NEW NONVOLATILE MEMORY TECHNOLOGIES - A SURVEY

Anton Biasizzo

Institit "Jožef Stefan", Ljubljana, Slovenia

Key words: nonvolatile memory, flash ram, feroelectric memory, magnetoresistive memory, ovonic unified memory

Abstract: The use of nonvolatile memory is growing rapidly. Their use ranges from the applications in computers, digital photography to mobile telephony. The drawbacks of current nonvolatile memory technology (flash ram) are relatively low endurance, low speed and the use of high voltages for program/erase operations. In recent years several new nonvolatile memory technologies have emerged based on different physical phenomenons. They tend to provide unlimited endurance, high speed and are easier to integrate into the process flow of standard logic circuit. The main challenging technologies to traditional flash ram are ferroelectric memory, magnetoresistive memory and ovonic unified memory. Several largest companies already developed memories based on this new technologies but currently only low capacity devices are available.

Pregled novih tehnologij neizbrisljivega pomnilnika

Kjučne besede: neizbrisljiv pomnilnik, pomnilnik flash, feroelektrični pomnilnik, magnetouporovni pomnilnik, sprememba faznega stanja

Izvleček: Neizbrisljiv pomnilnik se vedno bolj uporablja na različnih področjih. Uporabljamo ga tako v računalnikih, digitalnih fotoaparatih kot tudi v mobilnih telefonih ter tudi mnogih drugih napravah. Pomanjkljivosti obstoječe tehnologije neizbrisljivega pomnilnika so predvsem relativno kratka življenjska doba, počasnost ter uporaba relativno visokih napetosti pri zapisovanju vsebine. V zadnje času so se pojavile nove tehnologije, ki za shranjevanje podatkov uporabljajo druge fizikalne pojave. Nove tehnologije obetajo predvsem neomejeno življenjsko dobo, večje hitrosti delovanja kot tudi lažjo integracijo v proizvodnjo električnih vezij. Nove tehnologije neizbrisljivega pomnilnika so feroelektrični pomnilnik, magneto-uporovni pomnilnik ter pomnilnik na osnovi spremembe faznega stanja materialov. Nekateri največji proizvajalci so že razvili pomnilniška vezja osnovana na teh tehnologijah, a kapaciteta le-teh je še relativno majhna.

Introduction

Since the beginning of the use of computers, nonvolatile memory was their essential part. Semiconductor ROM and PROM were one of the first technologies of the nonvolatile memory however their lack of changing their contents limited its widespread use. The EPROM and EEPROM were their successors. This technology enabled to clear the contents of the device by submitting them to the ultraviolet light or high voltage. While these devices were suitable for storing the firmware and BIOS (Basic Input/Output System), which changes very rarely, they couldn't be used as a storage device.

It was the appearance of the Flash RAM that dramatically changed the use of nonvolatile memory. As one could expect the Flash RAM initially made its way to the computers BIOS and firmware of different peripheral devices. But due to its programmability it soon became a vital part of nearly every electronic system. In the cell phone the flash RAM holds the instructions to send and receive calls, stores the phone numbers, messages and phone settings. Electronic products of all types, from consumer appliances to the industrial machinery, use them for storing the operational instructions.

Flash RAM

Flash memory is a programmable, read-only, non-volatile memory similar to EPROM and EEPROM memory /1/. While it is a derivate of EEPROM the main difference is in the ERASE operation. Flash memory uses in-circuit wiring to apply erase function to the predetermined sections known as blocks or the whole memory.

Flash memory cell is composed of MOS transistor constructed with two gates: the control gate (CG) and the floating gate (FG), which is placed between the control gate and transistor's channel. Floating gate is isolated from the control gate and the channel by a thin oxide layer as shown in Figure 1. This isolation allows the floating gate to store the electrons (electric charge) when the power is cut off, so the memory is non-volatile.

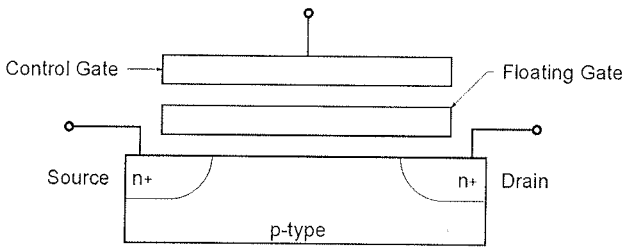

Figure 1: The structure of flash memory cell

Reading the value of the flash cell is relatively simple. The control gate is brought to logic 1 level. If there are no electrons at the floating gate then the transistor turns on. But if there are electrons at the floating gate, the charge blocks the impact of the electrical field induced by the control gate and the transistor remains in off state.

