Ultra-low-power non-volatile RAM for the Internet of Things (ULTRARAM)

Manus Hayne,[1*] Peter Hodgson,[1] Dominic Lane,[1] Peter Carrington,[2] Evangelia Delli[2] and J. Iwan Davies[3]

[1]Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom; [2]Department of Engineering, Lancaster University, Lancaster LA1 4YW, United Kingdom; [3]IQE plc, Pascal Close, Cardiff CF3 0LW, United Kingdom
*Corresponding author: m.hayne@lancaster.ac.uk

ABSTRACT

ULTRARAM™ is a novel memory technology that potentially combines the non-volatility of flash with performance that exceeds DRAM. We review the motivation for an ultra-efficient non-volatile memory for autonomous Internet of Things sensors and the state of the art in memory technology. We describe progress in ATTRACT Phase 1 in second generation devices, arrays and growth on Si substrates. For ATTRACT Phase 2 we propose a dramatic advance in the technology to ≥100 Mbit chips and three technology demonstrators. This ambitious goal is based on the rapid progress to date, on-going funding and strong industrial interest in the technology.

Keywords: Non-volatile memory; Internet of Things; sensors.

1. INTRODUCTION

The Internet of Things (IoT) is widely predicted to cause a profound technological and societal transformation, on a par with the technological innovations of microelectronics and the internet that will underpin it. The number of IoT devices increased 31% year-on-year to reach 8.4 billion by 2017, and it is estimated that there will be in excess of 20 billion such devices by 2020. The IoT, along with artificial intelligence and automated (robotic) production are expected to add $3.75 trillion to the world economy by 2025, whilst at the same time addressing the UN Sustainable Development Goals. However, there are considerable technological, societal and ethical challenges associated with the IoT. Examples include communication bandwidth (capacity), privacy and data security. Other, more fundamental and practical challenges exist at the level of the individual sensor and associated hardware embedded in the 'Things'. In particular, many of these sensors may be electrically isolated (autonomous), either having their own power source (battery) and/or be required to harvest or scavenge energy from their environment, for example from solar, wind, thermal or vibrational sources. Harvesting provides very little energy, so such autonomous sensors are required to have the lowest possible power consumption. They may even need to remain in a quiescent state slowly collecting or storing data for extended periods, before bursting into life when sufficient energy is available to connect to the outside world. A key component in achieving this is the devices' memory. Such a memory should work at low voltages, use as little energy as possible when being programmed, and be capable of robustly storing the data almost indefinitely with no power.

In the Phase 1 ATTRACT project, we advanced the development of a potentially ground-breaking type of novel compound-semiconductor memory, ULTRARAM™, for use in autonomous IoT sensors and other applications. ULTRARAM™ is predicted to have an intrinsic storage time that exceeds the age of the Universe [1], a single bit switching energy (per unit area) that is orders of magnitude lower than any conventional or emerging memory [2] and ultra-fast programming speeds [3]. ULTRARAM™ should be compatible with silicon or compound semiconductor platforms, allowing it to be integrated into a broad range of sensor types.

We have determined the changes needed in the design of previous first generation (Gen I) devices, allowing them to be connected in fully bit-addressable arrays in a compact design with the lowest normalised cell size of any memory technology, and have fabricated 4-bit arrays ready for testing. Substantial advances in growth on Si have been made, with material quality that exceeds reference samples grown on GaAs substrates.

2. STATE OF THE ART

The memory market is dominated by dynamic random access memory (DRAM, ~$70bn pa) and NAND flash (~$50bn pa). DRAM's high speed (compared to Flash) has made it the main system memory (RAM – random access memory) in computers and electronic devices. However, it has the significant drawbacks of volatility

Tab. 1. Benchmarking of ULTRARAM™ against conventional and emerging memory technologies [4, 5]. F is the feature size (node). The switching time for ULTRARAM™ is simulated.

