

The Future of Charge Trapping Memories

T. Mikolajick¹, M. Specht, N. Nagel, T. Mueller, S. Riedel, F. Beug, T. Melde and K.-H. Küsters

Qimonda Dresden GmbH&Co. OHG

¹ Chair of Electronic- and Sensor Materials at the Technical University of Freiberg, Germany

INTRODUCTION

Since the early days of non-volatile memories in the late 1960s, charge trapping transistors using silicon nitride as a charge storage layer have been an alternative to floating gate devices in EPROM, EEPROM and later Flash memories [1,2]. However, the success was very limited due to the fact that charge trapping memory cells are hard to erase by Fowler-Nordheim tunneling. During erase electrons are injected from the gate which leads to the well known erase saturation (see fig. 1 and [3]). In the late 1990s two approaches were explored to overcome this limitation. Hot hole erase was used in multi bit charge trapping memories and new materials were introduced to increase the ratio of tunneling current across the bottom oxide to tunneling current across the topoxide in charge trapping NAND. In floating gate memories, on the other hand, it becomes more and more difficult to maintain a high gate coupling ratio and keep floating gate interference limited [4]. Therefore charge trapping memories have attracted more and more attention and are observed in an increasing number of products. This trend is expected to hold on, since the charge trapping concept has significant advantages when looking into advanced device structures like FinFET or 3D stacked memory cells.

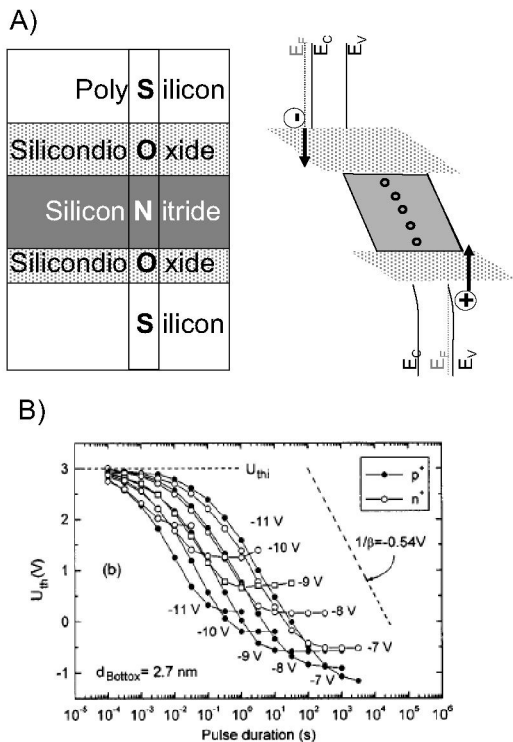


FIGURE 1. ERASE SATURATION IN A CHARGE TRAPPING MEMORY CELL. A) SONOS STRUCTURE AND BAND DIAGRAM DURING ERASE B) ERASE SATURATION FOR n^+ AND p^+ DOPED GATE ELECTRODES [3]

MULTI BIT CHARGE TRAPPING

In 1987 T.Y. Yan and co-workers introduced a memory cell that used a charge trapping storage layer and was programmed by hot electron injection [5]. Erase saturation was eliminated by using hot hole erase rather than tunneling erase. The combination of the localized storage in the nitride layer with a localized injection method resulted in very narrow charge storage regions self aligned to the drain of the device. This concept was used in the 1990s by B. Eitan, to construct the NROM (Nitride Read Only Memory) memory cell. NROM is a multi bit charge storage memory cell where both sides of the cell are used to store one bit of information [6]. Using a high enough drain voltage during a reverse read operation, where source and drain are interchanged compared to the programming operation, the information on the other side of the memory cell can be screened by punching through the area affected by the stored charge. To operate the cell in a symmetrical fashion, it is arranged in a virtual ground NOR array. This array architecture is a cross-point type of array leading in principle to $4F^2$ cell size. In a practical implementation, however, the channel length will limit one of the two pitches resulting in a physical cell size of about $6-8F^2$. The area efficient virtual ground architecture makes this cell not only attractive for the classical NOR market [7] but also allows addressing data type memory applications [8]. Fig. 2 shows the basic NROM concept and the TwinFlash implementation for three different technology nodes. By a proper device engineering the cell size could be shrunk very aggressively from about $0.08\mu m^2$ to about $0.01\mu m^2$ per bit.

