The structural, electronic and optical properties of GaSb/GaAs nanostructures for charge-based memory

M. Hayne[1], R. J. Young,[1] E. P. Smakman,[2] T. Nowozin,[3] P. Hodgson,[1] J. K. Garleff,[2] P. Rambabu,[2] P. M. Koenraad,[2] A. Marent,[3] L. Bonato,[3] A. Schliwa[3] and D. Bimberg[3,4]

[1]Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom
[2]Department of Applied Physics, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[3]Institut für Festkörperphysik, Technische Universität Berlin, Hardenbergstrasse 36, 10623 Berlin, Germany
[4] Also at King-Abdul-Aziz-University, Jeddah, Saudi Arabia

Abstract

The potential for GaSb nanostructures embedded in GaAs to operate as charge-based memory elements at room-temperature is introduced and explored. Cross-sectional scanning-tunnelling microscopy is employed to directly probe and optimise the growth of nanostructures by molecular beam epitaxy. The results of structural analysis are combined with electrical measurements made with deep-level transient spectroscopy, showing excellent agreement with theoretical calculations which model the electronic structure of the nanostructures using 8-band k·p theory. Hole-localisation energies exceeding 600 meV in quantum dots and near-100% material contrast between GaSb-rich quantum rings and the surrounding GaAs matrix are revealed (no intermixing). Optical measurements confirm the depth of the hole localisation, and demonstrate substantially lower inhomogenous broadening than has previously been reported. Multiple peaks are partially resolved in ensemble photoluminescence of GaSb/GaAs quantum rings, and are attributed to charge states from discrete numbers of confined holes.

1. Introduction

Semiconductor quantum dots[1] (QDs) are of interest in a great number of device applications as they can provide access to zero-dimensional electronic properties in device structures, and form a strong interface between light and electronics,[2] which can be exploited in optoelectronics. Incorporation of QDs into conventional structures promises to improve, for example, the performance of lasers[3] or the efficiency of photovoltaic cells.[4] Individual QDs can be optically or electronically addressed, which has led to numerous demonstrations of QD-based devices targeting the nascent field of quantum information processing (QIP); namely high-frequency single photon sources[5] and

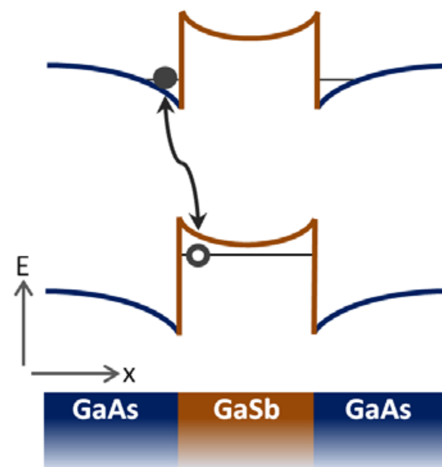

Figure 1. An illustration of valence (lower line) and conduction (upper line) bands formed when GaSb is embedded in GaAs. X indicates a spatial axis, e.g. the growth direction in molecular beam epitaxy, whilst E corresponds to energy. As illustrated, it is energetically favourable for an electron (solid circle) in the conduction band to be located in the GaAs, and for a heavy-hole in the valence band to be in the GaSb region. This asymmetric confinement of a bound electron-hole pair is referred to as type II.

detectors,[6] entangled photon sources[7] and memories.[8] QDs formed through the creation of GaSb nanostructures in a GaAs matrix have some unique and appealing properties that have yet to be fully exploited. This system has a type-II band alignment, as illustrated in Figure 1, providing strong spatial confinement for holes, and only binding electrons via the Coulomb interaction:[9] this contrasts with type-I systems which confine both electrons and holes. Note that there is no intrinsic advantage per se to type-II confinement (memories are unipolar devices), it is just that the difference in (unstrained) band gaps between GaAs and GaSb is taken up entirely by the valence band, so that the large hole localisation energy makes this system particularly interesting for room-temperature charge-based memories.[10,11]

In this review we report on progress in targeting applications of GaSb/GaAs nanostructures in charge-based memories. Section 2 briefly reviews conventional data storage technologies, while section 3 introduces the concept of QD memories. In section 4 we describe the molecular beam epitaxy (MBE) growth procedure used in this work. In section 5, cross-sectional scanning tunnelling microscopy (X-STM) is used to study the structures at the atomic level, and with it the growth procedure is refined. Section 6 discusses electronic measurements to quantify the degree of confinement of holes. Section 7 presents the results of optical measurements, demonstrating single-carrier charging and confirming the localisation energies measured electronically. In section 8 a 3D implementation of the 8-band k·p envelope function theory is combined with structural information from section 5 to assess the influence of morphology on confinement. In section 9 we bring all these considerations together, discussing the prospects for GaSb-based QD memories. Finally section 10 concludes.

2. Digital data storage

Digital data is stored using a variety of physical phenomena: magnetically (hard disk drives, tape), optically (CD, DVD, blu-ray), resistively (phase change memory, memristor) and electronically using charge storage. Often a distinction is drawn between data storage and memory. The former should be non-volatile and cheap, allowing huge amounts of information to be stored safely at low cost, but as a consequence it is invariably slow. On the other hand, a memory is usually fast, but may be neither cheap nor non-volatile. The Holy Grail of digital data storage is one which combines the best of both worlds, being cheap and non-volatile, but also fast. Such a memory, sometimes called universal memory or storage class memory is seen as a single technological solution to all data storage scenarios, and would have the additional advantages of removing the need for computers to boot-up as well as using less power: an important factor for mobile devices and of substantial environmental importance. However, despite considerable research into a wide variety of novel memory concepts, such a memory does not yet exist. Phase change memory, in which data is stored as a change in resistance in chalcogenide materials such as $GeSb_xTe_y$ (also used in optical storage media) is the most promising candidate, but has failed to make a significant impact to date, and is principally seen as a NOR Flash replacement: Si-based charge-storage memory continues to dominate. There are four main forms of Si-based memories, dynamic random access memory (DRAM), static RAM (SRAM), NOR Flash and NAND Flash. SRAM and NOR Flash are niche products. SRAM provides very fast but power-hungry and expensive random access, while NOR Flash can be thought of as a poor imitation of universal memory, providing non-volatile single bit access, but slow. It has been used extensively in mobile phones, but is now being replaced by DRAM plus NAND Flash. Traditional memory is therefore completely dominated by DRAM and NAND Flash, each performing very different and complementary functions. DRAM is the main RAM memory used in