Writing to the flash memory is actually a two-phase process. First the trapped electrons are removed from the floating gate of a block of memory (ERASE operation). Blocks can consists of several thousand transistors. The ERASE operation is performed using the Fowler-Nordheim tunnelling effect. The high voltage (Vpp=12V) is applied to the source of the transistor while the control gate grounded and the drain of the transistor is not connected as depicted in Figure 2. By applying this high voltage the electrons trapped on the floating gate are drained across the thin oxide isolation barrier to the source

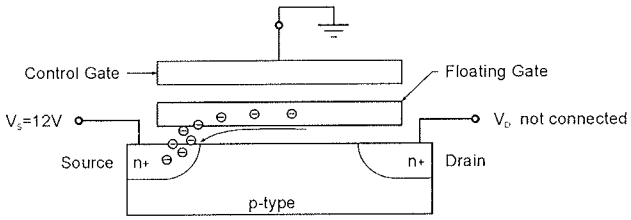

Figure 2: Erase operation using Fowler-Nordheim tunnelling effect

In the second phase the cell is programmed by transferring the electrons to the floating gates of the designated cells (WRITE operation). This is achieved using channel hot electron injection. The high programming voltage (Vpp=12V) is applied to the control gate, which forms the inversion region in the p-type substrate. The drain voltage is increased while the source is grounded. With the formed inversion region the current between the drain and source increases. This causes the electrons to gain sufficient energy to overcome the oxide barrier and collect on the floating gate as shown in Figure 3.

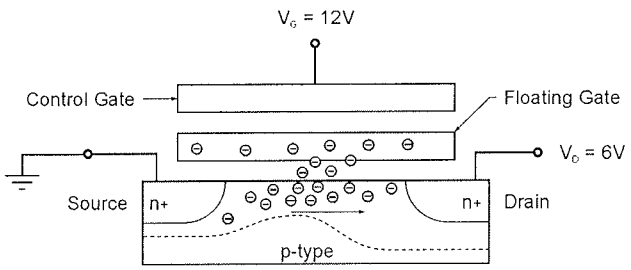

Figure 3: Write operation using the Channel Hot Electron Injection

This last operation is the cause for the main drawback of the flash memory, its endurance. Electrons shooting through the oxide barrier create defects in it that eventually cause the leakage between the floating gate and the channel. The flash memory sustains only a couple of ten thousands write cycles. This is not a problem for storing a computer BIOS, device firmware or even a cell phone, but it is a limitation for the use of digital cameras and other portable storage devices. This is especially notable with the use of nonvolatile disks.

Feroelectric memory

Ferroelectric memory (FeRAM or FRAM) technology /2,3/ is the most advanced rival of the flash technology. The core of the FeRAM cell is a capacitor formed by depositing a film of ferroelectric material in crystalline form between its plates. Lead-Zirconium-Titanate (PZT) or Barium-Strontium-Titanate (SBT) are usually used as a ferroelectric material. A model of a ferroelectric crystal is shown in Figure 4. A ferroelectric crystal has a mobile atom in the centre of the crystal. Applying electrical field to the lattice move the central atom in the field's direction. Reversing the field cause the atom to move in opposite direction. The central atom has two stable positions: at the top and at the bottom of the crystal lattice thus when the electrical field is removed the atom stays in the stable position.

