Introduction

Non-volatile memories (NVMs) are used to store data such that the data is held, or retained, even when power is turned off to the device. Traditionally the semiconductor implementation of such memories relies on a mechanism where electrical charge, representing the data value, is stored on a node, and is “trapped” there till the location is either erased (and likely re-written) or the charge has dissipated over a long period of time, typically in years, due to inherent leakage mechanisms. The memory is said to retain its data until such time as the stored charge no longer represents the correct data value associated with it when it was written.

Such charge-based traditional memory devices are known by several names based on the specific implementation of the underlying technology, e.g. Flash and SONOS. Depending on the functionality of the non-volatile memory, including data and control access, density and operation, such memories serve broader memory product classes like EEPROMs, NOR Flash and NAND Flash. In addition to stand-alone memory products, the same underlying charge-based technologies serve the purpose of embedded NVMs on systems-on-chip (SoCs). Figure 1 shows the structure of a flash memory cell based on floating gate technology.

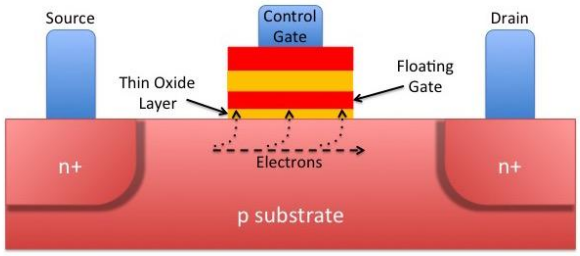

Figure 1. Flash memory cell structure based on floating gate technology.

Resistive-RAMs (ReRAMs) are a non-traditional type of NVMs in which the data is stored by way of modulating resistance between two electrodes by altering the physical dimensions of electrically conducting material between them. Figure 2 shows one kind of ReRAM cell, called a filamentary cell. In ReRAM technology program / erase cycles for the memory cause movements of electrically conducting atoms (ions) in a dielectric layer in-between

two electrodes, thereby altering the effective resistance between the electrodes. Circuits associated with the memory then interpret the resistance as a data value.

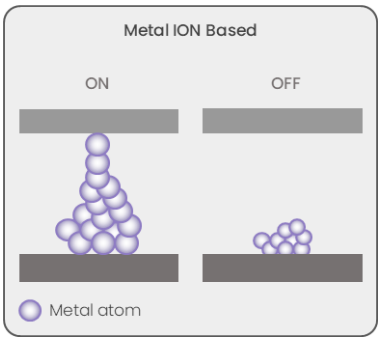

Figure 2. One type of ReRAM cell showing resistance modulation between the ON and OFF data states due to metal ion migration.

Manufacturing Complexity and Cost

The semiconductor manufacturing process is based upon creating patterned layers of

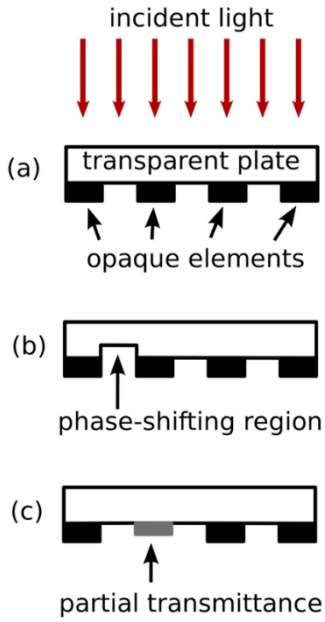

Figure 3. A schematic illustration of various types of masks: (a) a conventional (binary) mask; (b) an alternating phase-shift mask; (c) an attenuated phase-shift mask. Source: Wikipedia

different fabrication materials, typically one layer at a time. Patterning is accomplished through a process of deposition and/or removal of areas of every layer of material to match their own unique design. The pattern for each layer is represented by a mask (or photomask). Using a step called photolithography, each mask, unique to each layer, is used to create the desired pattern for that layer by shining a light through it that projects its pattern onto a material called photoresist on a silicon wafer (see figure 3). Masks define

areas where fabrication materials get deposited and/or removed on silicon based on their patterns. The number of masks used in a semiconductor fabrication process thus represents the complexity, and therefore the cost associated with the overall process.

