973/1, by the ATTRACT project funded by the EC under Grant Agreement 777222 and by the Joy Welch Educational Charitable Trust.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available in Lancaster University Research Data Repository at https://doi.org/10.17635/lancaster/researchdata/496, reference number 496.

Keywords

compound semiconductor on silicon, molecular beam epitaxy, nonvolatile memory, triple-barrier resonant tunneling heterostructure

Received: October 11, 2021
Revised: November 25, 2021
Published online:

[1] H. S. P. Wong, S. Salahuddin, *Nat. Nanotechnol.* **2015**, *10*, 191.
[2] F. Zahoor, T. Z. Azni Zulkifli, F. A. Khanday, *Nanoscale Res. Lett.* **2020**, *15*, 90.
[3] V. Sverdlov, S. Selberherr, *Phys. Rep.* **2015**, *585*, 1.
[4] D. Ielmini, A. L. Lacaita, *Mater. Today* **2011**, *14*, 600.
[5] T. P. Xiao, K. Chen, P. Santhanam, S. Fan, E. Yablonovitch, *J. Appl. Phys.* **2018**, *123*, 173104.
[6] M. C. Amann, W. Hofmann, *IEEE J. Sel. Top. Quantum Electron.* **2009**, *15*, 861.
[7] A. P. Craig, F. Al-Saymari, M. Jain, A. Bainbridge, G. R. Savich, T. Golding, A. Krier, G. W. Wicks, A. R. Marshall, *Appl. Phys. Lett.* **2019**, *114*, 151107.
[8] H. Hamada, T. Tsutsumi, H. Matsuzaki, T. Fujimura, I. Abdo, A. Shirane, K. Okada, G. Itami, H. J. Song, H. Sugiyama, H. Nosaka, *IEEE J. Solid-State Circuits* **2020**, *55*, 2316.
[9] J.-S. Park, M. Tang, S. Chen, H. Liu, *Crystals* **2020**, *10*, 1163.
[10] Y. B. Bolkhovityanov, O. P. Pchelyakov, *Open Nanosci. J.* **2009**, *3*, 20.
[11] S. F. Fang, K. Adomi, S. Iyer, H. Morkoç, H. Zabel, C. Choi, N. Otsuka, *J. Appl. Phys.* **1990**, *68*, R31.
[12] Y. H. Kim, J. Y. Lee, Y. G. Noh, M. D. Kim, S. M. Cho, Y. J. Kwon, J. E. Oh, *Appl. Phys. Lett.* **2006**, *88*, 241907.
[13] K. Akahane, N. Yamamoto, S.-i. Gozu, N. Ohtani, *J. Cryst. Growth* **2004**, *264*, 21.
[14] J. B. Rodriguez, K. Madiomanana, L. Cerutti, A. Castellano, E. Tournié, *J. Cryst. Growth* **2016**, *439*, 33.
[15] D. Lane, P. D. Hodgson, R. J. Potter, R. Beanland, M. Hayne, *IEEE Trans. Electron Devices* **2021**, *68*, 2271.
[16] M. Hayne, *United States Patent US10243086B2*, **2019**.
[17] D. Lane, M. Hayne, *IEEE Trans. Electron Devices* **2020**, *67*, 474.
[18] I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan, *J. Appl. Phys.* **2001**, *89*, 5815.
[19] O. Tizno, A. R. J. Marshall, N. Fernández-Delgado, M. Herrera, S. I. Molina, M. Hayne, *Sci. Rep.* **2019**, *9*, 8950.
[20] K. Prall, presented at 2017 IEEE Int. Mem. Workshop (IMW), Monterey, California, 14–17 May **2017**.
[21] S. H. Vajargah, S. Ghanad-Tavakoli, J. S. Preston, R. N. Kleiman, G. A. Botton, *J. Appl. Phys.* **2013**, *114*, 113101.
[22] J. B. Rodriguez, L. Cerutti, G. Patriarche, L. Largeau, K. Madiomanana, E. Tournié, *J. Cryst. Growth* **2017**, *477*, 65.
[23] U. Serincan, B. Arpapay, *Semicond. Sci. Technol.* **2019**, *34*, 035013.
[24] E. Delli, P. D. Hodgson, E. Repiso, A. P. Craig, J. P. Hayton, Q. Lu, A. R. J. Marshall, A. Krier, P. J. Carrington, *IEEE Photonics J.* **2019**, *11*, 2200608.
[25] G. A. Devenyi, S. Y. Woo, S. Ghanad-Tavakoli, R. A. Hughes, R. N. Kleiman, G. A. Botton, J. S. Preston, *J. Appl. Phys.* **2011**, *110*, 124316.
[26] E. Delli, P. D. Hodgson, M. Bentley, E. Repiso, A. P. Craig, Q. Lu, R. Beanland, A. R. J. Marshall, A. Krier, P. J. Carrington, *Appl. Phys. Lett.* **2020**, *117*, 131103.
[27] T. Ward, A. M. Sánchez, M. Tang, J. Wu, H. Liu, D. J. Dunstan, R. Beanland, *J. Appl. Phys.* **2014**, *116*, 063508.
[28] I. George, F. Becagli, H. Y. Liu, J. Wu, M. Tang, R. Beanland, *Semicond. Sci. Technol.* **2015**, *30*, 114004.
[29] E. Delli, V. Letka, P. D. Hodgson, E. Repiso, J. P. Hayton, A. P. Craig, Q. Lu, R. Beanland, A. Krier, A. R. J. Marshall, P. J. Carrington, *ACS Photonics* **2019**, *6*, 538.
[30] P. Pavan, R. Bez, P. Olivo, E. Zanoni, *Proc. IEEE* **1997**, *85*, 1248.
[31] S. W. Chang, X. Li, R. Oxland, S. W. Wang, C. H. Wang, R. Contreras-Guerrero, K. K. Bhuwalka, G. Doornbos, T. Vasen, M. C. Holland, G. Vellianitis, M. J. H. van Dal, B. Duriez, M. Edirisooriya, J. S. Rojas-Ramirez, P. Ramvall, S. Thoms, U. Peralagu, C. H. Hsieh, Y. S. Chang, K. M. Yin, E. Lind, L. E. Wernersson, R. Droopad, I. Thayne, M. Passlack, C. H. Diaz, presented at IEEE, International Electron Devices Meeting, Washington, DC, 9–11 December **2013**.
[32] W. Banerjee, *Electronics* **2020**, *9*, 1029.
[33] R. H. Dennard, F. H. Gaensslen, H. Yu, V. L. Rideout, E. Bassous, A. R. LeBlanc, *IEEE J. Solid-State Circuits* **1974**, *9*, 256.
[34] D. Lane, M. Hayne, *J. Phys. D: Appl. Phys.* **2021**, *54*, 355104.
[35] nextnano: Software for Semiconductor Nanodevices, https://www.nextnano.de/ (accessed: September 2021).
[36] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, P. Vogl, *IEEE Trans. Electron Devices* **2007**, *54*, 2137.
[37] K. Nishi, M. Yokoyama, H. Yokoyama, T. Hoshi, H. Sugiyama, M. Takenaka, S. Takagi, *Appl. Phys. Express* **2015**, *8*, 061203.
[38] M. Milojevic, F. S. Aguirre-Tostado, C. L. Hinkle, H. C. Kim, E. M. Vogel, J. Kim, R. M. Wallace, *Appl. Phys. Lett.* **2008**, *93*, 202902.