Technology	NV?	Switching energy	Switching time	Cell size
SRAM	No	1 fJ	~1 ns	>100F^2
DRAM	No	1-10 fJ	<10 ns	6F^2
Flash (NOR)	Yes	~100 pJ	0.01-100 ms	10F^2
Flash (3D NAND)	Yes	~10 fJ	0.1-100 ms	4F^2
RRAM	Yes	0.1-10 pJ	10-100 ns	(4 to 12)F^2
PCRAM	Yes	10-100 pJ	50-400 ns	(4 to 30)F^2
STT-MRAM	Yes	~0.1 pJ	10-50 ns	(6 to 50)F^2
ULTRARAM™	Yes	~10 aJ	<10 ns	<4F^2

and destructive read of data: it requires a continuous source of power otherwise data is lost. Flash, is non-volatile and low cost, making it suited to storing (large amounts) of data, but the voltages needed to program and erase data make it intrinsically slow, and damage the cells, limiting the number of program/erase cycles (endurance). Thus, none of the conventional memories are well-suited for autonomous IoT sensors. Alternatives to these conventional memories, so-called emerging non-volatile memories (NVM) such as resistive RAM (RRAM), phase change RAM (PCRAM) and spin-torque-transfer magnetic RAM (STT-MRAM) have been the subject of intense research for decades, with the ambition to combine the advantages of DRAM and flash, with none of their disadvantages, an ideal often referred to as 'universal memory'. With characteristics that include non-volatility, ultra-low energy consumption and high endurance, universal memory represents a 'holy grail' for IoT sensors, as well as for other applications. However, to date these emerging NVM have fallen short due to one or more of: limited endurance, inferior speed, large leakage currents and switching energies that are orders of magnitude higher than DRAM [4, 5].

3. BREAKTHROUGH CHARACTER OF THE PROJECT

ULTRARAM™ is a breakthrough memory technology with a set of characteristics that match or outperform other conventional and emerging memories (Table 1). Characteristics of the conventional memories are shown in the upper part of Table 1. Static RAM (SRAM) and DRAM are the highest performing conventional memories both in terms of speed and switching energy, but are volatile, making them unsuitable for data storage in IoT sensors. NAND flash is non-volatile, but is only suitable for data storage, so needs to be implemented with RAM to run any code. A typical combination for most computing applications is thus SRAM plus DRAM plus NAND Flash, however, this is overly complicated and expensive for low-cost applications like IoT sensors.

A widely-used alternative for such applications is a lesser-known flash variant, NOR flash, which can be used to run code and store small amounts of data as embedded flash, removing the need for DRAM. However, NOR has large switching energies and is slow. The second class of memories in Table 1 are the emerging memories. These substantially outperform (NAND/NOR) flash in terms of switching speed and have lower switching energies than NOR flash, making them attractive for IoT sensors and other embedded applications. Nevertheless, they fall far short of the performance of DRAM, leaving the field open for a disruptive memory technology that combes the non-volatility of flash and the emerging memories with the performance of DRAM (universal memory).

ULTRARAM™ is such a technology. Based on a compact flash-like patented concept [6] in which charge is stored in an electrically-isolated floating gate, it offers the "almost impossible" [4] combination of the non-volatility of flash, with the fast, low-energy switching of DRAM and SRAM. This is achieved by the use of a triple-barrier resonant-tunnelling (TBRT) structure (Fig. 1) that switches from opaque, in the absence of applied bias, to transparent with a gate voltage of ≤2.5 V [2, 3], four to ten times lower than required for flash. The extraordinary properties of the InAs/AlSb TBRT promises intrinsic (thermal) storage times that exceed the age of the Universe [6], in combination with peak program/erase currents that deliver ultra-fast switching times, in a low capacitance device with ultra-low switching energies, and low disturb [3].

4. PROJECT RESULTS

The ATTRACT Phase 1 project was concerned with the development of 4-bit ULTRARAM™ arrays and implementation on Si. The main results in these areas are described below.