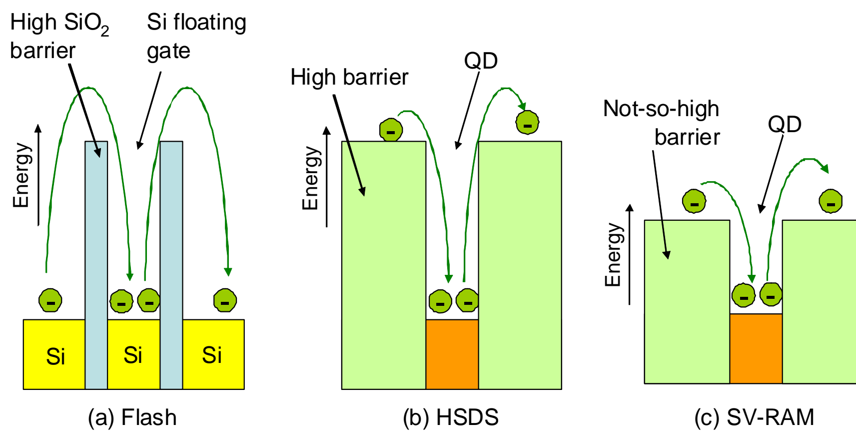

Figure 2. Schematic illustration of charge storage in (a) Flash and (b),(c) QD memory. In Flash charge is stored in a poly-silicon floating gate which is isolated from the channel by high SiO_2 barriers that must be overcome to write or erase the data. This is a slow process and eventually damages the device. In QD memory charge is captured in a potential well, which is an intrinsically fast process. Erase will still be slow, but unlike write, erase does not have to be done on demand. A second advantage of QD memory is that it is easy to tune the height of the barrier, so that different classes of memory can be made using the same basic design. Hence (b) is a non-volatile memory with fast write speed, which we call high-speed data storage (HSDS), while (c) is a semi-volatile RAM (SV-RAM) with improved refresh compared with DRAM and non-destructive read.

electronic devices, and provides fast (10 ns) single bit access (i.e. random access). Charge is stored capacitively, which means that it leaks away, so DRAM has to be refreshed approximately every 64 ms. A significant disadvantage of DRAM is that it has destructive read. NAND Flash, commonly just called Flash, is a remarkable story. Evolved from electrically-erasable programmable read-only memory (EEPROM), it has become the technology driver in electronics, taking over from DRAM in the 90 nm generation in 2003. Increases in bit density in NAND Flash are nothing less than extraordinary, and have been principally driven by physical scaling (traditional Moore's law), but recent advances have been as much to do with increasing the number of bits per cell. A state-of-the-art chip with features at 19 or 22 nm will have 3 bits per cell, with each level represented by as few as 15 electrons. According to the International Technology Roadmap for Semiconductors, physical scaling of Flash, as characterised by the half-pitch size, will decrease at a slower pace than needed to satisfy Moore's Law until 2022, when it will grind to a halt at 8 nm,[12] making research into modified or replacement technology an extremely active area of research. However, all contenders have so far not been able to compete with the low cost per bit (bit density) achieved by Flash, with current hopes pinned on a 3D version being implemented by a so-called bit-cost scalable (BiCS) process.

Flash is a modified field-effect transistor: charge is stored in a floating gate which alters the conductivity of the channel between source and drain. Despite its name Flash is intrinsically slow. To write or erase data requires pushing charge through the SiO_2 barrier separating the floating gate from the channel [Fig. 2(a)], which also gradually damages the device, giving Flash low endurance. The architecture of a Flash chip also makes single bit access (read, write or erase) very slow, i.e. it is not a RAM. However, this is not important for data storage, which is designed to give access to large amounts of data by accessing pages of data in parallel. This means that although Flash is thought of as memory, it is actually a data storage technology that performs faster than, for example, hard disk

drives, but at higher cost per bit. This also illustrates an important observation regarding a truly universal memory; the concept is a bit of a myth. RAM requires single bit access so will always have a more complex and expensive architecture than is needed for data storage. Hence, both memory (RAM) and data storage will always be required, but this does not exclude them being implemented using the same core information-storage technology.

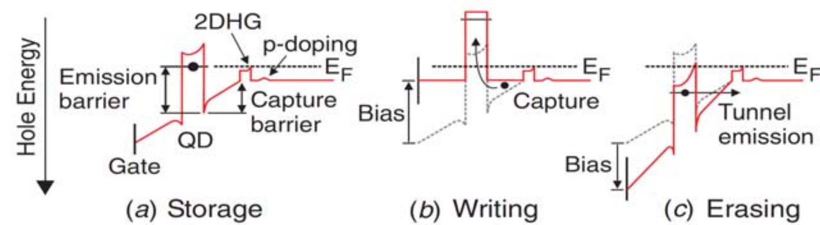

Figure 3. QD-memory device concept. Holes are trapped as a result of a large band-offset between semiconductor materials. The effective size of the barrier may be tuned by the application of a voltage. Readout is via the conductance of a channel that is in close proximity to the charge storage area.

3. Quantum dot memory

One potential contender as a new core technology for charge-based storage of information is quantum dot (QD) memory. QD memory, as conceived in the group of D. Bimberg at TU Berlin, involves the storage of charge in a QD, with readout by a channel which is placed in close proximity. The structure is therefore reminiscent of Flash: the QD plays the role of the floating gate, and is charged and discharged via the channel by applying a voltage to the (control) gate (Fig. 3). QD memory has a number of potential advantages over Flash, some of which are illustrated in Fig. 2. In the first instance, writing to the device involves capture of the charge in the potential well formed by the QD [Fig. 2(b)], which is an intrinsically fast process (a few ps).[13,14] Indeed low-temperature write times of 6 ns for QD memory have already been reported.[15] Erase will still be slow, but unlike write, erase does not have to be done on demand as long as it is recorded which data is redundant (the same system is used in hard disk drives). Carrier capture in QD memory should also result in improved endurance, and better reliability, as should the use of the atomically-perfect interfaces that are readily achievable in III-V heterojunctions. This is in stark contrast with Si/SiO_2 which is intrinsically peppered with defects and dangling bonds. Considering that a single bit may be represented by just 15 electrons, this is a significant point. In addition, the enormous flexibility offered by band engineering in III-Vs that is simply not available in Si/SiO_2 means that it is also possible to generate a RAM using the same Flash-like cell structure (but different architecture). RAM requires that read, write

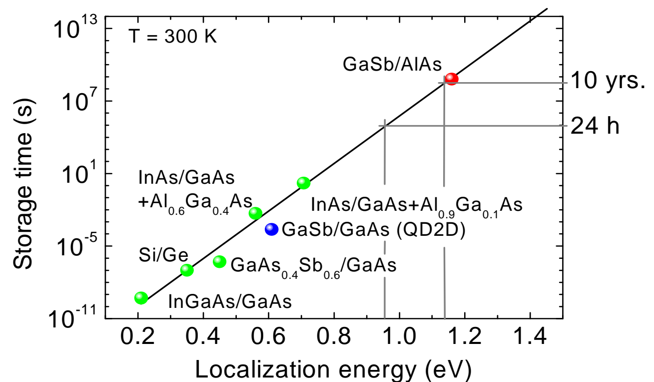

Figure 4. Room-temperature hole storage in QDs. The green dots are values reported in Ref 11. The blue dot (labelled QD2D) is the value for pure GaSb/GaAs QDs determined in the European QD2D project and reported here. The red dot placed at a storage time of about 20 years is a revised projected value for GaSb QDs with AlAs barriers based on the QD2D data point. The solid line has a slope which gives an order of magnitude increase in storage time for each 50 meV of additional localisation energy.

and erase are all performed fast at single-bit level, which means that non-volatility must be compromised to allow fast erase, resulting in a semi-volatile RAM (SV-RAM) [Fig. 2(c)]. Such a device would also have the considerable advantage of non-destructive read.

