TÉCNICO LISBOA

Nanotechnologies and Nanoelectronics

Homework #2

Duarte Miguel de Aguiar Pinto e Morais Marques, 96523
LEFT

Period 2, Semester 1

January 9, 2022

Professor Susana Cardoso de Freitas

1) Flash memories

Memory type: Transistor based RAM (Matrix-based memory, Non-volatile RAM) or UV erasable EPROM (Matrix-based memory, Re-programmable ROM)

Design of a typical architecture of the storage element: The sensing device which detects the presence or absence of stored electric charge should be in immediate proximity to the storage node. A field effect transistor (FET) is commonly used. A complete nonvolatile floating gate memory cell consists of a stack of metallic and insulating layers on the top of a FET channel (see Figure 1). In every semiconductor memory, each memory cell has to connect to two orthogonal signal lines, called a ***wordline*** **(WL)** and a ***bitline*** **(BL)**, forming a ***memory array***. The memory device is located at their point of intersection.

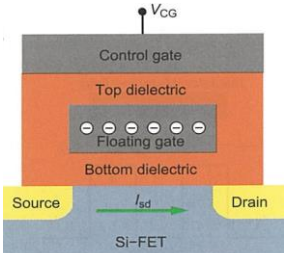


Figure 1: Generic cross-section of a floating gate memory cell

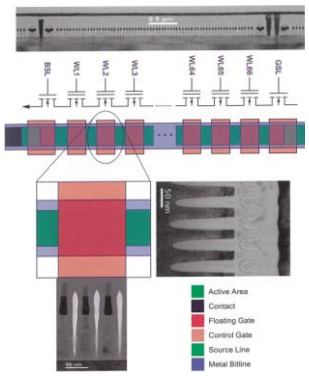


Figure 2: Layout and cross sections of a string of NAND memory cells

A straightforward way to implement a flash memory cell into such an array is to connect the control gate of every memory cell to the *wordline*, the drain of the FET to the *bitline* and all the sources of the memory cells to the ground – NOR array (different possibilities in Figure 3). It is also possible to connect the cells on the *bitline* in a serial connection resulting in a NAND type array. In practical applications, the number of cells in series has to be limited to keep the read current on an acceptable level, typically up to 66 cells. The random-access time is much faster in the NOR type array, since every cell is directly accessed by a *bitline*. In the NAND architecture, the serial connection of cells results in a high resistance to the current flowing (much lower read current and much slower random access). On the other hand, the NAND arrangement has a distinct size advantage. Figure 2 gives an overview of a typical NAND string including cell cross sections taken from a 25nm cell. The main operations - read, write and erase - are illustrated in Figure 4.

Some of the **materials** used are shown in Table 1. The bottom (tunnel) dielectric in floating gate cells is SiO_2 formed by thermal oxidation of the single-crystal silicon FET substrate. Its **thickness** is ≈8nm. The thickness of the top material is larger; for example, the physical thickness of the oxide-nitride-oxide composite film is not less than ≈15nm. However, practical values of thickness of dielectrics depend on the concrete reliability specifications. Aluminium oxide (Al_2O_3, k≈9) is considered a promising blocking dielectric in floating gate devices. For higher *k* (>10) and sufficiently high W_b (>1.7 eV), combinations of materials with different properties can be used. For example, $HfAlO_x$ layers of different compositions offer a useful trade-off between high-k properties and high W_b (*k* and W_b will be further discussed below).

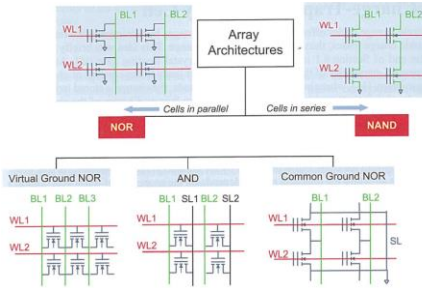


Figure 3: Array architectures for flash memory arrays

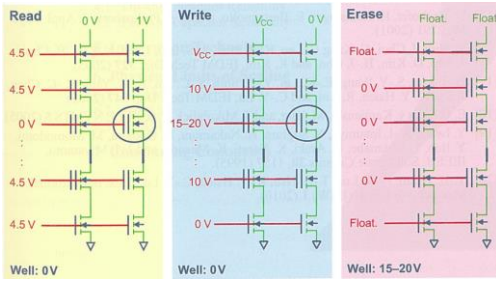


Figure 4: Read, write and erase operations of a NAND memory array

Material	Dielectric constant, *k*	Barrier height, W_b (with Si)	Effective electron mass, m^*	a_{min}
SiO_2	3.9	3.1 eV	0.50 m_0	5.0 nm
Si_3N_4	7.6	2.4 eV	0.43 m_0	6.0 nm
Al_2O_3	9	2.8 eV	0.30 m_0	6.8 nm

Table 1: Insulator material parameters and theoretical minimum insulator thickness for nonvolatile storage

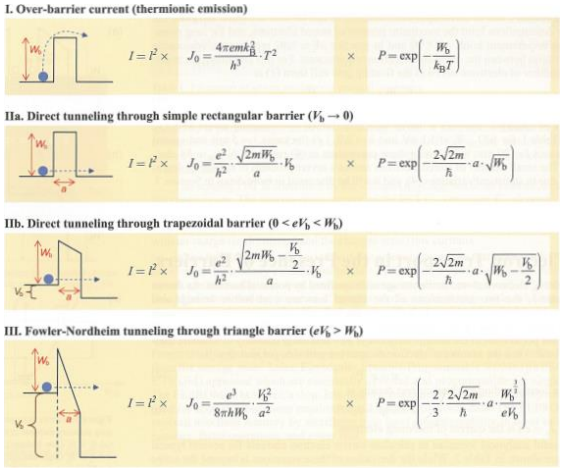


Table 2: Electron transport in the presence of barriers

Physical operation principle and basic equations: In a cell, the charge storage node is either on a conductive electrode surrounded by insulators (floating gate), in discrete traps within a defective insulator layer (charge trapping layer) or in nanocrystals embedded in the insulator layer. It is a **nonvolatile electron charge-based memory**, the basic building block of which is a floating gate cell. The term "flash" refers to the erase operation where many cells are cleared to one state (**erased**) in a large block **simultaneously**. A material structure to implement an electron-based memory must satisfy **charge retention** (very small flow of charge in store mode), **injection** (sufficiently large flow of charge during write mode) and **sensing** (read mode; the charge in the storage node should be sufficiently large, so that it can be detected by an electrometer type device). To prevent losses of stored charge, the node is defined by **energy barriers** of sufficient height $\boldsymbol{W_b}$ to retain charge.