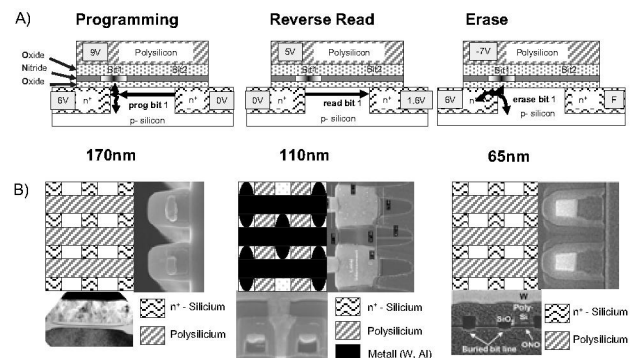


FIGURE 2. MULTI BIT CHARGE TRAPPING CELL CONCEPT (A) AND TWINFLASH IMPLEMENTATION USING 170NM, 110NM AND 65NM GROUNDRULES (B)

If the multi bit charge trapping concept is combined with a multi-level approach, 4 logical bits can be stored in the space of only one

physical bit [9]. With 4 levels on each side of the cell the four bits require a total of eight levels. In a pure multi-level approach, in contrast, $2^4=16$ levels are required to code 4 bits. Today the multi bit charge trapping memory cell is well established in high volume NOR type products [7,10] and has better scaling prospects than the classical floating gate based ETOX cell. For some time the overlap of the charge distributions between the two bits has been considered as a limit for multi bit charge trapping devices. However, even with a channel length as short as 100nm, an excellent separation of the 2 bits could be achieved by proper source-/drain and channel engineering [8]. For data type memories first products exist [7,8]. However, the existing products are still considerably lacking behind floating Gate NAND with respect to both minimum feature size and density. In addition, the hot carrier programming and erase mechanism is limiting the parallelism of program and erase operations. Therefore further innovations will be required to challenge the write/erase performance of NAND based memory concepts using the multi bit charge trapping approach.

CHARGE TRAPPING NAND

To implement the charge trapping concept into a classical NAND array, the erase has to be done by tunneling. Two interesting approaches have been presented in the last two years to solve the erase saturation using new material systems. The first implements a high-k topoxide combined with a high workfunction gate metal. The specific realization uses Al_2O_3 topoxide and a TaN gate electrode and is therefore called TANOS (Tantalum Nitride / Aluminum Oxide Nitride / Oxide / Silicon) [4,11]. The high-k topoxide reduces the voltage drop over the top oxide and the high workfunction gate increases the barrier for electrons from the control gate. Both effects help to reduce the electron injection across the topoxide during the erase operation. The combination of both shows performance results suitable for an implementation into a NAND architecture (see fig. 3a)). Recently a first 32Gb device was demonstrated using this implementation [11]. In the second approach the bottom dielectric is replaced by a sandwich of low-k / high-k / low-k dielectrics [12]. This was implemented in a silicon oxide / silicon nitride / silicon oxide stack, which circumvents the difficulties of new material integration associated with the TANOS approach. Since the bandgap of the high-k material is higher than that of the low-k material, the low-k / high-k / low-k dielectric stack creates a much higher effective barrier during retention than during erase (see fig. 3b, middle drawing). Therefore the hole tunneling during erase is enhanced without sacrificing on retention [13]. Since the effect is based on the different bandgaps of the high-k and low-k materials used in the bottom oxide stack, this approach is called BESONOS (Bandgap Engineered Silicon Oxide Nitride Oxide Silicon) by the authors. Replacing the floating gate cell by a charge trapping cell with Fowler Nordheim programming and erase can help to scale NAND devices further without significant changes in architecture and periphery circuits. However, there are still concerns whether both TANOS and BE-SONOS can reach the erase performance of today's floating gate NAND.