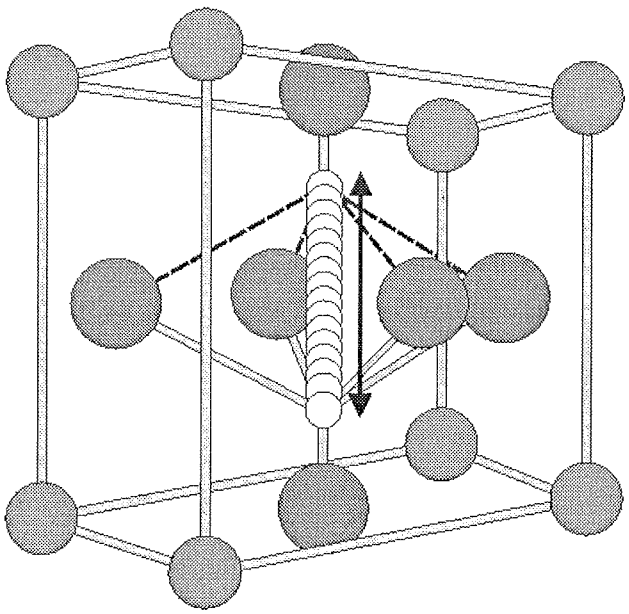

Figure 4: The structure of the ferroelectric crystal

As mentioned, the memory element of FeRAM is capacitor, however it does not store the data as linear charge but as electrical dipole. In order to read the memory cell the position of the central atoms has to be determined. Unfortunately this cannot be directly measured. Read operation is performed by employing the electrical field to the capacitor. If the central atoms in the lattice are in stable position opposite to the field direction they move in the field direction. If they are already in the stable position in the field direction, they stay put. As the atoms moves through the crystal a charge spike is emitted. A capacitor with switching electrical dipole of ferroelectric crystal produce larger charge spike than a capacitor where the dipole remains the same. The non-switching capacitor emits the charge equivalent to the charge emitted by ordinary capacitor. The memory circuit must determine whether the capacitor switched.

Read operation of the FeRAM cell involves the change of cell's state therefore the memory circuit has to restore the

initial cell state, just like in a DRAM cell. In fact FeRAM cell is similar to the DRAM cell. DRAM cell for the capacitor dielectric uses the silicon oxide and the data is stored as charge of cell's capacitor, which leaks into the substrate unless it is periodically rewritten. The state switch of FeRAM cell occurs in less than 1 ns and the complete circuit access is done in less than 50 ns.

Write operation of the FeRAM cell is very simple and does not require additional circuit. The data is applied to the FeRAM cell capacitor and the state of the ferroelectric crystal switches if necessary. The access time in the case of write operation is similar to the access time in read operation.

Initial FeRAM memory architecture consisted of two-transistors/two-capacitors (2T/2C) as shown in Figure 5.a. In this architecture for each data bit its own reference is provided achieving the robust data retention reliability. However the drawback of this architecture is relatively large cell size. With advances in ferroelectric materials and processing the need for internal reference capacitor was eliminated. As result one-transistor/one-capacitor architecture similar to DRAM architecture depicted in Figure 5.b was introduced.

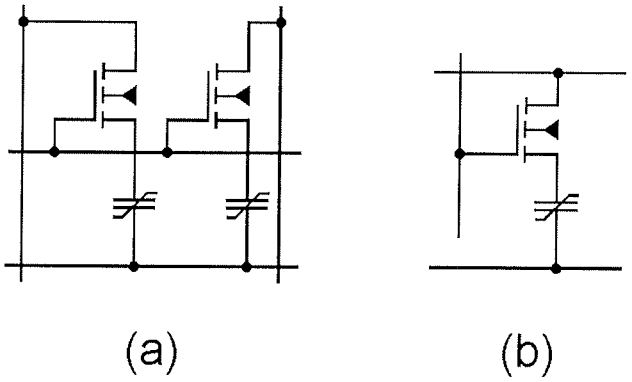

Figure 5: The structure of the FeRAM memory cell (a) 2T/2C and (b) 1T/1C

The major difference between FeRAM and flash memory is in difficulty of write operations. Writing the data to the flash memory involves erase and write operation in which electrons are pushed through the oxide barrier. This requires high voltages and relatively long time. Write time for flash memory is several tens ms while write access time for FeRAM is under 100 ns. Additionally flash memory allow very limited numbers of write operation due to the oxide barrier degradation. The endurance of the FeRAM memory is more then 10^{12} write cycles since write operation does not produce any stress. Finally no high voltages are needed for the write operation of the FeRAM memory.

While Ramtron International Corp. already fabricates and sells FeRAM devices since 1992 their capacity is still relatively small. Largest memories produced by Ramtron are 256Kb, which represents a small fraction of the gigabit chips offered by the major flash memory producers. Several other companies are also pursuing FeRAM as a replacement for today's non-volatile flash memory. Texas Instruments reported to produce a 64 Mb device for embedded applications /4/. Fujitsu, Samsung and others are also working on the development of larger FeRAM devices mainly for the use in the embedded systems.