A field effect transistor (FET) is commonly used as a sensor. It is controlled by the voltage V_{CG}, applied to the control gate (an external electrode). Measuring the FET current at a certain voltage allows one to distinguish between the state with **no charge (high current)** and the **charged state (low current)**. The floating gate storage node consists of a metallic storage node surrounded by layers of insulator. There are limitations on the number of stored electrons. When a number N_S of electrons is placed in the storage node, the node charge q=e· N_S results in an increase of the node potential $V_S = \frac{eN_S}{C_m}$ (C_m is the capacitance of the memory cell). Using the standard parallel-plate formula, cross-sectional area $A = l^2$, capacitor dielectric thickness *a*, dielectric constant of barrier material *k* and assuming a symmetrical barrier structure (not used in actual devices), the formula $\mathbf{C_m} = \frac{2\varepsilon_0 k}{a} l^2$ is obtained. The barrier deformations limit the maximum number of stored electrons. For long retention, the requirement holds $\mathbf{eV_S < W_b}$ and, in practice, $eV_S \approx 0.5W_b$ represents a reasonable compromise. For this condition, the maximum number of electrons stored in the floating gate cell is $\mathbf{N_{S_{max}}} \approx \frac{C_m W_b}{2e^2} = \frac{\varepsilon_0}{e^2} \cdot kW_b \cdot \frac{l^2}{a}$.

The two mechanisms of charge loss are over-barrier leakage and through-barrier leakage. In both, the leakage current from the storage node results from collisions of the stored electrons with the barrier walls. The number of electron escapes per unit area per unit time is $\mathrm{n} = \nu \cdot \mathrm{P}$ and the electron current density is $\mathrm{J} = \mathrm{J}_0 \cdot \mathrm{P}$ (ν is the frequency of collisions per unit area, *P* the probability of a barrier transition and $\mathrm{J}_0 = \mathrm{e} \cdot \mathrm{n}$ the current of incoming electrons). In Table 2, formulae to calculate barrier electron currents for several typical cases are shown.

Physical limits of scaling (limitations to reduce the size towards an ultrahigh density memory)

Among the main factors limiting scalability are the **control gate coupling ratio** and the **barrier insulator thickness**. While the coupling ratio issue can be, at least in principle, addressed by aggressive introduction of high-*k* materials, the barrier insulator thickness represents a fundamental scaling problem. Charge retention in memory cells depends on the parameters of the barrier surrounding the storage node, namely the height (W_b) and the width (a). For long retention (e.g., >10 years), the theoretical barrier width must be >5nm for all known dielectric materials (typically, >7nm in practical devices). Therefore, if one conducts a thought experiment on constructing a smallest possible floating gate memory cell from perfect materials, its minimum vertical and lateral dimensions mut be >2a=10nm. If a storage node of finite size is added, the minimum size will further increase. With only one electron be stored in the storage node, then the storage node size should be ≈3nm. Thus, the total size of the hypothetical minimal FG cell would be ≈13nm.

This analysis does not include scaling limits of the sensing FET, whose performance will degrade with decreasing gate length due to short--channel effects and a non-scalable gate dielectric. An additional limitation on the FET channel length arises if the channel hot electron injection is used from the write operation. In this case, applying a large voltage to a very short channel will result in **punch-through or junction breakdown**. In principles, the FET channel length issues could be circumvented using vertical FET structures. In combination with 3D stacked cell arrangement, this could be one way to extend the density increase of NAND flash memories beyond the limits described here.

An additional scaling challenge arises from the **array operation**. The reduced spacing between neighbor floating gates will lead to the strong capacitive coupling and therefore severe cross-talk between the cells. This calls for low-*k* materials to be used between the sidewalls of the neighbor FGs. The ultimate solution is using air gaps separating the cells.

How the retention time can be calculated and what are the key physical parameters that affect the retention of a bit state

The time to move an electric charge *q* between two terminals of an electric circuit is $t = \frac{q}{I}$ (where I is the electric current between the terminals). In the floating gate cell, the total current between the storage node and an external contact is $\mathrm{I} = \mathrm{I}_{\mathrm{o-b}} + \mathrm{I}_{\mathrm{T}}$. Assuming that 50% discharge of the storage node corresponds to an unwanted transition between the two states, if N_S electrons are stored, then the *store* or ***retention time*** can be estimated as:

$$\mathrm{t_r} = \frac{0.5\mathrm{e} \cdot \mathrm{N_S}}{2(\mathrm{I_{o-b}} + \mathrm{I_T})} = \frac{\mathrm{e} \cdot \mathrm{N_S}}{4(\mathrm{I_{o-b}} + \mathrm{I_T})}$$

The factor of 2 in the denominator appears because escape is possible over either of two barriers that confine the storage node. Table 1 shows parameters for several insulator materials and the corresponding theoretical minimum insulator thickness a_{min} for nonvolatile operation. The barrier scaling limit is one of several scaling limits of charge-based nonvolatile memories.

Reliability issues

The maximum number of electrons stored in the floating gate cell is $\mathrm{N_{S\,max}} \approx \frac{\mathrm{C_m W_b}}{2\mathrm{e}^2} = \frac{\varepsilon_0}{\mathrm{e}^2} \cdot \mathrm{kW_b} \cdot \frac{\mathrm{l}^2}{\mathrm{a}}$. As an example, considering a barrier structure formed by insulating layers of SiO_2, we have that $\mathrm{W_b}$=3.1 eV and k=3.9; using thickness t=5nm and l=20nm, we have that $N_{S_{max}}$≈53 electrons - this small number causes a severe problem of reliability degradation due to statistical variations.

One of the major issues regards **imperfect dielectrics**. Defects are always present in real dielectrics; there are additional leakage mechanisms for electrons stored in the floating gate, such as defect-assisted tunneling. In addition, the defect structure of dielectrics evolves with repeated write and erase operations, resulting in degradation of the insulating properties, which eventually limits the lifetime of the memory cell. For data retention and endurance, thicker dielectrics are needed due to enhanced leakage in imperfect dielectrics as compared to the ideal case. The top (blocking) dielectric is prepared by chemical vapor deposition and has many imperfections.

Regarding the floating gate cell FET, the **control gate issue** is very relevant. Differently from a conventional transistor, the degree of accessibility of the channel from the control gate is rather limited for two reasons. First, the control gate (CG) is far away from the channel, since the minimal thickness of both top and bottom dielectric layers is relatively large due to the retention requirements. Second, the CG controls the channel only indirectly, as the floating gate lies between the CG and the channel. The CG is connected to the channel through series capacitors formed by both gates, the dielectric layers and the FET. Maintaining adequate control gate coupling ratio is recognized as one of the most difficult problems in scaling the flash memory devices.

The aggressive introduction of high-k materials for the top dielectric layer, which have also lower interface barrier heights (W_b), is a common practice. Using the materials with higher *k* may result in increased leakage, thus requiring thicker layers (hence reducing the expected capacitance). Additionally, imperfections in practical high-*k* dielectric layers require thicker physical films, thus the effect of high-*k* on capacitance may be diminished. Also, the high-*k* films are usually prone to trapping making cell operation problematic.

Example of manufacture companies for such a memory device [1]

Samsung became the world's No. 1 flash memory brand in 2003. The Compound Annual Growth Rate for the flash memory market is projected to be 4.9% for the years between 2020 and 2025. The Flash Memory market was worth USD 62.06 Billion in 2019. It is expected to be worth USD 80 billion in 2025. Recently, Intel and Micron partnered up to create the company IM Flash, a new branch that's devoted to research and development to produce a new generation of flash memory cards.

The following were the largest NAND flash memory manufacturers, as of the first quarter of 2019: Samsung (29.9%), Kioxia-formerly Toshiba (20.2%)-, Micron Technology-Crucial (16.5%)-, Western Digital-SanDisk (14.9%)-, SK Hynix (9.5%), Intel (8.5%).