The most fundamental requirement for a III-V based QD memory is that the (room-temperature) charge storage time in the QD should be long enough to make the device viable, with 'long enough' being dependent on the target memory. Hence to compete with DRAM a data storage time better than the refresh time of 64 ms is required, whereas for non-volatile memory it is 10 years. Note that these figures refer to a chip, not to individual bits, so for statistical reasons a 1 Mb chip requires storage times at bit level that are approximately 1000× longer, taking into account some statistical loss which can be compensated for by standard error correction techniques. Thus for RAM we need single bit storage times of about a minute, and for non-volatile memory single bits should have a (projected) charge storage time of around 10 millennia. Figure 4 shows room-temperature hole storage times for a variety of different III-V QD systems as a function of localisation energy, which is the difference in energy between the QD ground state and the top of the barrier (activation energy). Note that storage of holes rather than electrons has the advantage that their large effective mass increases the density of states, allowing more charge to be stored in given dot. The fundamental building block of the QD telecoms laser, InAs/GaAs, is in the very bottom left-hand corner of the figure, with a carrier retention time of about 1 ns. Substantial gains (9 orders of magnitude) were made by adding Al barriers to the InAs/GaAs system, taking the storage time up to the present record value of 1.6 s.[16] This apparently extraordinary increase in room-temperature storage time is simply the direct result of the fact that it is exponentially dependent on the localisation energy (thermal activation process). This has been borne out by experimental studies of three different materials systems (InAs/$Al_xGa_{1-x}As$, Si/Ge and $GaAs_xSb_{1-x}$/GaAs), as shown in Fig. 4. It should be noted, however, that the storage time refers to a single carrier in the ground state per QD in the ensemble (maximum localisation energy). For a working memory with *N* charges per cell, the quantisation energy and the capacitive charging energy need to be taken into account.

Most promising are type-II systems where the band alignment is such that the difference in band gap between the materials is taken up entirely by either the conduction or valence band, thereby maximising it in a unipolar device, such as a charge-storage memory. In particular, GaSb/GaAs, which is a hole-confining type-II system, has attracted considerable attention. Despite the fact that the band gap of GaSb is twice that of InAs, the type-II band alignment of GaSb/GaAs leads to substantial improvement in localisation energy over InAs/GaAs. Early results with $GaSb_{0.6}As_{0.4}$ QDs[16] gave a factor-of-500 improvement (0.5 μs) over InAs/GaAs,[17] while an extraordinary storage time of 10^6 years was previously predicted for pure GaSb QDs in AlAs,[16] which would be more than sufficient for a non-volatile memory. The main objective of the work reported here was a thorough investigation of the growth, morphological, optical and charge storage properties of self-assembled GaSb/GaAs nanostructures in order to determine their suitability for charge-based memory.

4. Growth by molecular beam epitaxy

Numerous samples were grown by MBE, primarily consisting of a GaAs matrix containing GaSb nanostructures. The basic growth procedure followed to achieve high-quality growth of these structures is detailed in this section, and this is then built upon to realise various different structures in sections 5 to 8. The commonly-used Stranski-Krastanow growth mode is employed for nanostructure formation, but for GaSb on GaAs the lattice mismatch is 7.8% and the strain energy of

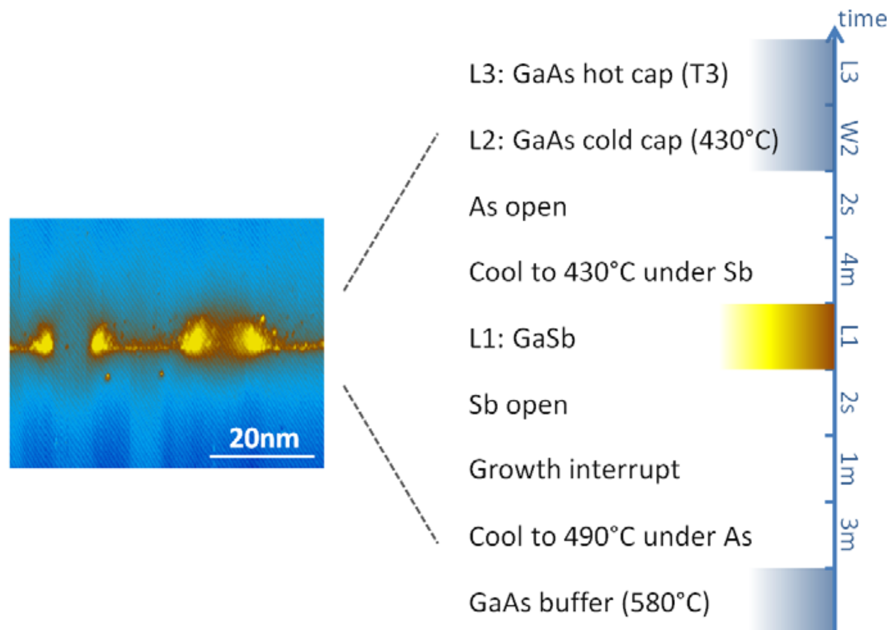

Figure 5. (Right) The molecular beam epitaxy growth procedure followed to form GaSb nanostructures embedded in GaAs. (Left) A false-colour X-STM image illustrating a cut through a pair of quantum rings. The influence of the thickness of the cold cap (W2), and the hot cap temperature (T3) on the morphology of the resulting nanostructures was investigated. The majority of samples contained nanostructures and for these 2.1 MLs of GaSb (L1) were deposited. Some deep-level transient spectroscopy samples containing just a GaSb wetting layer (0.9 MLs of GaSb) were also grown as a control.

an Ga-Sb pair is 1.87 eV; both of these are larger than in the more common InAs/GaAs dot system (7.2% mismatched and 1.40 eV for an In-As pair respectively).

Previous work has demonstrated that MBE growth of self-organised GaSb nanostructures embedded in GaAs tends to result in the formation of quantum rings (QRs), rather than dots.[18] An example of a cross-section through such ring structures is shown on the left-hand side of Figure 5. The final capping (or annealing) temperature has been thought to influence the ratio of major/minor axes of the rings that form[19] (ring formation will be investigated in more detail in section 5). The primary growth procedure followed in this study comprises a layer of GaSb nanostructures formed by the routine illustrated on the right of Figure 5. A GaAs buffer is first grown at high temperature (580°C) on a [001] oriented substrate, followed by the GaSb layer, in which 2.1 monolayers (ML) of GaSb were deposited at a pyrometer-measured temperature of 490°C, with a growth rate of 0.3 MLs^{-1} and a V/III ratio of between 5 and 10. This layer was 'cold' capped[20] with a 4 - 16 nm layer of GaAs at 430°C prior to growing the remaining 'hot' cap at 1 MLs^{-1} at a temperature of 480 - 580°C. The resulting areal number density of nanostructures is ~5 x 10^{10} cm^{-2}.