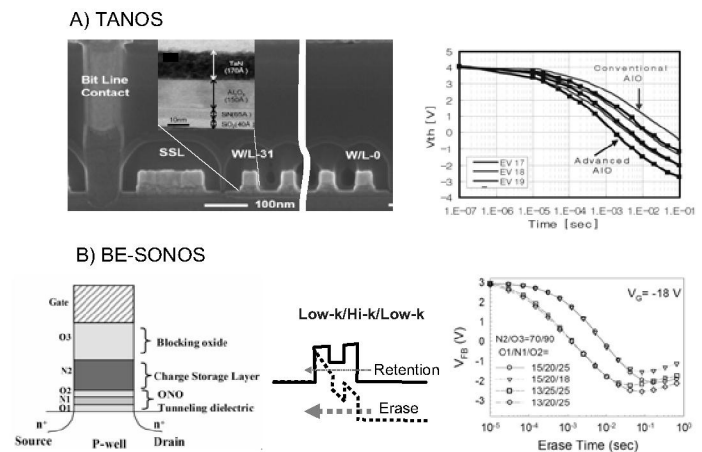


FIGURE 3. TANOS(TANTALUM NITRIDE / ALUMINUM OXIDE / NITRIDE / OXIDE / SILICON) MEMORY CELL (A) TAKEN FROM [11] AND BE-SONOS MEMORY CELL (B) TAKEN FROM [13]

FINFET NAND

When further scaling down NAND devices, the read current will become a major issue. A trigate FinFET type device as shown in fig. 4 can drastically improve the situation. Very promising results have been obtained when the charge trapping concept was combined with a FinFET type device [14]. Besides the enhanced read current, the erase saturation is also reduced when going to very narrow fins, since the curvature enhances the bottom oxide field. The use of p'-gates alone – without high-k topoxide – already resulted in NAND worthy erase performance and good retention (see right side of fig.4 and [14]). If a FinFET is combined with a floating gate device, the same effect will lead to a degradation of the coupling ratio, making charge trapping cells the better candidate for a FinFET type of architecture.

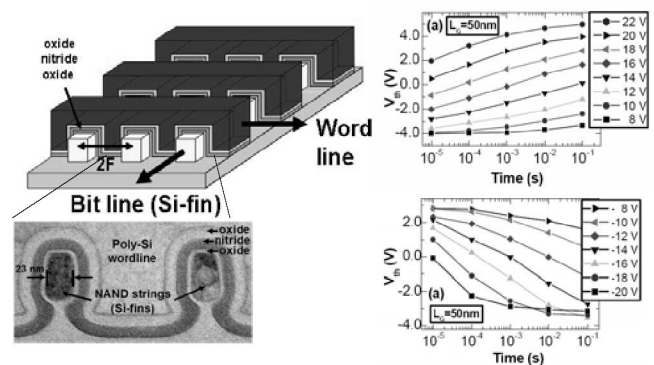


FIGURE 4. FINFET NAND MEMORY USING A CHARGE TRAPPING MEMORY CELL [14].

3D-NAND

Beyond the 30nm generation traditional scaling may not be sufficient to achieve the density requirements for data storage devices. This comes partly from the fact that the memory cell itself will be very hard to scale, but even more from the fact that lithography may not be available or at least will be too expensive to achieve the required cost structure [15]. Therefore the effective cell size has to be reduced below the lithographic limit. One approach is to store more than one bit into a physical cell by using multi bit or multi level approaches. Both multi bit as well as multi level are already being aggressively pushed. However, since 2^n levels are generally required to store n bits, it is not foreseeable, that more than 4 bits will be stored in a physical cell. Another approach to overcome the lithographic scaling limit is to stack memory cells on top of each other and to use the periphery circuits in common. Very recently it was demonstrated that the NAND architecture allows for such an approach with only 3 extra lithography layers, if a charge trapping memory cell is implemented [15]. In fig. 5 two NAND strings on top of each other are shown together with their program and erase characteristics. Good agreement between the program and erase characteristics of both layers was obtained. Again, floating gate is much harder to integrate into such an approach. It would definitely need more additional mask layers for the extra storage layer and therefore would negatively impact the cost position.