[1] https://www.marketresearchfuture.com/reports/flash-memory-market-986 (Market Research Future), https://www.samsung.com/semiconductor/minisite/ssd/worlds-no1-flash-memory/overview/ (Samsung)

2) Capacitive memories DRAM

Memory type: Volatile RAM (Matrix-based memory)

Design of a typical architecture of the storage element: The required dimensions for each cell are **2F** for the ***bitline*** and **4F** for ***wordlines*** (2F for the active *worldline* and 2F for the passing *wordline*), so the minimal achievable cell size is $\mathbf{8F^2}$, where *F* is the minimum feature size (for example, a line or a space width). No smaller cells are possible with the folded-*bitline* scheme (see Figure 5a). For cell sizes smaller than $8F^2$, the passing *wordline* has to be eliminated and the reference *bitline* cannot be taken from the adjacent cell, as it will be loaded by the charge of the neighboring cell, activated by the same *wordline*. In other words, every cell on the activated *wordline* is giving its information to its connected *bitline* in one array, so all *bitlines* carry information. The reference *bitline* has to be taken from a different array block or an extra reference cell. The smallest achievable cell size with this architecture is $4\mathrm{F}^2$, using 2F for the *wordline* and 2F for the *bitline* direction. First commercial-available products with sub-$8F^2$ cells had a cell size of $6F^2$, realized with stack capacitors (the cell architecture is shown in Figure 5b). Two cells share one *bitline* contact and the two node (=capacitor) contacts are separated by isolation gate.

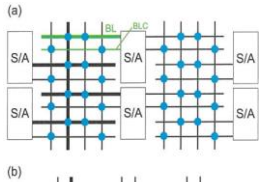


Figure 5: (a) Folded *bitline* scheme used for $8\mathrm{F}^2$ cell: a BL is compared to the adjacent BLC; (b) Open *bitline* scheme used for the $6\mathrm{F}^2$ cell: a BL is compared to one in the next array (BLC).

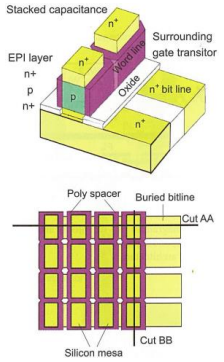


Figure 6: $4\mathrm{F}^2$ cell layout (a) and its bird's eye view (b).

In order to achieve the smallest theoretical cell size of $4\mathrm{F}^2$, the transistor channel was completely arranged vertically. A $4\mathrm{F}^2$ cell is shown in Figure 6. It does not have a body contact. It is difficult to achieve the full DRAM retention time of 64ms with floating body cells. Recently, new system architectures with more parallel internal data read/write operations were introduced. The realization with the 1T-1C DRAM cell was uncomplicated. With one active *wordline* in the array, the information of 512 to 1024 cells can be delivered or written in parallel.

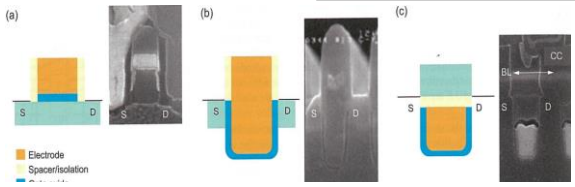


Figure 7: 3-D array devices (schematic views and SEM pictures): (a) planar transistor; (b) recessed channel device; (c) buried *wordline* (BWL) device (even the electrode is buried in the Si substrate).

DRAM	GR > 40 nm	GR < 40 nm	
Dielectric	HfO_2/ZrO_2 based	STO	Al doped TiO_2
Electrode	TiN	$SrRuO_3$	RuO_2
Barrier layer needed	no	yes	yes
T-budget [°C]	500 °C	500 °C	500 °C

Table 3: Typical DRAM dielectric, electrode, barrier and plug materials

A **DRAM access transistor** serves as a switch for connecting the capacitor to the *bitline* in the 1T-1C cell during reading, writing and refreshing of information. In the activated conducting state of the device, the capacitor is charged or discharged within a few nanoseconds. Assuming a 25fF cell capacitance, the access device must pass at least 10µA within a *bitline* voltage of 1.0V. In the non-activated state, the OFF-current of the device has to be below 1fA in order to keep the charge loss low enough during the refresh time. The DRAM access device has to cover eight orders of magnitude between ON- and OFF- currents (much higher demand than for FETs in *logic* gates).

Symmetric planar devices (see Figure 7a) were used as access devices for DRAM generations down to around 120nm. The highest electric fields across the p-n junction within these devices are reaching critical values of 0.5 MV/cm due to high channel doping, causing retention failures due to trap assisted tunneling. With the use of **3D devices** (see Figure 7b), the device length was turned into the third dimension and therefore independent of the footprint of the device. Gate induces drain leakage (GIDL) becomes now the most critical leakage mechanism in the array device. To suppress the GIDL, **elevated source drains** were introduced to pull up geometrically the node junction to the thicker *wordline* spacer region or thicker spacers introduced in the recessed channel (see Figure 7c). With the introduction of this buried *wordline* technology in 2008, the *wordline* became a continuous metal line below the silicon surface. In each cell, it is directly used as a gate of the array device, resulting in a buried array transistor. With increasing height of the stack capacitor, long *bitline* contact parallel to each capacitor were achieved. Accordingly, due to this contact, not the whole cell area could be used for the capacitor, which became a critical factor, such that this concept was replaced by a capacitor over *bitline* (COB) approach. Later on, after the introduction of TiN bottom electrode, free standing cylinders were possible to be manufactured with dielectric deposition on the inner and outer sides of the cylinder structure, which doubled the available surface area. Since the structures became more fragile, a stabilization layer at the top of the capacitor was introduced. For generations <40nm, the cylinder structure is replaced by a pedestal structure (the bottom electrode is a simple post with a square or circle cross section).

Physical operation principle and basic equations: The structure of the 1T-1C DRAM cell is shown in Figure 8. The access **transistor** T acts as a **switch** and is addressed by the ***wordline*** WL controlling the gate. The memory **capacitor** C_S represents the charge storage element for the information and is connected to the ***bitline*** BL via the transistor. When the switch is closed (conducting), the information on the *bitline* is written to the capacitor. The charge on the capacitor represents the binary information – a '1' with a *bitline* voltage V_{DD} and a '0' represented by 0V on the ***bitline*** ($V_{DD}/2$ is the reference voltage). Following this write operation, the capacitor is disconnected by opening the transistor switch. For reading the memory state, the charge must be sensed with an external circuit. A flip-flop is commonly used to translate the charge back into digital information. To initiate the sensing, first an equalizing signal of $V_{DD}/2$ is applied to both differential inputs (flip-flop now in an unstable state). By opening the transistor switch, the charge on the corresponding capacitor is redistributed between C_S and the bit capacitance C_{BL}, leading to a voltage change. Depending on the previous write operation which charged C_S to $V_{DD}/2$ (for '1') or -$V_{DD}/2$ (for '0'), **the *bitline* is pulled to**

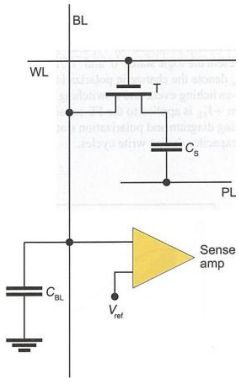


Figure 8: DRAM memory cell (C_S is a dielectric)

$$\mathrm{V_{BL}} = \left(1 \pm \frac{\mathrm{C_S}}{\mathrm{C_S} + \mathrm{C_{BL}}}\right)\frac{\mathrm{V_{DD}}}{2}, \qquad \begin{cases} + \text{ for '1'} \\ - \text{ for '0'} \end{cases}$$

The voltages $V_{BL}^{(0)}$ and $V_{BL}^{(1)}$ designate '0' or '1'. During the read operation, the charge in the cell capacitor must be restored for a subsequent read operation. With the flip-flop swinging to a stable state, the information in the cell can be rewritten. In order to reduce the electric voltage stress over the dielectric of the capacitor, the *plateline* PL as opposite electrode is kept at $V_{DD}/2$. This reduces the applied capacitor voltage to $\pm V_{DD}/2$ instead of 0V and full V_{DD}.