5. Structural properties

The influence of two growth parameters on the morphology of the GaSb nanostructures was probed by growing a series of samples, each of which contained a number of GaSb layers grown under different conditions. The width, W2, of the cold cap (L2 in Figure 5) was varied from 4 nm to 16 nm, and the temperature, T3, of the GaAs hot cap (L3 in Figure 5) was varied from 480°C to 580°C. X-STM was employed to probe the resulting nanostructures. All X-STM measurements were performed in an Omicron STM1 with a TS-2 scanner, under ultra-high vacuum conditions (pressure ~6 x 10^{-11} mbar) at room temperature. Electrochemically-etched tungsten tips were used. The STM

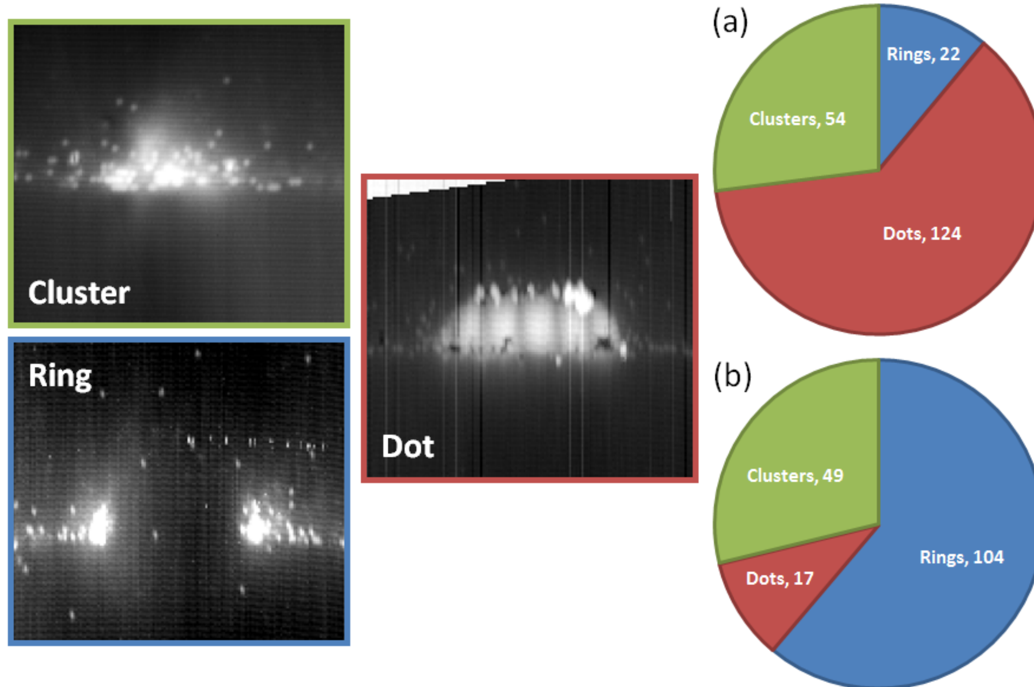

Figure 6. (Left and centre). X-STM images illustrating three different categories of GaSb-rich nanostructures that are found in our studies, each square section has lateral dimensions of 25 by 25 nm. (a) and (b) report statistics of the type of nanostructures that form as the growth temperature of L3 (Figure 2) is varied: in (a) it is 480 to 530°C, and in (b) it is 580°C.

was operated in constant current mode to scan in-situ-cleaved cross-sectional (110) surfaces. The applied bias ranged from –2.5 to –4.0 V, with tunnelling currents between 25 and 50 pA.

The large lattice mismatch between GaSb and GaAs leads to a high pair energy, and strong local strain fields that are relaxed via various mechanisms, producing three distinctly different morphologies. Examples of each of these, labelled clusters, rings and dots, are shown in Figure 6. Clusters have a strongly inhomogeneous Sb concentration, with widely varying morphologies. For some clusters pure areas of GaSb are observed with GaAs in-between. Rings show up as two lobes of a cleaved torus, with a hole in the centre that is often entirely free of Sb; significantly different from InAs/GaAs rings where In material is found inside the ring.[21] Dots have the approximate shape of a truncated pyramid, with a nearly pure GaSb content. The intermixing between As and Sb is substantially less than for GaSb QDs grown by metal-organic chemical vapour deposition (MOCVD)[16] and the In-As intermixing in InAs/GaAs QDs.

The statistics for the three categories of nanostructure, shown on the right hand-side of Figure 6, demonstrate that a small change in the growth conditions can lead to a majority of the nanostructures forming as dots [Figure 6(a)], or as rings [Figure 6(b)]. In this case the growth temperature of the hot cap L3 (T3 in Figure 5) was altered, but a similar shift was induced by varying the thickness of L2.[22]

In previous work we presented evidence to indicate that the evolution of GaSb nanostructures from dots to rings happens after the layer has been deposited and while the nanostructures are capped.[19] During capping, local strain is initially enhanced, as the dots are encapsulated by material with a reduced lattice constant.[23] The strain that has accrued can relax by mass transport: Sb is removed from the dot by processes involving lateral diffusion, As/Sb exchange, and Sb segregation. Diffusion of Sb occurs at regions with the largest local strain, in the centre of a dot, for example. The

instability arises at the position where the equilibrium of the interfacial surface tensions between the substrate, the dot, and the cap layer are disturbed. As/Sb exchange and Sb diffusion empty those regions of Sb, leading to rings with very little intermixing. It is apparent that rings form during the capping process, when there is sufficient energy available to enable mass transport. Their formation is favoured in the presence of a large strain field, with smaller uncapped QDs remaining as dots, often decorated with threading dislocations or evolving into clusters during the capping process.[22]

6. Electrical properties

To directly probe the confinement potential that the GaSb nanostructures present to holes, they were incorporated into device structures [Figure 7(a)] to facilitate deep-level transient spectroscopy (DLTS) studies.[24,25,26] The GaSb-based layers incorporated into these structures were grown either under conditions that favoured dot formation (in order to facilitate comparison with theory), or with a sub-critical thickness of GaSb (L1 from Figure 5 was 0.9 MLs), to leave just a wetting layer (WL) in the active area as a control. A conventional DLTS graph from a sample containing predominantly dots and one containing just a WL is shown in Figure 7(b). To acquire these traces V_{pulse} was set to 0 V and V_r to 3 V, so as to ensure the potential is completely filled and emptied during the process. The peak at ~280 K is a consequence of the discharging of all holes from the QDs during the measurement. This trace was taken with a τ_{ref} of 80 ms and the peak corresponds to an activation energy of 520 meV with an apparent capture cross-section of $\sigma_\infty = 7\text{x}10^{-16}$ cm^2. In contrast, the WL does not show any feature throughout the full temperature range, which indicates that the confinement in the WL is not sufficient to be resolved in the measurement.

