ULTRARAM: A Compound-Semiconductor Floating-Gate Universal Memory

X. Xia[1], S. Tekin[1], P. D. Hodgson[1,2] and M. Hayne[1,2]

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

[2]*Quinas Technology*, *Department of Physics, Lancaster University, Lancaster LA1 4YB, United Kingdom*

m.hayne@lancaster.ac.uk; m.hayne@quinas.tech

The ~$160bn per year memory chip market is dominated (>90%) by dynamic random access (DRAM) and flash. Flash is intrinsically slow and has poor program/erase cycling endurance, but is non-volatile and very cheap, so well suited for data storage. DRAM is relatively fast, with low switching energies and excellent endurance, so is the main (working) memory in computing systems. However, because the charge representing data in each DRAM bit is stored in a capacitor, it is non-volatile, reading the bit removes the charge (destructive read), it needs constant refreshing and it cannot be scaled beyond 10-nm node, worsening the memory bottleneck. All of these generate significant inefficiencies. The deficiencies of both DRAM and flash have stimulated intense activity into alternative memory technologies, so called 'emerging memories', with the objective of combining the non-volatility of flash with the speed and endurance of DRAM, a so-called 'universal memory'. Unlike DRAM and flash, which use charge to store data, most emerging memories use the movement of atoms or ions to change resistance. Whilst this delivers non-volatility, emerging memories have struggled to match the performance of DRAM or the cost of flash.

ULTRA**RAM** [1] is a charge-storage emerging memory. In essence, it is a compound semiconductor version of flash where the oxide barrier between the channel and floating gate is replaced by a triple-barrier resonant tunnelling structure (TBRT) [2], nominally consisting of two ultrathin InAs quantum wells of unequal thickness (2.4 and 3.0 nm) bounded by even thinner AlSb barriers (1.8, 1.2 and 1.8 nm). In the absence of control gate bias, charge is robustly stored in the floating gate with a retention time of at least 1,000 years [3]. Inherently fast (ns) [4] program and erase proceeds via the use of resonant tunnelling on application of just ±2.5 V between control gate and channel, compared to μs for flash at 10 to 20 V. Its low capacitance and low program/erase voltages deliver the lowest switching energies of any memory technology, 100× lower than DRAM, 1,000× lower than flash and at least 10,000× lower than other emerging memories (projected to 20-nm node). The low voltage operation and low switching energy result in high endurance, no degradation is observed after 10^7 cycles [4] (actual endurance is unknown), while flash has a typical endurance of just 10^4. Notwithstanding the use of the 6.1 Å family of semiconductors (AlSb, GaSb, InAs), ULTRA**RAM** has been implemented on GaAs [5] and Si [3] substrates, with lattice mismatches of 7.8 and 12.3% respectively. The memory is thus demonstrably highly tolerant of threading dislocations, which manifest at densities of ~10^8 cm^{-3} [3], *i.e.* hundreds per μm-scale device. We will report on recent progress in developing a dry-etch process for fabrication of ULTRA**RAM** memories at the nanoscale and the production of 16-bit arrays.

Acknowledgements: This work was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie project Quantimony, grant agreement n° 956548, and ATTRACT, grant agreement n° 101004462, by the Engineering and Physical Sciences Research Council, UK, Impact Acceleration Account (EP/X525583/1) and by Innovate UK, project n°10078273.

References

[1] M. Hayne and D. Lane, US11929120B2 (2024).
[2] M. Hayne US10243086B2 (2019).
[3] P. D. Hodgson *et al.* Adv. Electron. Mat. **8**, 210103 (2022).
[4] D. Lane and M. Hayne, IEEE Trans. Electron Devices **67**, 474-480 (2020).
[5] O. Tizno *et al.*, Scientific Reports **9**, 8950 (2019); D. Lane and M. Hayne, IEEE Trans. Electron Devices **68**, 2271-2274 (2021).