The geometrical dimensions can be approximated very well by using the equation $\mathbf{C_S = \varepsilon_0 \varepsilon_{r,eff} \frac{A_S}{t_{phys}} = \varepsilon_0 \varepsilon_{r,SiO_2} \frac{A_S}{t_{eq}}}$, with $t_{eq} = \frac{\varepsilon_{r,SiO_2}}{\varepsilon_{r,eff}} t_{phys}$ and $\varepsilon_{r,SiO_2} = 3.9$; A_S is the total area of the capacitor, t_{phys} is the physical thickness of the dielectric and $\varepsilon_{r,eff}$ its effective relative permittivity. Some of the most relevant requirements for DRAM capacitors include: high permittivity; physical thickness of the MIM film stack; the material has to be homogeneously deposited over large areas of production type wafers (the atomic layer deposition technique is the most suited); all the processes have to be compatible with CMOS process technology. Regarding the properties of high permittivity materials: the relevant property for the DRAM cell switching from the state '0' to the state '1' (or vice versa) is the **total charge difference, ΔQ_S**, on the cell capacitor: $\Delta Q_S = \frac{A_S \varepsilon_0}{t} \int_{-V_{DD}/2}^{+V_{DD}/2} \varepsilon_r(V) dV$ (an average permittivity obtained by integration). The temperature dependance of the low frequency permittivity of bulk STO obeys the Curie-Weiss law: $\chi'_e(T) \approx \varepsilon_r(T) = \frac{C}{T - T_0}$. The charge loss in DRAM capacitors originates from two different phenomena: a non-zero conductivity of the dielectric, which results in leakage currents, and dielectric losses, which can be characterized by dielectric relaxation currents. Generally, the dielectric relaxation currents in partially disordered or amorphous dielectrics follow the empirical Curie-von Schweidler relation, where the current density is $j_{relax}(t) = j_0 \left(\frac{t}{t_0}\right)^{-\alpha}$, $0.5 \leq \alpha \leq 1$. Usually, α is very close to 1. The leakage currents are due to electronic conduction through dielectric thin films and they become dominant towards longer times.

Physical limits of scaling (limitations to reduce the size towards an ultrahigh density memory): There are several challenges for the DRAM array, especially how to arrange all required elements of a DRAM cell on a footprint $\mathbf{4F^2}$ (four contacts and their isolations are necessary, therefore at least two contacts have to be stacked on each other), the maintaining of the **ON-current** of around 10 µA at V_{DS} =1V while still keeping the **device length**, **series resistances** (such as contact resistances) and **cross-talk effects**. The sense amplifier is becoming more and more critical since the signal margin (ΔV_{sense}) can realistically not be further reduced. Thus, the area gain from one to the next shrink generation is significantly cut, since the sense amplifier for $4F^2$ cells below 40nm covers already almost the same area as the cell arrays itself. The capacitor geometry of a thin film capacitor can have various shapes; for a pedestal structure with a cylinder of diameter *F* as bottom electrode post, the capacitance can be estimated *by* $C_S = \varepsilon_0 \varepsilon_{r,max} \frac{A_S}{t_{min}} = \varepsilon_0 \varepsilon_{r,max} \frac{\pi \left(X_{AR} + \frac{1}{4}\right) F^2}{t_{min}}$, where t_{min} denotes the minimal thickness of the capacitor dielectric and X_{AR} its aspect ratio. Many further simplifications need to be made to this expression. While approaching the minimum limit for thickness, R_{TE} tends to infinity. This results in an ultimate physical scaling limit at a technology node of F=12 nm. However, it is worth noting that the ideal material properties and geometries have been assumed in the estimation of the ultimate limit of scaling. In a more realistic approach, many constraints have to be taken into account. The thickness of the top electrode with a layer thickness of twice 1nm will probably lead to R_{TE} values far larger than those of the pure metal. In addition to the technological challenge to fabricate structures with very high aspect ratios, a realistic capacitor structure not perfectly vertical due to the limitations of the dry etch process and accordingly a small taper and bowing of the structure need to be accounted for. Additional limitations to the down scaling stem from the access transistor.

How the retention time can be calculated and what are the key physical parameters that affect the retention of a bit state: DRAM cells lose data because capacitors leak charge over time. The amount of time that a cell can store data before loss occurs is the cell's *retention time*. Typically, all DRAM cells are required to have one greater than 64ms. For DRAMs in certain conditions, the average time between refresh calls is 7.8µs. Since refresh logic refreshes each row at the same rate, they must be by every 8192nd refresh command, since 8192 x 7.8µs ≈ 64ms.

The retention time of a DRAM cell depends on the **leakage current** for that cell's capacitor and access transistor, as well has many other factors, two of which haven't been adequately explored by prior work: **data pattern dependence** and the **variability** of the retention time (VRT). In the first case, a retention failure occurs when too much charge has leaked away from a DRAM cell capacitor. However, the overall voltage change that occurs on the *bitline* is also affected by noise. The two major sources of this noise are *bitline-bitline* coupling (electrical coupling between adjacent lines) and *bitline-wordline* coupling. The second case refers to the fact that, even though the DRAM cell's leakage current is assumed stable over time, there are actually multiple retention time states. [2]

[2] https://www.pdl.cmu.edu/PDL-FTP/NVM/dram-retention_isca13.pdf

Reliability issues: In considering the reliability of the overall system, the **long-term stability** of all the devices, for example the capacitor, the transistor and the metal lines on a chip need to be evaluated. Only the reliability of the **cell capacitors** will be focused on. The steady-state leakage current regime is generally followed by a period of resistance degradation, resulting in increasing leakage currents and subsequently by a soft (slow) or hard (abrupt) **breakdown** (complete loss of the insulating properties). These phenomena determine the ***lifetime*** or ***breakdown time*** (and thus the reliability of the capacitor), which can be defined as the time at which the leakage current has increased by a factor of ten compared to the value of the steady state minimum. The benchmark for the lifetime of capacitors to be used is 10 years at V_{DD}≈0.5 V and 110ºC. Generally, the data for the lifetime have to be collected in accelerated tests (in order to avoid extremely long testing times) - that means at temperatures and/or voltages much higher than the benchmark values. The breakdown time can be extracted and statistically evaluated by plotting it versus the cumulative failure rate in a Weibull graph. A power law dependence as a function of the applied bias best fits the experimental results. Some types of breakdown events show a temperature dependance of the lifetime that can be described by Arrhenius characteristics, that is, by a thermally activated failure mechanism. Typically, the extracted activation energies for thin films vary between 0.2eV and 1.6eV; ZrO_2showed a low value of 1.0 eV. Overall, the energy depends on various parameters; it decreases with increased electrical field and with increased film thickness. Due to the complex functional dependance of the lifetime on many parameters, exact extrapolations to the benchmark condition are impossible and so only rough estimates can be given.