Magneto-resistive memory

The magneto-resistive memory (MRAM) is an emerging memory technology that stores information using the magnetic moments of a thin ferromagnetic layer /2,5/. Atoms in a ferromagnetic material act like tiny magnets and respond to the external magnetic field aligning themselves in its direction. Similar to the ferroelectric materials they form domains. Applying an external magnetic field to the ferromagnetic material its domains line up with the external magnetic field and remain oriented in the same direction when the magnetic field is removed. When the magnetic field in opposite direction is applied, the domains flip over. MRAM uses the magnetic moments of a thin ferromagnetic material to store information.

The core of MRAM cell is Magnetic Tunnel Junction (MTJ) composed of two ferromagnetic layers, storage layer and reference layer, separated by thin insulating layer. If the insulation layer is ticker then few nanometers, the insulator stops the passage of the electrons. However insulation layer of the tunnel junction is thinner than 2 nm and some electrons can tunnel across it from one ferromagnetic layer to another. The magnetic moment of the reference layer is fixed while the magnetic moment of the storage layer change its direction in response to the applied magnetic field. In order to read the memory cell the orientation of the magnetic moment of the storage has to be determined. As in the case of FeRAM cell this cannot be measured directly. The orientation of the magnetic moment of storage layer can be parallel or anti-parallel with the orientation of the magnetic moment of the reference layer. When the magnetic moments are parallel the resistance of the MTJ is smaller then when they are anti-parallel.

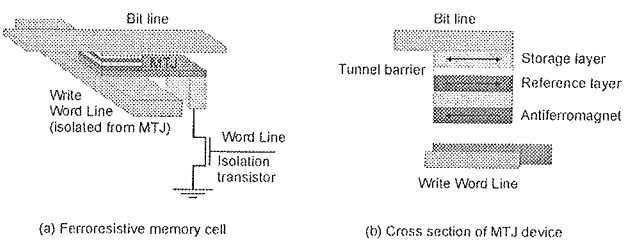

Figure 6. The structure of the Magnetic Tunnel Junction

Reading the MRAM cell is straightforward, small electric current is passed via word and bit lines through the MTJ of the selected cell. When the orientation of the magnetic moments is parallel the resistance is low and if the orienta-

tion is anti-parallel the resistance is high. To determine whether the resistance is high or low, the resistance is compared to the resistance of the reference memory cell along the same wordline. As reported by Motorola one reference cell is needed for every 64 memory cells in their 1Mb MRAM sample.

Writing the data into the cell is achieved by passing the current pulses through the wires close to (but not connected to) the magnetic cells. The magnetic field generated by the current in wires is used for changing the direction of magnetic moment in MTJ. The wires and cells are arranged in cross-point architecture depicted in Figure 7. The set of wires called bit lines run in parallel above the magnetic cells. The other set of wires called word lines runs under the magnetic cells perpendicular to the bit lines. Word lines are isolated from memory cells. MTJ are set up at intersections of the word and bit lines. To write a data to chosen memory cell the current pulse is applied to both its bit line and its word line. Only joint magnetic field of both lines is strong enough to change the direction of magnetic moments thus other MTJ along the bit and word lines remain unaffected.

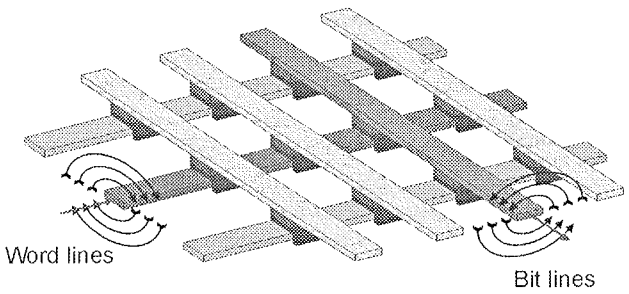

Figure 7: Cross-point architecture of the memory

MRAM has many advantages over other types of memory. It is faster than Flash memory, actually it could be even faster then DRAM and almost as fast as SRAM (Static RAM) /6/. MRAM has also the potential of extremely dense packaging. The only limiting factor is the density of the set of wires. Several MRAM layers can also be placed on top of a single die. The endurance of MRAM is also extremely high (10^{14} write cycles) and at last MRAM is nonvolatile since the magnetic moments maintain their alignment even when the power is switched off.