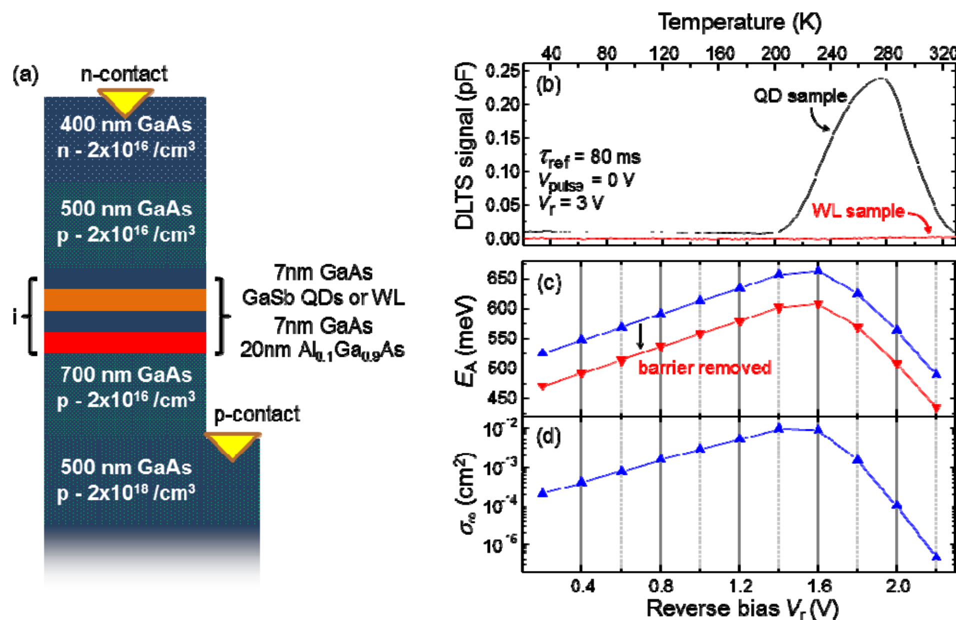

Figure 7. (a) An illustration of the sample structure which was grown for devices fabricated for deep-level transient spectroscopy (DLTS) measurements of either quantum dots (QDs) or the wetting layer (WL). The diameter of the mesa is 400 μm. (b) Conventional DLTS spectrum for a sample containing QDs (black) compared to a sample containing just a WL (red). (c) Activation energies (E_A) plotted as a function of measurement bias, before (blue) and after (red) compensating for the presence of the $Al_{0.1}Ga_{0.9}As$ barrier in the device. (d) The apparent capture cross-section, σ_∞, corresponding to each activation energy.

To make charge-selective[16,17] DLTS measurements of QDs, the measurement bias *V*r is set to different voltages in 0.2 V steps while setting the charging pulse voltage to $V_{pulse} = V_r$ - 0.2 V, so as to induce, on average, the capture or release of just one hole per QD during the pulse. This way the electronic structure of the QD ensemble can be probed, starting with the higher hole levels and proceeding to the deeper hole levels down to the ground state. A consequence of ensemble broadening and many-particle effects are multi-exponential decay transients, and these were analysed using the double-boxcar method.[27] The work cycle was then repeated for different temperatures, resulting in a temperature-dependent curve, with maxima at temperatures where the emission time constant coincides with the reference time constant τ_{ref} of the boxcar window. Tunnelling processes are negligible in this voltage range and the holes trapped in the potential are emitted only due to thermal excitation.[28] An Arrhenius plot of the DLTS peaks for different reference time constants is used to obtain the activation energies for different charging states of the dots. This is plotted as a function of measurement bias V_r in blue in Figure 7(c) alongside its apparent capture cross-section, in Figure 7(d). The device structure, shown in Figure 7(a), contains an $Al_{0.1}Ga_{0.9}As$ barrier that is expected to contribute 54 meV to the measured confinement energies;[29] the red trace in Figure 7(c) compensates for this.

In Figure 7(c), (d), as the reverse bias exceeds 1.6 V, the activation energies and the apparent capture cross sections begin to decrease. This effect is attributed to emission processes from unwanted potentials, such as defect states, or states related to the growth of the dots (perhaps clusters, as shown in Figure 6, or relaxed dots), but not to the emission of holes from the dots themselves. This interpretation is supported by the range in which the apparent capture cross-section for the ground state of dots is typically found ($\sigma_\infty = 10^{-13}$ cm^2 to 10^{-12} cm^2).

The localisation energy, i.e. the activation energy of the ground state, of the dot ensemble measured by charge-selective DLTS here was ~600 meV,[30] this is much larger than previously reported,[17] confirming the growth of nearly pure GaSb nanostructures.

7. Optical properties

For optical studies a simple structure was grown;[31] comprising a single layer of GaSb quantum rings, capped with 250 nm of GaAs. Optical measurements were performed at 1.4 K by placing the sample in a ^{4}He cryostat. A stabilised 532-nm continuous-wave laser optically pumped the GaAs via an optical fibre, and a second fibre collected the emitted photoluminescence (PL) from the quantum ring ensemble and passed it into a grating spectrometer. The excitation spot size on the sample was ~2 mm and the laser power was varied from 2 μW to 500 mW. A typical spectrum showing emission from the ring ensemble, wetting layer and bulk GaAs is shown in Figure 8(a). Emission from the nanostructures is strong when compared to other emitters in the sample, indicating that the rings have good optical properties, and the peak emission energy of the rings is below 1 eV, illustrating the strong confinement potential for this material system. The localisation energy of the electron in the ring system is thought to be small, and therefore the separation between the GaAs bulk peaks and the rings' peak can be used to estimate the hole localisation energy in this system: for example, for the sub-peak at 1340 nm labelled as 2H in Figure 8(b) and (c) we obtain 610 meV, which is comparable with hole localisation energy obtained from the DLTS measurements of the QDs.

Clear oscillations in the emission from the quantum rings can be seen in Figure 8(a). These were analysed by fitting a series of fixed-width Gaussian sub-peaks to the measured spectrum, as shown in Figure 8(b). Next, the excitation power was varied over >5 orders of magnitude, as shown in Figure 8(c), and the fit was successful in closely matching the experimental data throughout this power-range. As the laser power is increased the centre of mass (COM) of the emission from the

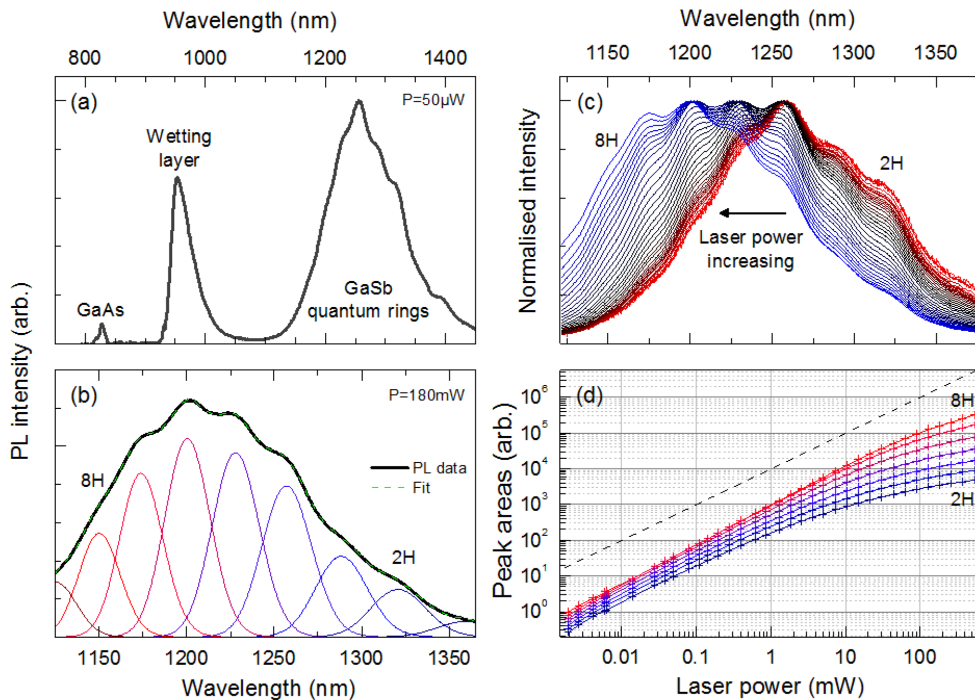

Figure 8. Photoluminescence (PL) from a layer of quantum rings taken at low temperature (1.4 K). (a) A spectrum taken with a wide wavelength range, showing emission peaks from the rings, the wetting layer and the GaAs substrate. (b) Ring peak only, a fit to the data is performed by modelling the substructure as a series of fixed-width Gaussian peaks. (c) Power dependence of the ring peak. In (d) the area of peaks from the fit in (b) are normalised, offset and plotted as a function of laser power. In (b) to (d) 2H and 8H refer to occupancies of 2 and 8 holes respectively.