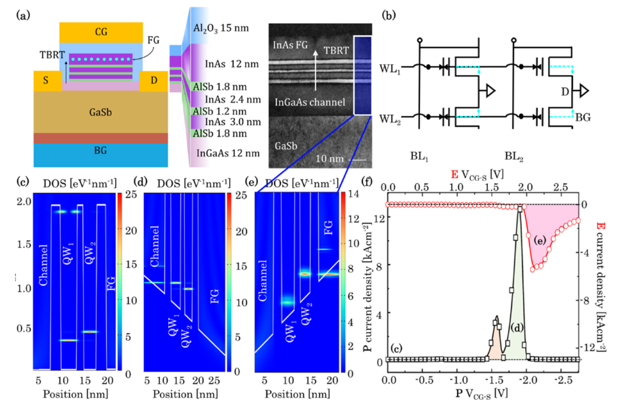

Fig. 1. (a) Second generation (Gen II) ULTRARAM device structure; (b) basis of architecture with two pairs of devices with drain contacts shorted to common back gate (BG); simulations of TBRT - (c) store, (d) program, (e) erase and (f) program/erase current density versus applied bias.

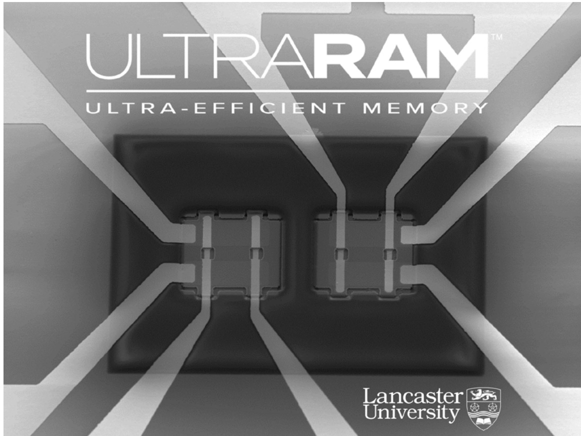

Fig. 2. Two 2×2 bit ULTRARAM™ arrays.

4.1. Second generation devices and arrays

First generation (Gen I) ULTRARAM™ devices reported in Ref. [4] demonstrated room-temperature operation of non-volatile memories with low voltage (2.5 V) program and erase, and non-destructive read. However, these devices suffered from poor 0/1 contrast of ~10%, and had channels that were always conductive, inhibiting their use in arrays. Both effects were caused by a combination of the use of a wide InAs n-doped channel between source and drain, and a large hole leakage current, caused by a relatively featureless valence band across the structure [2].
During the ATTRACT Phase 1 project we have made substantial progress in the demonstration of Gen II devices in which these issues are fully resolved by reversing the device (Fig. 1(a)) such that charge is transferred into (out of) the floating gate from (to) the channel. The channel itself is formed by a narrow (Ga)InAs quantum well that is non-conductive in the absence of an applied gate bias of a magnitude that depends on the occupancy of the floating gate [3, 7]. This scheme (Fig. 1(b)) allows the drain contacts of pairs of devices to be grounded to the back gate, resulting in an efficient design that only requires source and control gate contacts connected as bitline (BL) and wordline (WL), respectively (Fig. 2) [3, 7]. Indeed, we have recently designed a 64-bit array with a record-breaking normalised cell size of $3.4F^2$.