Example of manufacture companies for such a memory device: It was patented in 1967 and introduced into the market by Intel Corporation in 1972. It was the driving force for the exponentially growing large-scale integration in memory chips, while promoting similar advances in logic chips. The simple structure and functionality guaranteed the long-lasting success and growth of the market. There are only a handful of semiconductor manufacturers with the capability to produce DRAM chips, which include Micron (Crucial), Samsung, and Hynix [3].

[3] https://www.crucial.com/support/articles-faq-memory/truth-about-memory-manufacturers (Crucial)

3) Magnetic memories MRAM

Memory type: Resistive RAM (Matrix-based memory, Non-volatile RAM)

Design of a typical architecture of the storage element

In a **Stoner-Wolfarth MRAM (SW-MRAM)**, the write selectivity is achieved by combining two orthogonal pulses of magnetic field (see Figure 9). During writing, the selection transistor is open - no current flows through the MJT. During reading, the selection transistor of the addressed cell is closed, current flows through the MJT and the magnetic state of the memory point is derived from the measured resistance of the stack.

In a **Toggle MRAM**, a storage layer of a synthetic antiferromagnetic (SAF) layer is used. Such SAF consists of two magnetic layers separated by a thin non-magnetic layer of Ru. For proper choice of the Ru thickness (between 0.5nm to 1nm), a coupling (named RKKY coupling) exists between the two magnetic layers which is antiparallel and its strength can be adjusted by finely tuning the Ru thickness. Figure 10 shows a cross section of a MRAM cell. The MTJ is grown above the CMOS levels and electrically connected to the selection CMOS transistor by vertical via.

In a **Spin-Transfer Torque RAM**, spin-transfer torque is used to switch the magnetization of the storage layer; the memory cell is shown in Figure 11. The effective anisotropy may be increased by giving the cell an elongated shape (shape anisotropy of an elongated ellipsoid), but this works only up to aspect ratio of the order of 2. In MJT stacks with perpendicular-to-plane magnetization, an increased spin-torque efficiency is achieved by integrating in the stack two pinned layers on both sides of the storage layer, as seen in Figure 12. The two spacer layers separating the storage layer from the two pinned layers must have different resistances to avoid compensation of their magneto-resistive effects. It offers better thermal stability and reduced write current. Using two oppositely magnetized hard layers more than doubles the efficiency and balances the magnetostatic stray field created by each of the hard layers on the storage layer. The storage layer is separated from these hard layers on one side by a tunnel barrier, which can be of lower *RA* than the first one or metallic (Cu, for example). The Ti or Ta layers absorb the B out of the CoFeB alloys. It is important to attract the B away from the MgO interfaces during annealing because it would reduce the TMR and interfacial perpendicular anisotropy amplitudes.

As for the **TA-MRAM**, the storage layer material in the MTJ is chosen so that its magnetization switching field strongly decreases in the temperature range between room temperature and ≈200ºC. An exchange biased storage layer (i.e., a ferromagnetic layer exchange coupled to an antiferromagnetic layer) with a blocking temperature of the order of 200ºC can be used for this. Figure 13 shows an example of realization of the approach consisting in combining TAS with a pulse of magnetic field. Figure 13a represents the write procedure and figure 13c the MJT stack.

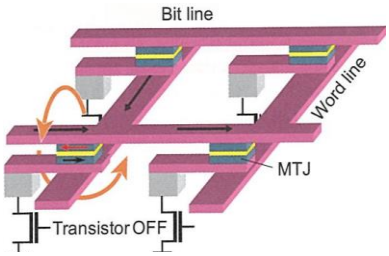


Figure 9: First generation of MTJ based MRAM using field induced magnetization switching (FIMS) as write scheme

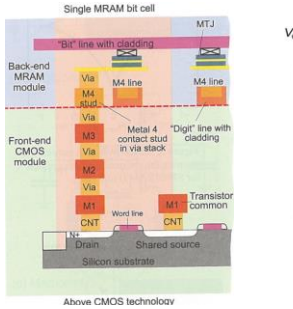


Figure 10: Cross section of a Toggle--MRAM cell showing the two stages technology: front-end CMOS module and above back-end magnetic module

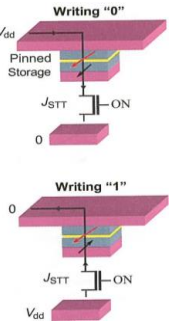


Figure 11: Memory cell written by spin-transfer

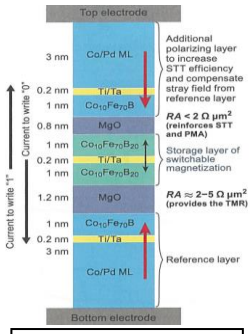


Figure 12: Improved perpendicular STT-RAM cell

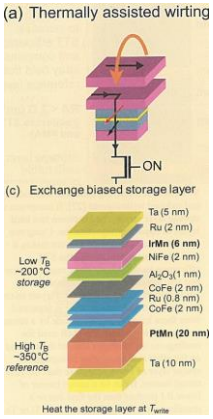


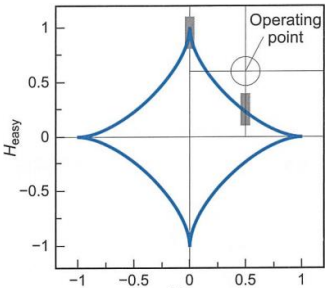
Figure 13: Write scheme combining heating with application of a pulse of weak magnetic field: (a) Write principle; (c) Typical stack used to implement this approach either with Alumina or MgO based tunnel barrier.

Physical operation principle and basic equations

Spintronics has more recently found applications in non-volatile memories (MRAM=*Magnetic Random-Access Memory*). In MRAM, each cell consists of a magnetic tunnel junction (MJT) connected in series with a CMOS selection device, most often a MOSFET. The information is encoded via the orientation of the magnetization of the MJT storage layer relatively to that of the MJT reference layer. The first generations of MRAM were based on magnetic field writing. The spin-transfer phenomenon provides a new write scheme in MRAM, yielding a much better scalability of these devices. Thermal assistance provided by the Joule heating around the tunnel barrier improves the MRAM performance.