rings blue-shifts with increasing pump intensity; by more than 50 meV over powers probed here. An excitation-power-induced blueshift is a feature that has been reported in numerous different quantum-confining type-II systems. A bending of the band-structure[32] at the interface between the different carrier types, and capacitive charging by the strongly confined carriers (in this case the holes) have both been suggested as contributing to this shift. In QDs the latter plays a much greater role than the former.[33,34]

We consider several potential causes of the sub-peaks measured in PL:

i. Occupation of excited quantum-confined states[35] in the rings due to state filling. If this were the case then increasingly super-linear behaviour of the peak intensity with power would be expected with increasing energy: The higher-energy states would become rapidly occupied at increasing excitation power. This is not observed. Figure 8(d) plots the normalised, offset, integrated-areas under each of the sub-peaks as a function of power. The dependence of all peaks is remarkably linear, with the lower energy states only beginning to saturate at high powers. Furthermore, occupation of high-energy (high-degeneracy) quantum-confined levels by state filling would require unfeasibly large numbers of confined holes.
ii. Monolayer fluctuations[36] in one dimension of the size of rings, e.g. their heights. This can be excluded on 3 counts. Firstly, if it were the case the structures with the smallest physical size, i.e. largest emission energy, should saturate with increasing pump intensity first. The reverse of this behaviour is seen in Figure 8(d). Secondly, the separation between peaks is almost invariant (21% variation), whereas we would expect it to rapidly increase with decreasing monolayer thickness. For example, for the quantum wires studied in Ref 36 the energy separation between adjacent PL peaks due to monolayer fluctuations changes by a factor of 8. Finally, in contrast to the ribbon-like geometry of the quantum wires, the morphology of the rings (Figures 5 and 6) is not really compatible with a well-defined change in size by 1 ML.
iii. Fabry-Pérot cavity modes. Though it is not readily visible in Figure 8(c), the sub-peak positions display a clear energy dependence as the pump power is increased, and a similar shift is recorded when a magnetic field is introduced in the Faraday geometry. Optical cavity modes could not account for either of these two effects. Also there is no sign of intensity modulation in the WL PL.
iv. Discrete hole charging. A similar, but charge-asymmetrical, version of (i). As the

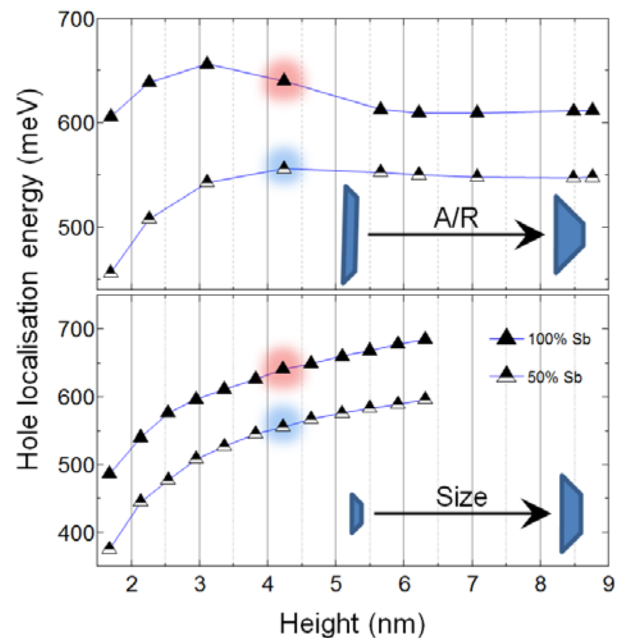

Figure 9: Calculations illustrating the effect on the hole localisation energy of changing the quantum dot size (constant aspect ratio, lower panel), and its aspect ratio at constant volume (upper panel). In both cases two lines are shown, the uppermost corresponding to a dot composed of 100% GaSb, and the lower $GaSb_{0.5}As_{0.5}$. The highlighted points have the same parameters in the top and bottom graphs.

laser power increases the rings fill with holes, but the weak confining potential for electrons prevents further capture of electrons. The emission remains dominated by ground-state recombination, and the increased hole-hole Coulomb charging energy explains the separation between the sub-peaks. A simulation of this effect was probed in Ref 33, though the magnitude of their result was somewhat weaker. Their theoretical study focused on much larger, more dilute GaSbAs nanostructures with a correspondingly smaller charging energy than the ~24 meV peak separation we observe. Indeed, a charging energy of 13 meV was reported in Ref 17 for GaAsSb QDs, which are again more dilute and with a larger volume. Accepting this interpretation of the data, it is now possible to label the sub-peaks with the hole occupancy,[31] as indicated in Figure 8(b)-(d).

The capacitive charging energy determines the amount of charge that can be stored in the dot or ring,[37] so its determination, even by optical methods, is of clear relevance to memories. A capacitive charging energy of 24 meV is uncomfortably high for a charge-based memory: when 10 holes are stored the escape time for one of them will be, on average, about 4 orders of magnitude faster than with single hole occupancy. Indeed the filling up of the rings with holes and its effect on the localisation energy is very graphically illustrated in the PL data of Figure 8(c) and (d).

8. Calculations

To shed light on the relation between the directly-measured morphology of the nanostructures studied in this work and experimental measurements of their confinement, electronic structure modelling was carried out with varying size, shape, and composition of the QDs. Initial parameters were derived directly from X-STM measurements, taking into account the actual morphology and the resulting strain, including the piezoelectric fields, by using a 3D implementation of eight-band k·p envelope function theory, as outlined in Ref. 38. The hole localisation energy is shown as a function of size and aspect ratio for QDs with 100% and 50% GaSb content in Figure 9.

To investigate the influence of the QD height on the localisation energy the QD height was varied between 1.4 and 6.3 nm in our calculations, while keeping the QD shape constant with a height-width aspect ratio of 0.2 (Figure 9, lower panel). In this large range the confinement energy of the pure GaSb dot increases from ~490 meV to ~690 meV with increasing size. The result for $GaSb_{0.5}As_{0.5}$ dots is diminished by roughly 100 meV. This result can be understood in terms of simple quantum confinement, and the smaller valence band offset as the Sb content is reduced.

The capability of X-STM to determine the QD shape is hampered by the random position of the cleave-plane within the quantum dot, which tends to lead to a systematic overestimation of the vertical aspect ratio for truncated pyramidal-shaped QDs as presented in this work. Hence, the aspect ratio of the quantum dot width and height was varied between 0.05 and 0.42, as shown in the upper panel of Figure 9. The volume was kept constant in these calculations to eliminate size quantisation effects. The results show that the hole localisation energy exhibits only a moderate variation between 600-660 meV for GaSb dots, and 450-550 meV for $GaSb_{0.5}As_{0.5}$ dots respectively. The increase of hole localization with decreasing height (from 6 to 3 nm height) for GaSb dots is counterintuitive, and is a result of the redistribution of strain from being predominantly hydrostatic for tall dots, to biaxial for flat structures. For dots with heights less than 3 nm, the height quantisation dominates, leading to a decrease of localisation energy. This effect is more pronounced

for the $GaSb_{0.5}As_{0.5}$ structures because the overall strain is small, as the lattice-mismatch is only half of that for GaSb dots.