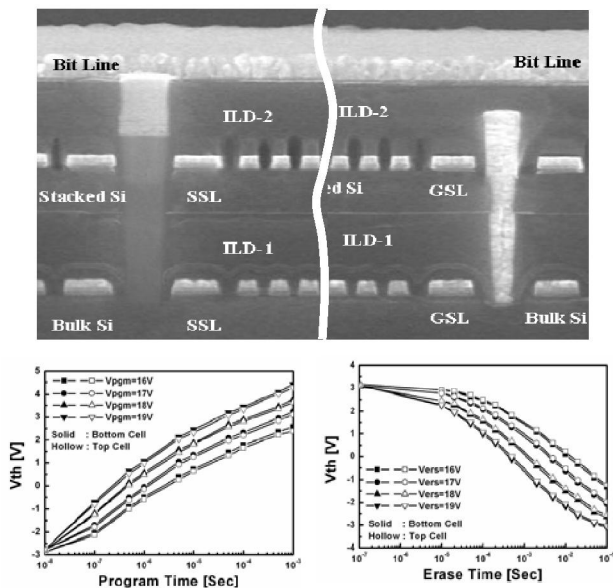


FIGURE 5. 3D NAND MEMORY USING A CHARGE TRAPPING MEMORY CELL [15].

DISCUSSION

The above described concepts allow charge trapping memories to scale into the 30nm feature size and multi 100Gb density arena. A large portion of these concepts depend on the distinct features of the charge trapping memory cell like easy integration and localized storage. Therefore charge trapping opens up much more scaling opportunities than floating gate. However, the possible transition

point is hard to estimate. For the NOR type memories, charge trapping already has achieved a considerable market share using the multi bit charge trapping concept. NAND type memories are currently still dominated by the floating gate. Due to the scaling limitations of floating gate by capacitive coupling and loss of coupling ratio combined with the advances achieved in solving the erase saturation issue by material innovations, it can be expected, that NAND memories will switch to charge trapping in one of the upcoming generations. The exact transition will depend on the further progress with both charge trapping and floating gate memories and may differ at different competitors. In principle also nanocrystal memories [16] are an option to take over from floating gate. As illustrated in fig. 6, nanocrystals are a special case in between floating gate and charge trapping, offering many of the advantages that charge trapping memories can offer. Additionally the erase saturation issue is less pronounced, since both tunneling from the gate electrode and tunneling out of the nanocrystal have the same voltage dependence. Thus higher erase voltages do not lead to higher saturation levels like in charge trapping devices [17]. However, from today's point of view the fluctuation in number of nanocrystals per device remains a major obstacle in nanocrystal devices. If a very controlled fabrication method can be found, nanocrystals may once pave the way to tailor the trapping properties of the memory cell. However, from today's point of view silicon nitride based charge trapping memories offer the better overall properties and the simpler manufacturing.

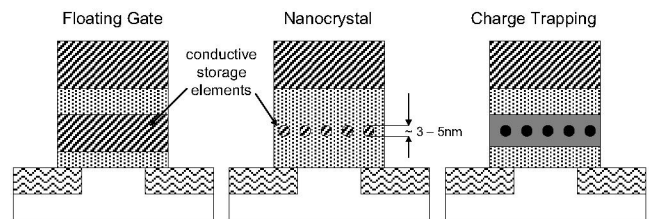


FIGURE 6. COMPARISON BETWEEN FLOATING GATE, NANOCRYSTAL AND CHARGE TRAPPING MEMORY CELLS

In summary charge trapping is the most promising concept for the medium time frame. Since many years also memories based on other switching mechanisms [18] have been explored. These memory concepts rely on the integration of new switching materials into a CMOS process. Fig. 7 gives an overview over the possibilities based on the choice of switching materials. Organic and molecular memory concepts offer interesting prospects. However, organic and molecular memories are still in a very early development stage with mainly single cell results demonstrated so far. In the field of inorganic switching materials, the most prominent and most advanced candidates are ferroelectric RAM, magnetoresistive RAM and phase change RAM. Additionally many resistive switching materials are investigated under the general term of resistive RAM. Ferroelectric RAMs and magnetoresistive RAMs are both in production for low density NVRAM memories. But both concepts are difficult to scale to smaller cell sizes. Only phase change memories currently hold the promise to result in both competitive cell size and reasonable performance for the short to medium timeframe by demonstrating a 512M device in 90nm technology with a cell size of $5.8F^2$ [19]. If a

multi-level capability can be added or if the device could be shrunk more aggressively very rapidly, this could challenge existing NOR type floating gate memories as well as multi-bit charge trapping memories for the classical NOR type application. For the NAND type application this cell size is still too large and a real cross point type array with 3D stacking capabilities would be required to challenge the existing technologies. Only basic results are available for such a technology [20]. The Experience from ferroelectric memories and phase change memories shows that there is significant effort ahead to bring these technologies out of the basics into production.