MRAM also has several potential drawbacks. It requires relatively high power to write and the MRAM layer may interfere with heat dissipation. Main problems in MRAM production are related to the thin insulation layer a key element of each MTJ. It must be uniform across the wafer since the resistance of MTJ rises exponentially with its thickness. To ensure reliable readout of the cell the difference in the resistance of the parallel and anti-parallel magnetic moments in MTJ must be significant. At the beginning of the development of MRAM cell it was around 5% but with the use of different materials in MTJ and reducing insulation thickness the 40% change of resistance was achieved.

The research and development of MRAM memory started in IBM, where similar techniques were introduced in design of hard disks heads which drastically increased the capacity of new hard drives. Infineon joined IBM in development of new MRAM chip /7/. It is expected that first prototypes of capacity 256Mb should be available in year 2004. Motorola, considered a leader in the development of MRAM technology, introduced their 1Mb MRAM chip in April 2002 /8/ and new 4Mb MRAM chip in October 2003 /9/. NEC and Toshiba joined in the development of their own 256Mb MRAM device.

Ovonic unified memory

The Ovonic Unified Memory (OUM) technology is another promising nonvolatile technology. It uses the reversible structural phase-change from the crystalline phase to amorphous phase as a storage mechanism /2,10/. Several chalcogenide alloys exhibit reversible phase-change and are used in optical fibers and rewriteable CDs and DVDs. The OUM technology uses a thin film chalcogenide alloy of germanium, antimony and tellurium ($Ge_xSb_yTe_z$). These alloys are suitable for memory elements because it can rapidly switch, when heated and cooled, between both amorphous and crystalline phase. Each phase state represents different digital value. When alloy is in amorphous phase it has low light reflectivity and high resistance, while in crystalline phase its reflectivity is high and the resistance is low. This property of the chalcogenide alloys is also used in rewriteable optical media.

The schematic illustration of the OUM memory cell cross section is shown in Figure 8. It consists of a layer of the chalcogenide alloy and a resistive heating element placed in between connecting electrodes. The portion of the thin chalcogenide alloy in the junction with the heater is heated and cooled in order to switch the phase of the alloy. This switch in the phase of material causes significant change in alloy resistance. Because of the ultra small size of the media programmed thermal time constant of the device are very shot, on the order of a nanosecond. Since the write energy also scale with the programmed media volume the power consumption of the write operation is very low.

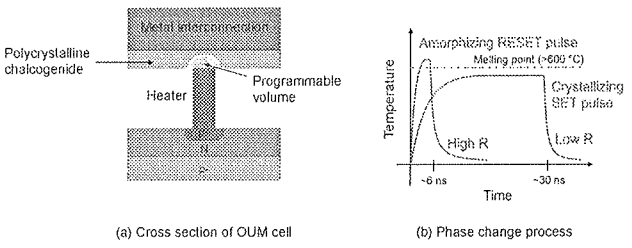

Figure 8: (a) schematic cross section of the OUM memory cell, (b) phase change process

Reading the content of the OUM memory cell is similar to reading the MRAM memory cell. It is done by measuring

its resistance however the change of the resistance is much larger, approximately a factor of 100. Due to the large range of programmed resistance a multi-bit per cell storage capability is possible.

Writing the data into OUM memory cell is achieved by heating the chalcogenide near its melting point and then rapidly cooling it. The actual parameters of the heating process depend on the value to be programmed into the device. The set pulse is applied to change the phase of chalcogenide to crystalline phase. During this pulse the temperature of the material is held just below the melting point for a time sufficient for crystal growth. The reset pulse change the phase of the material to amorphous phase. This pulse has slightly higher amplitude and cause the programmed material to melt and eliminates the crystalline structure. When the reset pulse is terminated the material "freeze" in the disordered amorphous structure.

OUM memory technology is more endurable and faster then Flash memory. Potentially, it can be even denser than both MRAM and FeRAM memory technology. It has faster read/write access than FeRAM technology however MRAM cells are faster. On the other hand the power dissipation is lower then the power dissipation of MRAM cell. Its main advantage over their competitors is easy integration into conventional CMOS process technology.

OUM memory technology is based on proprietary technology originally developed by and exclusively licensed from Energy Conversion Devices, Inc (ECD, Inc). The Ovonyx, Inc. the wholly owned subsidiary of ECD, Inc is further developing the OUM technology. The largest device made to date is 4Mb array built with partner Intel Corp. /11/. Ovonyx is also working with other semiconductor producers namely STMicroelectronics, which has announced to present their first OUM samples in 2005.