One of the write schemes is called **Stoner-Wohlfarth approach (SW)**. The write selectivity is based on the combination of two perpendicular pulses of magnetic field (known as "Stoner-Wolfarth approach"). The SW model gives the relationship between the strength of the field required to switch the magnetization and the direction of the applied field. The angular dependence of the condition for switching of the magnetization of the storage layer is given by $\mathrm{H_X^{2/3} + H_y^{2/3} \geq H_K^{2/3}}$, where H_x and H_y are the fields generated respectively by the *word* and *bitlines* and H_k is the magnetic anisotropy field of the storage layer (see Figure 14). For a successful write operation, minimum power consumption and optimum write margins, the combination of H_x and H_y should fall in the area defined by the circle. Any appropriate combination of positive H_{hard} (*bitline* field) and H_{easy} (*wordline* field) encodes a '1'; with a negative H_{easy}, a '0' is encoded. In the operating region, application of either H_x or H_y alone would not result in data encoding.

Figure 14: Stoner-Wohlfarth asteroid showing the optimum operation region for SW-MRAM



The **Toggle MRAM** uses a synthetic antiferromagnetic (SAF) storage layer. In zero magnetic field, the tri-layer can have a quasi-null remnant magnetization. At a critical field – the spin-flop field (H_{sf}) - the two magnetizations that were antiparallel at low field suddenly rotate to be orthogonal to the applied field, while slightly scissoring in the direction of the field. The activation energy of the system in the half-selected configurations (i.e., when only H_1 and H_2 are applied) initially increases up to a field amplitude of ≈H_{sat} /2 (H_{sat} is the field necessary to saturate the two magnetizations parallel to each other), then it decreases to reach zero at ≈H_{sat}.

In **STT-RAM**, spin-transfer torque is used to switch the magnetization of the storage layer, instead of magnetic field. Writing a '0' is done by sending a current pulse through the stack. Writing a '1' is done by sending a pulse of current of opposite polarity. The critical current density for which the magnetization switches due to the STT, where *P* is the current polarization, *K* the uniaxial anisotropy of the storage layer, M_S its spontaneous magnetiza-

tion, α its Gilbert damping and t_F its thickness influence, is given by $J_{WR\ in-plane} = \left(\frac{2e}{\hbar}\right)\frac{\alpha t_F}{P}\left(\frac{\mu_0 M_S^2}{2} + 2K\right)$. To push further the superparamagnetic limit, MTJ stacks having perpendicular-to-plane magnetization are used. In **perpendicular MTJ**, the current density for switching can be very reduced; the critical current density is then $J_{WR\ out-of-plane} = \left(\frac{2e}{\hbar}\right)\frac{\alpha t_F}{P}(2K_{eff})$ (K_{eff} is the effective perpendicular anisotropy which takes into account the perpendicular anisotropy of bulk or interfacial origin minus the demagnetizing energy of the layer).

Thermally assisted writing in MRAM consists in combining a temporary heating of a memory bit with the application of a magnetic field or of a spin-transfer torque to select and write the magnetic bit. In order to write in a cell, the transistor of the selected bit is first set in passing mode to let a heating current flow though the MTJ. The inelastic relaxation of the hot tunneling electrons heats up the storage layer, enabling the switching of its magnetization. Then, either a weak pulse of magnetic field is applied or the same pulse of current used to heat is also used to exert a spin torque on the magnetization.

Physical limits of scaling (limitations to reduce the size towards an ultrahigh density memory): Despite the success of the Toggle-MRAM, all MRAM technologies based on field induced switching schemes are **poorly scalable**, because the **energy barrier $K_{eff}V$** between the two states has to be kept above $70k_BT$ to guarantee 10 years data retention, which implies either maintaining the magnetic volume of the storage layer (*V*) or increasing the effective magnetic anisotropy (K_{eff}). In both cases, the switching fields for SW write or the H_{sf} for Toggle write increases. As the **cross-sections** of the *bitlines* and *wordlines* **decrease**, the **current density** then **drastically increases**, up to the electromigration limit of the order of $10^7 A/cm^2$. Besides, the write power continuously increases, which makes these concepts not viable at small technological nodes. The Toggle-MRAM is not predicted to operate at nodes smaller than 45nm. One major advantage of the STT write approach is the down-size scalability that it provides (good scalability down to cell size of the order of 45nm). However, as the size decreases, the volume decreases and the relation $KV > 50k_BT$ (associated with the retention time) becomes more and more difficult to fulfill. The effective anisotropy may be increased by giving to the cell an elongated shape (shape anisotropy of an elongated ellipsoid), but this works only up to aspect ratio of the order of 2. Above, the switching proceeds by nucleation/propagation of domain walls rather than coherent rotation. Furthermore, trying to use materials with higher uniaxial anisotropy leads to a correlated increase in the Gilbert damping α - higher write current density, which is not good (large selection transistors and excessive electrical stress across the tunnel barrier). It is believed that the perpendicular STT-RAM (pSTT-RAM) could be downsize scalable to sub-20nm nodes. It may however be limited by the total current required to write the memory while maintaining a sufficient thermal stability. The total current flowing through the MJT is given by $I_{WR\ out-of-plane} = \left(\frac{4e}{\hbar}\right)\frac{\alpha k_B T \Delta}{P}$, where Δ is the thermal stability factor $\Delta E/k_BT$, which must be maintained above 50 for a single bit or even above 67 for a 32Mbit chip to insure a 10-year retention. To further improve the down-size scalability of MRAM, TA-MRAM were proposed. This method still needs further optimization but it allows a very compact cell design and would offer the ultimate scalability in MRAM as in magnetic recording technology.

How the retention time can be calculated and what are the key physical parameters that affect the retention of a bit state: For a storage application, it is of primary importance that, once written, the magnetization orientation of the storage layer remains stable for several years (typically **10 years** is chosen for criterion). This defines the retention of the memory. In magnetism, the characteristic time to switch above a barrier of height ΔE is given by an Arrhenius law of the form:

$$\tau = \tau_0 e^{\frac{\Delta E}{k_B T}}$$

Where $\tau_0 \approx 10^{-9}$s represents an attempt time, ΔE=KV (*V* is the volume of the storage layer and *K* its magnetic anisotropy per unit volume) and *T* the temperature. This implies that, to avoid accidental switching of the magnetization of the storage layer during 10 years in standby due to thermal fluctuations, the barrier height must fulfill the relationship $\frac{\Delta E}{k_B T} > \log\frac{10\ years}{10^{-9}s} \approx 50$, meaning $KV > 50k_BT$. As the size decreases, the volume decreases and the relation's more difficult to fulfill. In TA-MRAM, the dynamics of cooling and heating is quite fast. The heating time varies inversely proportional to the heating power sent through the MTJ. A good compromise between maximum heating current density and duration of heating phase yields typical heating duration of the order 3 to 5ns. The cooling time is determined by the material properties and is of the order of 10 to 20ns. As a result, the write cycle time can be of the order of 30ns.