4.2. Growth on Si

A vital step in the roadmap towards cost-competitive ULTRARAM™ memories is their integration onto large Si wafers. However, the heteroepitaxial growth of III-V materials on Si is a non-trivial task due to several material incompatibilities, such as differences in lattice constants, thermal expansion coefficients and the differing polar/non-polar nature of the crystals. Nevertheless, we have made significant progress towards this goal, and have produced III-V on Si material with quality approaching, and in some cases exceeding, that of previous samples grown on GaAs substrates [2].
Growth of III-Vs was by molecular beam epitaxy (MBE) on a Veeco GENxplor system. Samples were grown on two distinct types of substrate: (A) Ge/Si wafers with a 4° off-cut towards [110], and (B) Si wafers with a 4° off-cut towards the [0-11] direction. The former substrates were supplied by IQE plc, with the Ge layers grown by chemical vapour deposition. For both substrate types, the initial aim was to achieve a high quality, monocrystalline GaSb surface on which to deposit the layers of the memory structure. This was done using a variety of III-V buffer structures and multiple growth techniques, including migration-enhanced epitaxy, interface misfit epitaxy and 2-step temperature MBE growth. As can be seen in Table 2, the (A) buffer layer sample's surface has a higher defect density and roughness than the GaSb/GaAs control sample and requires further optimisation. However, the (B) buffer layer sample is of extremely high quality, exhibiting a defect density lower than the control and, therefore, was chosen for the implementation of ULTRARAM™ on Si. To this end, initial progress is promising, with the successful inclusion of an InAs back-gate into the buffer layers and an InGaAs channel on top (Table II). Work is ongoing to grow, fabricate and test a full ULTRARAM™ structure on Si.

Tab. 2. Comparison of material quality for GaSb (and successive layers) grown on GaAs and on Si.

Layers grown	Substrate	Surface defect density	RMS surface roughness
GaSb/GaAs	GaAs	5.6×10^{8} cm^{-2}	1.1 nm
GaSb/GaAs/Ge/Si	(A)	7.8×10^{9} cm^{-2}	3.5 nm
GaSb/AlSb/Si	(B)	2.1×10^{8} cm^{-2}	1.8 nm
InGaAs/GaSb/AlSb/Si*	(B)	2.0×10^{8} cm^{-2}	1.7 nm

*GaSb layer includes 50 nm AlSb/ 50 nm InAs back gate layers

5. FUTURE PROJECT VISION

5.1. Technology Scaling

The current status of the technology is room-temperature operation of single bit Gen I ULTRARAM™ devices demonstrating non-volatile data storage with low-voltage, low-energy program/erase and non-destructive read. Gen II devices are now fully simulated, with single devices and 4-bit arrays fabricated and ready for testing, and a highly compact ($3.4F^2$) 64-bit array designed. Significant progress has been made in implementation on Si substrates. All work to date has been at the $F = 10$ µm node.
The objective of a Phase 2 ATTRACT project would be fully complete and qualified ULTRARAM™ memory chips (TRL 8), requiring (i) massive up-scaling of arrays

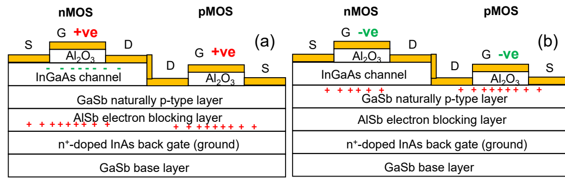

Fig. 3. Basis of proposed III-V CMOS logic with (a) positive, and (b) negative, gate bias.

from 4-bit to ≥100 Mbit, (ii) full integration of novel III-V CMOS logic for addressing, (iii) validated industrial process for III-V growth on Si substrates (at least 6”), and (iv) scaling of feature size (node) F to ≤100 nm (ideally 20 nm). This ambitious target takes into account progress made between ATTRACT Phase 1 and Phase 2 using secured and pending funding. (i) will be achieved by going from 4 to 64 bits with external addressing (in progress), then to 1 Mbit with integrated addressing, which is the actual size of cell arrays in DRAM. These 1-Mbit blocks will be repeated to generate chips ≥100 Mbit. The integrated III-V CMOS addressing logic (ii) will be implemented using the scheme summarised in Fig. 3, utilising layers that are already part of ULTRARAM™. (iii) will build on work described in section 4.2 (on-going) and scaling for (iv) is due to start soon.