Since NVM is usually an added feature of any baseline semiconductor process (one without NVM) the incremental cost added to the process depends on the number of additional masks required to implement NVM. **Flash memory, typical of broadly used NVMs, can add 10 or more masks for its implementation. As mentioned in this article in Embedded.com, this adds 20-25% to the cost of a silicon wafer, while embedded ReRAM technology requires only two additional masks for a cost-adder of about 10% to a silicon wafer**. This provides a significant improvement in the overall complexity and cost of the process. **A simpler process generally allows for fewer failures and a more robust product.**

Data Retention

There are inherent mechanisms associated with NVMs that cause data loss when the data is held beyond a certain duration of elapsed time. This period is called **retention** (or **data retention**), and it specifies how long data may be expected to be held as written into the memory. Retention for commercial grade memories can range roughly from a year to ten years, though that number is quite sensitive to environmental factors, particularly temperature. Quite obviously, the higher the retention for an NVM the better it is considered to be. Retention becomes critically important in applications where the data being held in memory is mission critical and therefore non-compromisable.

As mentioned previously, data retention for NVMs is very sensitive to temperature, and it degrades rapidly as temperature rises. This degradation is based on the concept of **activation energy,** usually denoted by **Ea** and expressed in **electron volts (eV).** In the context of semiconductor reliability, it refers to the minimum amount of energy required to trigger a temperature accelerated failure mechanism such as a retention failure in NVM. This EESemi writeup provides additional details.

The **Arrhenius Equation** provides the relationship between the rate at which a failure mechanism occurs, the activation energy of the mechanism, and the temperature according to:

$$R = Ae^{(-Ea/kT)}$$

where **R** is the rate at which the failure mechanism occurs, **A** is a constant, **Ea** is the activation energy of the failure mechanism, **k** is Boltzmann's constant and **T** is the absolute

temperature at which the failure mechanism occurs. Using the equation to determine the acceleration for retention failure rate for NVM at a higher absolute temperature (T_{high}) compared to a lower absolute temperature (T_{low}) we get the acceleration factor as:

$$F = e^{Ea/k(1/T_{low} - 1/T_{high})}$$

As the above equation suggests, when activation energy is a positive number, failure rates increase with increasing temperature, and the higher the activation energy, the more rapid the rise in failures with temperature. Another way to look at this phenomenon, since NVMs are typically qualified at elevated temperatures (frequently as high as 125C or 150C), if both incumbent NVM and ReRAM have identical retention specifications at elevated temperatures, then the failure rate of ReRAM will be different than the incumbent for lower temperatures, particularly as temperatures drop to a range more in line with human environments, if ReRAM activation energy is higher than incumbent NVMs. For NOR flash memory data retention activation energy is typically around 1.0eV (see NXP Engineering Bulletin for specific published numbers as examples). For ReRAM, the activation energy is typically around 1.5eV (see IEEE IEDM publication for a specific published number as example).

To illustrate the point above, if we have Flash and ReRAM NVMs, both qualified for a 10-year retention specification at 125C, **the retention for Flash with activation energy of 1.0eV would be 70 years at 100C, while the retention for ReRAM with activation energy of 1.5eV would be 188 years at 100C. ReRAM provides superior retention at temperatures in line with human environments.**

Radiation Tolerance

Data storage on charge-based NVMs makes them susceptible to radiation, since radiation can create mobile electron hole pairs that can dissipate the stored charge as shown in Figure 4. When the charge is dissipated, the data state associated with the storage node is no longer representative of the data that was originally written to that node as shown in Figure 5. This puts restrictions on how much radiation charge-based memories can be exposed to without losing data.

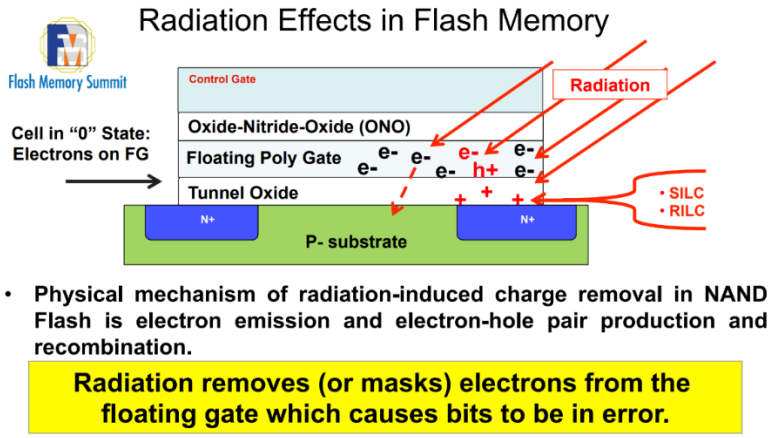

Figure 4. Radiation induced charge loss in Flash.