Summary

Besides the presented new memory technologies there is ongoing effort in developing other memory technologies as well. Polymer memories, carbon nanotube memory, single electron memories and nitride storage memory are only some of them. However they are still more or less at the beginning of the development phase.

The well-established technologies as flash RAM are also further developed to overcome their weakness. Multi-bit storage is that would rise the device density is considered in several technologies. Also the use of diode instead of transistor or even cells without transistor are developed to achieve higher memory density.

Presented technologies are adopted by major semiconductor producers and, while each of them provides technical challenges to be overcome, first product can be expected soon. The comparision of described technologies with proven memory technologies is given in following table.

As one can see new technologies outperform the flash memory and can even be compared with DRAM. However their current capacity is too small to present a serious threat for both flash and DRAM memory. Currently they are mainly used in embedded systems where the memory is placed on a same die as processor and other devices. Due to the rapid development of these new as well as traditional technologies the given comparison may soon become inaccurate. The race for predominant memory technology is still open.

References

/1/ B. Pavan, R. Bez, P. Olivo, and E. Zanoni, "Flash Memory cell-An overview", Proc. IEEE, vol. 85, no. 8, Aug. 1997, pp. 1248-1271

/2/ L. Geppert, "The New Indelible Memories", IEEE Spectrum, March 2003, pp.49-54.

/3/ Ramtron Inc., "Ferroelectric Technology Backgrounder", Technology Note, Dec 2003, http://www.ramtron.com/lib/literature/FRAMTechBkgder.pdf

/4/ "Texas Instruments Eyes Embedded FRAM As Next Generation Non-Volatile Memory Technology", http://focus.ti.com/docs/pr/pressrelease.jhtml?prelId=sc02272

/5/ R. Butner , "Magnetic RAM cures your computer of short-term memory loss", http://www.research.ibm.com/thinkresearch/pages/2001/20010202_mram.shtml

/6/ R. Desikan, C. R. Lefurgy, S. W. Keckler, and D. Burger, "On-chip MRAM as a High-Bandwidth, Low-Latency Replacement

	SRAM	DRAM	Flash	FeRAM	MRAM	OUM
Capacity [Mb]	4		4096	64	4	4
Structure	6T	1T/1C	1T	1T/1C	1T/1MTJ	1T
Cell size [F^2]	50-80	6-12	7-11	18	10-20	5-8
Access time [ns]	1	20	50	20	15	20
Write time [ns]	1	20	2 ms	20	15	20
Voltage [V]	3.3	3.3	3.3 and 12	3.3	3.3	3.3
Write Power	high	medium	high	medium	medium	low
Retention	volatile	volatile	> 10 yrs	> 10 yrs	> 10 yrs	> 10 yrs
Endurance			10^5	10^{13}	10^{15}	10^{13}
Multi-bit	no	no	yes	no	no	yes
Availability	now	now	now	now	2004	2004

for DRAM Physical Memories", Technical Report TR-02-47, Department of Computer Sciences, University of Texas at Austin, 2002.

/7/ IBM Inc., "IBM, Infineon Develop Most Advanced MRAM Technology to Date", http://www.research.ibm.com/resources/news/20030610_mram.shtml

/8/ M. Durlam et al., "A low power 1Mbit MRAM based on 1T1MTJ bit cell integrated with Copper Interconnects", IEEE Symp. on VLSI Ckts, June 2002, pp.158-161.

/9/ M. Durlam et al., "A 0.18mm 4Mb Toggling MRAM", IEDM Digest of Technical Papers, 2003, session 34, paper 6, 2003

/10/ Ovonyx, Inc., "Ovonic Unified Memory", http://www.ovonyx.com/tech_html.html

/11/ S. Lai, Tyler Lowrey, "OUM - A 180 nm Nonvolatile Memory Cell Element Technology For Stand Alone and Embedded Applications", Session 36.5, IEDM Technical Digest, 2001.

/12/ M. Gill, T. Lowrey, J. Park, "Ovonic Unified Memory- A High-performance Non volatile Memory Technology for Stand Alone Memory and Embedded Applications", TD 12.4, Proceedings of ISSCC, 2002.

Anton Biasizzo
Institit "Jožef Stefan", Ljubljana, Slovenia
phone: 4773 297 fax: 2519 385
e-mail: anton.biasizzo@ijs.si

Prispelo (Arrived): 18.03.2004 *Sprejeto (Accepted): 20.06.2004*