Reliability issues: The SW write scheme is efficient as long as all the bits constituting the array have identical or very similar magnetic properties. The distribution of the anisotropy fields H_K over the array should be as tight as possible to avoid write errors. The main parameters ruling the distribution widths are the **uniformity of the chemical composition of the MTJ** and the **accuracy of the patterning process**. Any defect or dispersion in the shape of the MTJ immediately results in a broadening of the switching field distribution. Similarly, the correct control of the shape upon scaling is very critical to tightly control the switching distribution. If the switching field of some bits gets too close to the asteroid curve, its magnetic stability is reduced and it might be accidentally written due to thermal fluctuations. This problem is known as the ***half select instability*** and may also occur due to some irreproducibility in the switching process in the magnetic elements. Avoiding write errors requires very narrow distribution of the switching field, which imposes stringent conditions on the process of fabrication of the memory elements and prevents the realization of large memory arrays. In Toggle-MRAMs, when the shape of the bit is more elongated in the direction of the easy axis, the saturation boundary along *x* is reduced and the write margin is reduced. When the field excursion gets closer to H_{SF}, the risk of soft errors is increased as the activation energy falls to zero along the easy axis when $H=H_{SF}$. Below cell sizes of ≈45nm, STT-RAMs have issues with the thermal stability of the information written in the cell. In both STT-MRAMs and TAS-STT-RAMs (thermal assistance combined with STT, TAS+STT), the main issues concern **processing and reliability**. The reliability associated with resistance to electrical breakdown needs to be further improved.

Example of manufacture companies for such a memory device: Several large IC manufacturers (Freescale, IBM, NEC, Toshiba, Samsung, …), attracted by the potential of MRAM as a universal memory, rapidly entered the MRAM arena and started their initial developments. Toggle--MRAMs have been extensively studied by Motorola, Freescale and Everspin (the spin-off of Freescale that industrialized the MRAM).

4) Phase change memories PCM

Memory type: Resistive RAM (Matrix-based memory, Non-volatile RAM)

Design of a typical architecture of the storage element

Figure 15 shows the two technological implementations. Technologically useful phase change materials show pronounced differences between the **optical** and **electrical properties** of the **amorphous** and **crystalline phases**, which excludes metals. Dielectric materials are also excluded due to their very low absorption rate at typical laser wavelength and high resistances. $\mathbf{Ge_2Sb_2Te_5}$ is an exemplary **phase change material**. It is amorphous when deposited at room temperature, typically by sputter deposition. Heating it will lead to a phase transformation at the crystallization temperature, T_x, which is a function of the heating rate and is typically 150-160ºC. The crystal structure that is formed at this temperature is a distorted rock-salt structure with Te atoms on one sublattice, and Ge, Sb and about 20% vacancies on the other sublattice in random atomic placement. Figure 16 shows such a structure obtained by molecular dynamics calculations. For $Ge_2Sb_2Te_5$, the bandgap in the amorphous phase is ≈0.7eV and ≈0.5eV for crystalline phases.

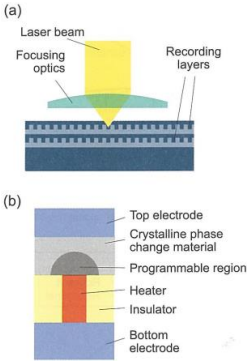


Figure 15: (a) Schematic of a dual layer Blu-ray recording disc. The recording is done in the grooves to isolate adjacent tracks; (b) Schematic of a PCRAM so-called mushroom cell, owing its name to the shape of the programmable region above the heater.

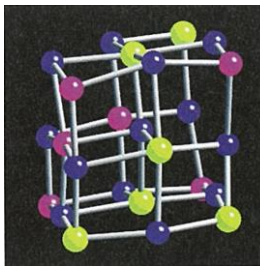


Figure 16: Crystalline structure of $Ge_2Sb_2Te_5$ obtained by molecular dynamics calculations. Te atoms are blue, Sb magenta and Ge yellow.

Most phase change materials are **p-type semiconductors**. The resistivity in the amorphous phase is a strong function of material composition and typically ranges from 1Ω·cm (e.g., Sb_2Te) to 10^4Ω·cm (e.g., N-doped $Ge_2Sb_2Te_5$). Phase change media have to fulfil five main data storage requirements, which can be translated to media requirements shown in **Table 4**.

Table 4: Requirements for phase change media

Storage requirements	Materials requirements	Material property for optical recording	Material property for PCRAM
Writability	Glass formation possible	Low melting point, high optical absorption	Low melting point, high resistivity of amorphous phase, threshold switching
Readability	Large signal/ noise ratio	Large optical contrast	Large electrical contrast, stable resistance values
Erasibility	Fast recrystal-lization	Simple crystalline phase, low viscosity, high optical absorption	Simple crystalline phase wit relatively high resistivity
Archival storage, operation at elevated temperature	Stable amorphous phase	High activation energy, high crystallization temperature	High activation energy, higl crystallization temperature
Cyclability	Stable material composition	Low stress	Low stress, no elemental segregation or electromigration

Physical operation principle and basic equations

Phase change materials can be rapidly and reversibly **switched** between the amorphous and crystalline **states**. Almost any material can be prepared in both, but, in phase change materials, the crystallization of the amorphous state is particularly fast and the relevant properties of the states differ significantly. This combination of properties is restricted to a group of semiconductor alloys mostly comprised of **chalcogenides** (selenides and tellurides) and several **antimony** compounds such as Ge-Sb and Ga-Sb based materials. The switching from the crystalline to the amorphous phase occurs during melting and by rapidly quenching the material so that it solidifies in the amorphous plane.

Phase change random access memory (**PCRAM**) relies on the **large difference in electrical resistivities between the two phases**. Switching and reading is done using electrical pulses. To amorphized, a high laser/current pulse with short trailing edge is applied for melt-quenching. To crystallize, longer and lower intensity laser/current pulses are used and even lower laser/current pulses measure the reflectivity/resistivity without causing any phase changes. To achieve high data transfer rates, materials should have fast phase transitions. A phase transformation can only proceed if it reduces the free enthalpy *G* (Gibbs free energy). The driving force of a phase transition is ***ΔG*** of the free enthalpy of the two phases. The velocity of the phase transformation will be strongly influenced by the height of the activation barrier, which has to be surmounted for the transformation to proceed. We have that G=H-TS (thermodynamics), where *H* is the free energy, *T* the temperature and *S* the entropy. Neglecting the temperature dependance of *S* and *H*, with T_m and T_G being the melting point and glass transition temperature (respectively), the **transition** from the crystalline to the liquid phase and from the amorphous to the crystalline phase have driving forces (respectively) of

$$\Delta G = \Delta H \frac{T_m - T}{T_m} \quad \text{and} \quad \Delta G = \begin{cases} \Delta H \frac{T_m - T}{T_m}, & T > T_G \\ \Delta H_{ac}\left[1 - \frac{T}{T_G}\left(1 - \frac{\Delta H\,(T_m - T_G)}{\Delta H_{ac} T_m}\right)\right], & T \le T_G \end{cases}$$

Where H_{ac} describes the exothermic energy of the transformation from the amorphous to the crystalline state. The equation for temperatures $T<T_G$ is of no relevance, since at these low temperatures the transformation proceeds extremely slowly. This is desirable, since amorphous bits could otherwise already be erased by thermal fluctuations below T_G.

To form a nucleus with a different phase, the total energy change is given by $\mathbf{\Delta G_{total}(r)} = \Delta G_V \frac{4\pi}{3} r^3 + 4\gamma\pi r^2$ for a spherical nucleus of radius *r*, where G_V is the free enthalpy per volume and *γ* the specific interface energy, and neglecting the elastic energy. For radii smaller than the **critical nucleus** $\boldsymbol{r_c}$, a further nucleus growth is unfavourable. For $r >= r_c$, a further growth is energetically more favourable than the decay. Nuclei with $r < r_c$ are denoted as ***embryos***, nuclei with $r > r_c$ as ***nuclei*** and nuclei with $r=r_c$ as ***critical nuclei***.