9. Discussion

The primary result of this work is a procedure to grow high-quality, near-pure GaSb nanostructures by MBE that provide hole-localisation energies that are very close to their theoretical maximum. Direct electrical (DLTS) measurements of dots are in good agreement with calculations based on structural measurements, confirming the localisation energy as ~600meV, while optical (PL) experiments on rings infer a very similar number. Taking into account the capture cross-section this gives a room temperature charge storage time of the order of 0.2 ms. This is the longest storage time for any of the QD systems studied to date (Fig. 4) in the absence of Al barriers, and is about 1000× longer than the previous result for $GaAs_{0.4}Sb_{0.6}$ QDs grown by MOCVD. This clearly demonstrates the benefit of the growth of GaSb QD/QRs by MBE. Given that the nanostructures we have studied here are pure GaSb, and that the localisation energy is largely determined by the valence band offset and smallest dimension, i.e. the height, further optimisation of their size, shape and composition will provide little improvement in this value. The addition of AlAs barriers is expected to increase the storage time to around 20 years (see Figure 4), meaning that the previous prediction of 10^6 years was somewhat optimistic. This overestimate is a direct result of the commonly-used value for the GaSb/GaAs band offset being too large; instead the value proposed by Wei and Zunger[39] should be used. A consequence of this is that GaSb/$Al_xGa_{1-x}As$ nanostructures are unable to match the requirements for a non-volatile memory.

The storage time in GaSb/AlGaAs is, however, more than sufficient for a RAM. As discussed in section 2, RAM based on QD memory would have the considerable advantages of non-destructive read and semi-volatility. The extent of semi-volatility that can be achieved will be a compromise between the conflicting requirements of long storage time and fast erase, and will need to be the subject of further research. Using the rule-of-thumb of an order of magnitude improvement in storage time for every 50 meV of additional localisation energy, achieving a single bit storage time of about a minute would require $Al_{0.4}Ga_{0.6}As$ barriers, which is a much lower Al content than has already been implemented in a functioning (low-temperature) QD memory cell in the InAs system.[11] Creative design of tunnelling barriers that exploit the extreme flexibility for band engineering in III-Vs may also allow the breaking of the relationship between storage time and erase time, leading to fast memories with storage times in the region of hours or days. Achieving this would be a significant breakthrough. Note that the work presented here has only addressed the fundamental question of charge storage in GaSb nanostructures, a working GaSb/GaAs memory cell, i.e. with a channel to read out the QD/QR has yet to be developed.

Next we turn to the question of rings versus dots. The optical properties of GaSb QRs are significantly improved compared to those of GaSb QDs: the rings provide a degree of electron localisation in the GaAs at their centre,[31] which acts to improve the oscillator strength while still keeping the type-II band alignment. This makes GaSb/GaAs QRs very attractive for optical applications such as telecoms-wavelength lasers and QIP, but is less relevant for memories. Of greater significance for the latter application is that the formation of QRs relaxes the strain, thereby allowing dislocation-free growth. Indeed, the growth of defect-free 10-fold stacks of GaSb/GaAs QRs has recently been demonstrated.[40] The disadvantage of QRs (compared to QDs) is that the reduction in volume due to their geometry increases the capacitive charging energy, and hence the amount of

charge that can be stored: here we report a charging energy of 24 meV for MBE-grown GaSb QRs compared to 13 meV for GaAsSb QDs grown by MOCVD[17,41]. However, given that we are now considering GaSb nanostructures as a basis for SV-RAM, which implies a single bit per cell (programming of multi-bit cells is slow), it may be that only a few holes need to be stored, in which case capacitive charging energy becomes less important. The answer to this should be clear with the implementation of a working GaSb-QR memory cell, as discussed in the preceding paragraph.

Beyond this, two further significant challenges remain, ordering of QDs and implementation on Si. Fabrication of an entire memory chip on GaAs is conceivable and may also be commercially viable if the inevitable additional cost of such a product offered substantial other advantages, such as semi-volatility. Nevertheless, implementation on Si offers lower cost and additional possibilities such as integration of QD memory on a processor chip. The other issue is that QD/QRs grown by the Stranski-Krastonow method are randomly distributed across the sample surface, which is incompatible with top-down processing. It should be noted that the dimensions of self-assembled QD/QRs are comparable with minimum feature sizes in Si processing, so realistically we need to achieve exactly one QD/QR per memory cell. It possible that both ordering and implementation can be resolved simultaneously using selective area growth (SAG). This is already a very active area of (industrial) research for the implementation of III-Vs with Si-CMOS technology.[42]

10. Conclusions

We have presented a comprehensive investigation of the structural, optical and electronic properties of self-assembled GaSb nanostructures in GaAs grown by molecular beam epitaxy (MBE), with a view to assessing their suitability for charge based memories. MBE growth tends to favour the formation of quantum rings with a remarkably pure GaAs core, but with careful control of the growth conditions it is also possible to grow quantum dots. Structural investigations reveal that for both dots and rings there is very little intermixing, i.e. the nanostructures are nearly pure GaSb. The formation of quantum rings relaxes the strain, allowing the growth of dislocation-free material and rings also have intrinsically better optical properties. Electrical, optical and theoretical studies all confirm that the ground-state localisation energy for holes is a little over 600 meV, which corresponds to a storage time for a single hole of about 0.2 ms in GaSb/GaAs, and around 20 years in GaSb/AlAs. This is insufficient for non-volatile memory, but very promising for a RAM. Compared with incumbent Si-based DRAM technology GaSb-based RAM would have the dual advantages of non-destructive read and semi-volatility, making it clearly deserving of further research.

11. Acknowledgements

The authors are indebted to T. Mikolajick (NaMLab, TU Dresden) and M. van Duuren (NXP) for illuminating discussions on memories over several years. The work reported here was primarily supported by the QD2D project (www.QD2D.eu) with a grant from the European Commission in the framework of NanoSci-E+, regionally this was managed by the EPSRC (EP/H006419), FOM and DFG (BI284/30-1), and the BMBF in the framework of the VIP program HOFUS. R. J. Y. and M.H. would both like to thank the Royal Society for support, as a Research Fellow and through a Brian Mercer Feasibility Award, respectively, and P.H. acknowledges support of QinetiQ (Agreement No: 3000127730). Additional support comes from DFG Contract No. BI284/29-1.

[1] D. Bimberg, M. Grundmann, and N. Ledentsov, *Quantum Dot Heterostructures* (Wiley, New York, 1999).