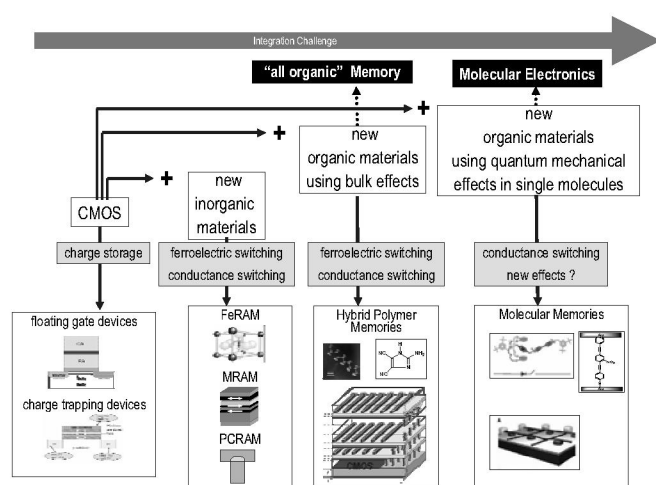


FIGURE 7. HIERARCHY OF NONVOLATILE MEMORY CONCEPTS OBTAINED BY INTEGRATING NEW SWITCHING MATERIALS INTO A CMOS PROCESS

SUMMARY

With floating gate memory cells running into scaling limitations caused by reduced gate coupling and excessive floating gate interference, charge trapping in its two variants multi bit charge trapping and charge trapping NAND is the most promising technology for the mid term. For NOR type applications also Phase change RAM could appear as a competitor in a few years, but some considerable development is still down the road. Concepts to challenge NAND type applications are still in the early stage. Therefore charge trapping is expected to be the technology of choice for code storage in the short to mid term and for data storage in the mid term timeframe.

ACKNOWLEDGEMENT

Parts of the work for this paper were supported within the scope of technology development by the EFRE fund of the European Community and by funding of the State Saxony of the Federal Republic of Germany.

REFERENCES

- [1] H.A.R Wegener et al., IEDM Digest of Technical Papers, 1969, p.70
- [2] J.T. Wallmark and J.H. Scott, RCA Review, 1969, p. 335-365
- [3] H. Bachhofer et al., Journal of Applied Physics Vol. 89, 2001, 5, p. 2791
- [4] K. Kim, Proceedings of the NVSMW, 2006, p. 9
- [5] T.Y. Chan, K.K Young and C. Hu, IEEE Electron Device Letters Vol. 8, 1987, 3, p. 93
- [6] B. Eitan et al., IEEE Electron Device Letters Vol. 21, 2000, 11, p. 543
- [7] M. van Buskirk, Proceedings of the NVSMW, 2006, p. 8
- [8] N.Nagel et al., to be published in Proceedings IEEE VLSI-TSA, 2007
- [9] B. Eitan, IEDM Digest of Technical Papers, 2005, p.22.1.1
- [10] http://www.spansion.com/flash_memory_products/mirrorbit.html
- [11] Y. Park et al, IEDM Digest of Technical Papers, 2006, p.2.2.1
- [12] M. Specht, M. Städele, F. Hofmann, Proceedings of ESSDERC, 2002, p. 599
- [13] H.-T. Lue et al., IEDM Digest of Technical Papers, 2006, p.18.5.1
- [14] C. Friederich et al., IEDM Digest of Technical Papers, 2006, p.37.2.1
- [15] S. M. Jung et al., IEDM Digest of Technical Papers, 2006, p.2.3.1
- [16] S. Tiwari et al., IEDM Digest of Technical Papers, 1995, p. 521
- [17] R.A. Rao et al., Solid-State Electronics Vol. 48, 2004, 9, p. 1463
- [18] C. U. Pinnow and T.Mikolajick, Journal of the Electrochemical Society Vol. 151, 2004, 6, p. K13
- [19] J.H. Oh et al., IEDM Digest of Technical Papers, 2006, p.2.6.1
- [20] I. G. Baek, IEDM Digest of Technical Papers, 2005, p.31.4.1