5.2. Project Synergies and Outreach

ULTRARAM™’s unprecedented characteristics and the exceptionally rapid progress in development compared with emerging memories has already secured substantial interest from a number of organisations covering all competencies required to reach the end-goals of the ATTRACT Phase 2 project. In addition to ATTRACT Phase 1 partners Lancaster University and IQE, these include Warwick University (electron microscopy) IMEC (microelectronics research), Aixtron (epitaxy equipment manufacturer), IBM (research & end-user), Huawei (research & end-user), CML Microsystems (chip design), British Telecom (BT, end user). We will actively seek out further potential partners in order to maximise the chances of a successful bid and research outcome, and will cluster with other ATTRACT Phase 1 funded projects, such as those concerned with electronics or IoT (*e.g.* DIBIS, ECOTAGS and SPACC) for technological demonstration (see section 5.3).

ULTRARAM™ technology has already generated huge amounts of public interest globally. Highlights include more than 250 on-line articles with a combined reach of >78 million people, a piece on BBC World Service’s Digital Planet, front cover of Electronics Weekly and a two-page feature article in the leading UK high-street magazine PC Gamer. We will continue to issue press releases, directly contact journalists and conduct media interviews during ATTRACT Phase 2. We have also developed a portable demo kit, intended to be used at the ATTRACT Final Conference, which will be supplemented with pull-up banners and glossy leaflets for public outreach events.

5.3. Technology application and demonstration cases

ULTRARAM™ has the potential to become a core underpinning information and communication technology, with implementation in devices that range from embedded memory in (autonomous IoT) sensors and appliances, smartphones, laptops and personal computers through to large-scale computing infrastructure. Details are yet to be finalised, but a likely approach will be technology demonstrators at three scales; (IoT) sensor, smartphone and large-scale computing. The first of these would be in cooperation with an ATTRACT Phase 1 partner, such as those mentioned in section 5.2. The smartphone demonstrator would involve the replacement of the DRAM and flash chips with a single ULTRARAM™ chip, with Huawei as a potential partner, and the large-scale computing demonstrator would involve simulating the implementation of ULTRARAM™ in the computing infrastructure of an industrial end-user, *e.g.* BT, or a large-scale Research Infrastructure, *e.g.* CERN. These demonstrators have the potential to bring concrete benefit to Societal Challenges, for example via the ubiquitous IoT, which is recognised as being able to address the UN Sustainable Development Goals. However, it is ultra-efficiency that is at the core of ULTRARAM™, and this will be most dramatically demonstrated in the context of large-scale computing infrastructure: ever-increasing demand for computing and data, including the IoT, risks drowning the planet in ‘tsumani of data’ that could consume 20% of global electricity by 2025.

5.4. Technology commercialization

Commercialisation of ULTRARAM™ will require large-volume, low-cost manufacture so that it can compete economically, as well as technically, with the conventional (Si) and emerging memories. The technology scaling that we have proposed for the Phase 2 ATTRACT project is an ambitious goal that, if successful, will take it to TRL 8. At this point its technological performance can be fully evaluated at chip level, and directly compared to existing products. We anticipate that ULTRARAM™’s performance will be at least as good as DRAM, but with the additional advantages of non-volatility and ultra-low energy switching, in which case manufacturing cost will be a primary concern. We have good reasons to believe that cost/bit will actually be lower for ULTRARAM™ than for DRAM. The first reason is associated with the economics of Si chip manufacture, almost all of which is in the microelectronics fab. This explains why 300 mm substrates are used (to process as much material in parallel as possible), and why most companies outsource