[2] A. J. Shields, Nature Photon. **1**, 215 (2007)
[3] Y. Arakawa, and H. Sakaki, Appl. Phys. Lett. **40**, 939 (1982)
[4] A. Luque and A. Mati, Phys. Rev. Lett. **78**, 5014 (1997).
[5] J. R. Gell, M. B. Ward, R. J. Young, R. M. Stevenson, P. Atkinson, D. Anderson, G. A. C. Jones, D. A. Ritchie, and A. J. Shields, Appl. Phys. Let. **93**, 081115 (2008)
[6] A. J. Shields, M. P. O'Sullivan, I. Farrer, D. A. Ritchie, R. A. Hogg, M. L. Leadbeater, C. E. Norman, and M. Pepper, Appl. Phys. Lett. **76**, 3673 (2000)
[7] R. M. Stevenson, A. J. Hudson, R. J. Young, P. Atkinson, K. Cooper, D. A. Ritchie, and A. J. Shields, Optics Express **15**, 6507 (2007)
[8] R. J. Young, S. J. Dewhurst, R. M. Stevenson, P. Atkinson, A. J. Bennett, M. B. Ward, K. Cooper, D. A. Ritchie, and A. J. Shields, New J. Phys. **9**, 365 (2007)
[9] M. Hayne, J. Maes, S. Bersier, V. V. Moshchalkov, A. Schliwa, L. Müller-Kirsch, C. Kapteyn, R. Heitz, and D. Bimberg, Appl. Phys. Lett. **82**, 4355 (2003).
[10] A. Marent, M. Geller, and D. Bimberg, Microelectr. J. **40**, 492 (2009)
[11] A. Marent, T. Nowozin, M. Geller, and D. Bimberg, Semicond. Sci. Tech. **26**, 014026 (2011)
[12] International Technoogy Roadmap for Semicondcutors 2012 Update, http://www.itrs.net/Links/2012ITRS/Home2012.htm, accessed 26/03/2012.
[13] T. Müller, F.F. Schrey, G. Strasser, K. Unterrainer, APL **83** 3572-3574 (2003)
[14] M.Geller, A. Marent, E. Stock, D. Bimberg, V. I. Zubkov, I. S. Shulgunova, and A. V. Solomonov, APL **89** 232105 (2006)
[15] M. Geller, A. Marent, T. Nowozin, D. Bimberg, N. Akçay, and N. Öncan, Appl. Phys. Lett. **92**, 092108 (2008)
[16] A. Marent, M. Geller, A. Schliwa, D. Feise, K. Pötschke, D. Bimberg, N. Akcay, and N. Öncan, Appl. Phys. Lett. **91**, 242109 (2007)
[17] M. Geller, C. Kapteyn, L.Müller-Kirsch, R. Heitz, and D. Bimberg, Appl. Phys. Lett. **82**, 2706 (2003).
[18] R. Timm, H. Eisele, A. Lenz, L. Ivanova, G. Balakrishnan, D. L. Huffaker, and M. Dähne, Phys. Rev. Lett. **101**, 256101 (2008)
[19] M. Ahmad Kamarudin, M. Hayne, R. J. Young, Q. D. Zhuang, T. Ben, and S. I. Molina, Phys. Rev. B **83**, 115311 (2011).
[20] M. Ahmad Kamarudin, M. Hayne, Q. D. Zhuang, O. Kolosov, T. Nuytten, V. V. Moshchalkov, and F. Dinelli, J. Phys. D: Appl. Phys. **43**, 065402 (2010).
[21] P. Offermans, P. M. Koenraad, J. H. Wolter, D. Granados, J. M. Garcı´a, V. M. Fomin, V. N. Gladilin, and J. T. Devreese, Appl. Phys. Lett. **87**, 131902 (2005)
[22] E. P. Smakman, J. K. Garleff, R. J. Young, M. Hayne, P. Rambabu, and P. M. Koenraad, Appl. Phys. Lett. **100**, 142116 (2012)
[23] R. Blossey and A. Lorke, Phys. Rev. E **65**, 021603 (2002)
[24] D. V. Lang, J. Appl. Phys. **45**, 3023 (1974)
[25] R. Timm, H. Eisele, A. Lenz, S. K. Becker, J. Grabowski, T. Y. Kim, L. Müller-Kirsch, K. Pötschke, U. W. Pohl, D. Bimberg, and M. Dähne, Appl. Phys. Lett. **85**, 5890 (2004)
[26] R. Magno, B. R. Bennett, and E. R. Glaser, J. Appl. Phys. **88**, 5843 (2000)
[27] D. S. Day, M. Y. Tsai, B. G. Streetman, and D. V. Lang, J. Appl. Phys. **50**, 5093 (1979)
[28] T. Nowozin, A. Marent, M. Geller, D. Bimberg, N. Akçay, and N. Öncan, APL 94, 042108 (2009)
[29] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J.Appl. Phys. **89**, 5815 (2001)
[30] T. Nowozin, A. Marent, L. Bonato, A. Schliwa, D. Bimberg, E.P. Smakman, J. Garleff, P.M. Koenraad, R.J. Young, and M. Hayne, Phys. Rev. B **86**, 035305 (2012)
[31] R. J. Young, E. P. Smakman, A. M. Sanchez, P. Hodgson, P. M. Koenraad, and M. Hayne, Appl. Phys. Lett. **100**, 082104 (2012)
[32] N. N. Ledentsov, J. Böhrer, M. Beer, F. Heinrichsdorff, M. Grundmann, D. Bimberg, S. V. Ivanov, B. Ya. Meltser, S. V. Shaposhnikov, I. N. Yassievich, N. N. Faleev, P. S. Kop'ev, Z. I. Alferov, Phys. Rev. B **52**, 14058 (1995)
[33] K. Gradkowski, T. J. Ochalski, D. P. Williams, S. B. Healy, J. Tatebayashi, G. Balakrishnan, E. P. O'Reilly, G. Huyet, and D. L. Huffaker, Phys. Status Solidi B **246**, 752 (2009)
[34] P. D. Hodgson, R. J. Young, M. Ahmad Kamarudin, P. J. Carrington, A. Krier, Q. D. Zhuang, E. P. Smakman, P. M. Koenraad and M. Hayne, to be published.
[35] K. Gradkowski, N. Pavarelli, T. J. Ochalski, D. P. Williams, J. Tatebayashi, G. Huyet, E. P. O'Reilly, and D. L. Huffaker, Appl. Phys. Lett. **95**, 061102 (2009)

[36] J. Maes, M. Hayne, Y. Sidor, B. Partoens, F. M. Peeters, Y. González, L. González, D. Fuster, J. M. Garcia, and V. V. Moshchalkov, Phys. Rev. B **70**, 155311 (2004).
[37] Quantum confinement is a less important factor because of the degeneracy of higher quantum states.
[38] A. Schliwa, M. Winkelnkemper, and D. Bimberg, Phys. Rev. B **76**, 205324 (2007)
[39] S.-H. Wei and A. Zunger, Appl. Phys. Lett. **72**, 2011 (1998)
[40] P. J. Carrington, R. J. Young, P. D. Hodgson, A. M. Sanchez, M. Hayne, and A. Krier, Cryst. Growth Des. **13**, 1226 (2013).
[41] Note that this does not imply that GaAsSb QDs grown by MOCVD would be better for memories than MBE-grown QRs, since the increased band offset in the pure QRs more than compensates for the larger charging energy.
[42] M. Paladugu, C. Merckling, R. Loo, O. Richard, H. Bender, J. Dekoster, W. Vandervorst, M. Caymax, and M. Heyns, Cryst. Growth Des. **10**, 4696 (2012).