For the critical radius, ΔG_{total} reaches a maximum: $r_c = -\frac{2\gamma}{\Delta G_C}$ and $\Delta G_C = \frac{16\pi}{3} \cdot \frac{\gamma^3}{\Delta G_V{}^2}$. The negative sign in r_c reflects the fact that only negative values of ΔG_C lead to a phase transition; it represents an energy barrier that first has to be overcome for critical nuclei to form and is essential for the rate at which critical nuclei are formed. The number of nuclei with radius *r* and its critical value are given by $N(r) = N_0 e^{-\frac{\Delta G_{total}(r)}{k_B T}}$ and $N_C = e^{-\frac{\Delta G_C}{k_B T}}$, respectively. The growth rate is $U = \xi \cdot n \cdot \nu \cdot p \cdot B \cdot V_{atom} e^{-\frac{W_{21}}{k_B T}} \left(1 - e^{-\frac{\Delta G_{atom}}{k_B T}}\right)$, where *n* denotes the areal density at the interface, *p* the probability of a change to the other phase, *B* the probability for an atom to remain in the new phase, *ξ* considers the fact that not all areas of the interface can adhere atoms, *ν* describes the attempt frequency, while ΔG_{atom} is the difference of the free enthalpies (per atom) between the two phases. A maximum occurs at a higher temperature than the maximum for nucleation.

Physical limits of scaling (limitations to reduce the size towards an ultrahigh density memory)

The ultimate **limit** of PCRAM scaling is given by the size at which phase change materials **stop being phase change materials** - in other words, when they **lose data storage capabilities** because of scaling to small dimensions.

Phase change nanoparticles were reported down to sizes of even about **2nm**, stable at room temperature in the amorphous phase and crystallizing at size dependent temperatures which increased when particle size was reduced. This is the optimum scaling behaviour, because higher crystallization temperatures will lead to better stability of the amorphous phase. It appears that the physical limits, set by the fact that at least a few atoms are required to form a crystalline phase, are reached in the 2nm range, just about three times the lattice constant.

The increase of storage density is crucial to ensure the future success of optical recording. This is a serious challenge for optical recording. The improvements in storage density were achieved by moving to shorter wavelength and larger NAs. However, there are no new laser diodes available in the short term to move to even shorter wavelength and the numerical aperture of 0.85 is close to the possible physical limit. In the long term, new concepts for optical storage need to be developed - one of them is near-field microscopy. However, in this method, only a small photon flux is available for the characterization and modification of material and the distance regulation is quite slow. This is a further obstacle for efficient and fast storage and manipulation schemes employing such light sources. Both drawbacks of conventional near-field microscopes can however be circumvented by a different near-field light source design.

Retention time and what are the key physical parameters that affect the retention of a bit state [5]: The requirements for data retention at higher temperatures are much more stringent for automotive applications (150ºC for 10 years) or pre-coded chips that need to pass a solder bonding process (250-260ºC for tens of seconds). In these cases, phase change materials with much higher crystallization temperatures compared to $Ge_2Sb_2Te_5$-based alloys are required. Using PCM to replace DRAM is a big challenge, because **very fast switching times** in the nanoseconds range and **extremely high cycle numbers** of $\approx 10^{16}$ present a combination of requirements that **have not been achieved** by phase change materials. The high cycle number remains a problem, but it appears that scaling to smaller dimensions of the phase change material is beneficial for cycling.

The main failure mechanisms of PCM devices include **elemental segregation**, in particular Sb enrichment in the switching region caused by **electromigration**. This leads to poor data retention when the cell can no longer be switched to the high resistance state or does not remain in the high resistance state since Sb-rich alloys have **low crystallization temperatures**.

A higher crystallization temperature T_x of phase change materials means that unintentional crystallization is less likely, and thus a longer data retention time prevails. Measurements on $Ge_2Sb_2Te_5$ nanowire devices have shown that nanowires with smaller diameters have smaller activation energy, indicating a shorter data retention time.

The retention period refers to the maximum period up to which the information stored on the device is guaranteed to be retained without once powering up the device. It is measured by *artificial aging*. The devices are exposed to high temperature and voltage for months on end. Batches are removed from running and are tested at increasing time intervals. An appealing attribute of PCM is that the stored data are retained for a very long time (typically 10 years at room temperature), but is written in only a few ns. This property could enable PCM to be used for nonvolatile storage such as Flash and hard disk drives, while operating almost as fast as a high-performance volatile memory such as DRAM.

[5] http://poplab.stanford.edu/pdfs/Raoux-PCMreview-mrsbull14.pdf, https://www.sciencedirect.com/topics/materials-science/phase-change-memory (Science Direct)

Reliability issues

To fully integrate PCRAM into large memory arrays, billions of memory cells need to meet the requirements for data retention and write cycles with very low defect levels; 10^{13} cycles can be achieved with PCRAM cells. With **aggressive scaling of the PCRAM cells** to smaller and smaller phase change **cells and** contact size **cyclability** can **deteriorate**. There are two main failure mechanisms, so-called ***stuck SET*** **and** ***stuck RESET***. In the first case, the cell cannot be switched anymore to the high resistance state, while in the stuck RESET case it cannot be switch anymore to the low resistance state.

Stuck SET is often caused by a change in composition of the phase change material in the active, switching region. Electromigration and elemental segregation is an issue in PCRAM technology. Both effects can be very strong when the phase change material is in the molten state and a very high current density that can reach 10^7-$10^8 \mathrm{A}/cm^2$ is flowing though the molten material. It is often observed that the active switching area becomes rich in Sb. Sb-rich alloys have a low crystallization temperature, so even if the cell can still be switched, it will not retain the data in the amorphous phase.

The **stuck RESET** failure is often caused by void formation over the bottom electrode or delamination between the phase change material and the heater so that the electrical path is interrupted. Long and high RESET pulses lead to an earlier failure of the cells. In order to reliably reset all cells with the same RESET pulse, a large pulse is used to over-RESET many cells, leading to earlier failure.

Example of manufacture companies for such a memory device: The first memory chips with phase change materials as the storage media were introduced into the market in 2011. The global phase change memory market is forecast to reach USD 46.52 Billion by 2026. The proliferation of smartphones in the developing regions is the primary driving factor for the PCM market. Key participants include IBM, Micron Technology, Samsung Electronics, Hewlett-Packard, Toshiba, BAE Systems, STMicroelectronics, GlobalFoundries, Taiwan Semiconductor Manufacturing Company, United Microelectronics Corporation, Intel and Western Digital, among others. IBM is working on a three-bit per cell PCM chip, which is expected to provide more storage and stability than its previous research, which demonstrated 1 bit per cell options [6].

[6] https://www.globenewswire.com/news-release/2019/11/13/1946534/0/en/Phase-Change-Memory-Market-To-Reach-USD-46-52-Billion-By-2026-Reports-And-Data.html (GlobeNewswire)

Note: In order to prevent wasting space, the information in this homework was not presented in tables.

Bibliography: Apart from the links used for certain topics (indicated throughout this document), the materials available in the course's *webpage* were used to research about the topics at hand.