production to specialist manufacturers such as TSMC. The unit area cost of the substrate is only a very minor factor. ULTRARAM™ incorporates all of the active materials properties into the epitaxial growth as a single technological step, and requires no lateral changes in active material properties, such as the n-p-n and p-n-p junctions needed in MOSFETs. This makes processing vey considerably simpler, and cheaper. Secondly, its compact form should reduce the cost/bit. Finally, its non-volatile nature will eliminate the need for DRAM's additional electronics for refresh and to compensate for destructive read. A study of ULTRARAM™'s expected manufacturing cost will be incorporated into the Phase 2 project. Once proven technologically and economically viable, commercialisation largely depends on what happens with the intellectual property (IP) [6, 7]. The main options are licencing or assigning the IP to an established global player or generating a start-up as a vehicle for commercialisation. Given the industrial interest in the technology at this early stage, and that we have been contacted by three different investment companies about it, both of these routes should be available.

5.5. Envisioned risks

The main risks are scientific/technological and limited lab access due to Covid-19, both of which will have the effect of limiting progress within the timescale of the project. The main mitigation will be to reduce the interdependency of different technological aspects (scaling, growth on Si, III-V CMOS). If scaling <100 nm proves challenging, then work will continue at that node: a 100 Mbit chip at the 100 nm node with full addressing logic will only occupy ~4 mm^2. If progress with growth on Si is slow, research can continue using GaAs substrates, and if III-V CMOS logic has issues, chip bonding memory arrays with Si CMOS will be explored. Reduced access to laboratories due to Covid-19 will be mitigated by either re-assigning the work to a different partner in the consortium, and/or by prioritising work that can be most efficiently executed with the facilities that are available. A comprehensive risk register will be produced to govern the risk assessment and mitigation process, and will be regularly updated.

5.6. Liaison with Student Teams and Socio-Economic Study

The consideration of the potential societal impact of ULTRARAM™ is already fully embedded in our research programme, although not via the route provided by ATTRACT Phase 1. We have a formal and on-going collaboration with the Institute for Social Futures, Lancaster University, and have secured funding from the Materials Social Futures doctoral training programme for a PhD student who will receive social science training alongside his physics-based PhD research on III-V CMOS for ULTRARAM™. We are thus very well placed to incorporate socio-economic studies into an ATTRACT Phase 2 project, with direct access to the expertise required, and consider it an important part of our on-going activity.

6. ACKNOWLEDGEMENT

This work was supported by the Engineering and Physical Sciences Research Council, UK, in the framework of the 2017-2020 IAA Funding Allocation to Lancaster University under grant EP/R511560/1, a scholarship under grant EP/N509504/1, the Future Compound Semiconductor Manufacturing Hub grant EP/P006973/, by the ATTRACT project funded by the EC under Grant Agreement 777222 and by the Joy Welch Educational Charitable Trust.

7. REFERENCES

[1] T. Nowozin, T., Bimberg, D., Daqrouq,K., Ajour, M. N., & Awedh, M., 2013. Materials for Future Quantum Dot-Based Memories, J. Nanomaterials, 215613.

[2] Tizno, O., Marshall, A.R.J., Fernández-Delgado, N., Herrer, M., Molina S.I., & Hayne, M., 2019. Room-temperature Operation of Low-voltage, Non-volatile, Compound-semiconductor Memory Cells, Scientific Reports 9, 8950.

[3] Lane, D., & Hayne, M., 2020. Simulations of Ultralow-Power Nonvolatile Cells for Random-Access Memory, IEEE Trans. Electron Devices 67, pp 474-480.

[4] Yu, S., & Chen, P.-Y., 2016. Emerging Memories Technologies – Recent Trends and Prospects, IEEE Solid-State Circuits Magazine (Spring issue) 2016, pp. 43-56.

[5] Prall, K., 2017, Benchmarking and Metrics for Emerging Memories, 2017 IEEE International Memory Workshop (IMW), Monterey, 2017, pp. 1-5.

[6] Hayne, M., 2019. Electronic Memory Devices, US10243086B2.

[7] Hayne, M. & Lane, D. 2019. International (PCT) patent pending (unpublished).