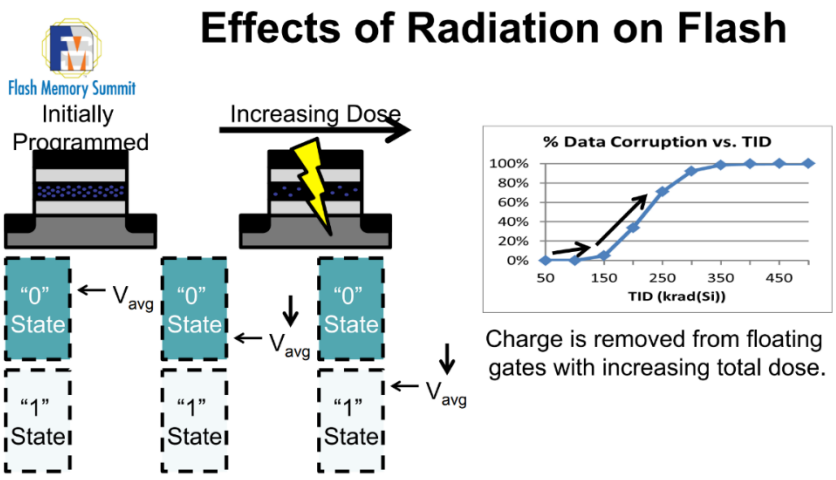

Figure 5. Charge loss leading to data corruption in Flash technology.

As described previously, ReRAM cells store data by modulating the physical dimensions of electrically conducting material in a dielectric film in-between two electrodes, providing different resistance values representing different data states. These materials remain virtually unaffected in the presence of radiation, thereby providing extremely high tolerance to radiation in terms of data retention.

Radiation that can cause failures in charge-based NVM consists of two broad components: charged particles like alpha particles (ionized Helium atoms), and electro-magnetic radiation like X-rays. In general, the earth's magnetic field provides protection against charged particles, while the atmosphere provides protection against X-rays, so the higher up in atmosphere we go, the more prone charge-based NVM becomes to retention losses due to the amount of radiation naturally present at higher altitudes.

At the terrestrial level, natural radiation is at a low enough level to be able to cause retention losses in traditional charge-based NVMs. However, there are human-made cases of radiation generation that can lead to loss of data in charge-based NVMs. These include nuclear events, somewhat of a rarity, but they also include X-ray sources used for electronic board manufacturing quality checking. Additionally, some semiconductor packaging materials are a source of alpha particles that can also cause retention failures in charge-based NVMs.

A commonly used unit of radiation measurement is the rad, which signifies the amount of radiation energy absorbed by a body per unit mass: 1 rad = 0.01 Joules / kg. **Some charge-based NVMs can start failing for retention at radiation levels of 25krad**. This radiation level is prevalent at altitudes deemed to be in the low-earth orbit (LEO) range. **ReRAM NVM, on the other hand, can generally tolerate radiation levels in the hundreds of krad, making it suitable even for deep space missions, well beyond LEOs.**

In summary, traditional NVM technology is based on charge-based storage. This makes it prone to loss of data upon exposure to radiation. Data retention in ReRAM is virtually unaffected in the presence of radiation levels experienced even during deep space missions. Additionally, there are human-made radiation sources that may go beyond the threshold of radiation tolerance for charge-based technology while not affecting ReRAM NVM. **Where security and integrity of stored data is mission-critical, ReRAM provides extra safety and insurance against radiation-based data loss.**

Security

There are two broad aspects to the overall security of NVM, namely: 1) intrinsic failures in retaining data such as those explained in earlier sections; and 2) resistance to malicious external attacks specifically targeting either gaining access to critical data stored in it or destroying the data altogether. This section will focus on (2).

There are several characteristics of charge-based NVM technologies that are exploited to infer critical data stored in them.

Direct Charge Measurement

This publication on direct charge measurement for Flash technology using a Scanning Electron Microscope (SEM) involves backside sample preparation and Passive Voltage Contrast techniques. These techniques result in successful extraction of Flash memory contents as described in the paper. However, **these techniques do not apply to ReRAM since there is no charge associated with data stored in ReRAM.**

Photon Emission Analysis

When transistors, including those that are the foundational charge-storing building blocks of Flash memory, switch states, they emit photons in a phenomenon known as photon emission. These emissions can be observed using Photon Emission Analysis (PEA). This reference demonstrates how PEA is employed to retrieve data from Flash memory embedded in a microcontroller. Once again, **since ReRAM technology does not rely on charge-based data storage, it is immune PEA based data detection.**

Electron Microscopy

Figure 6 shows the structure of a ReRAM cell and how it integrates within layers of metal in a CMOS process.

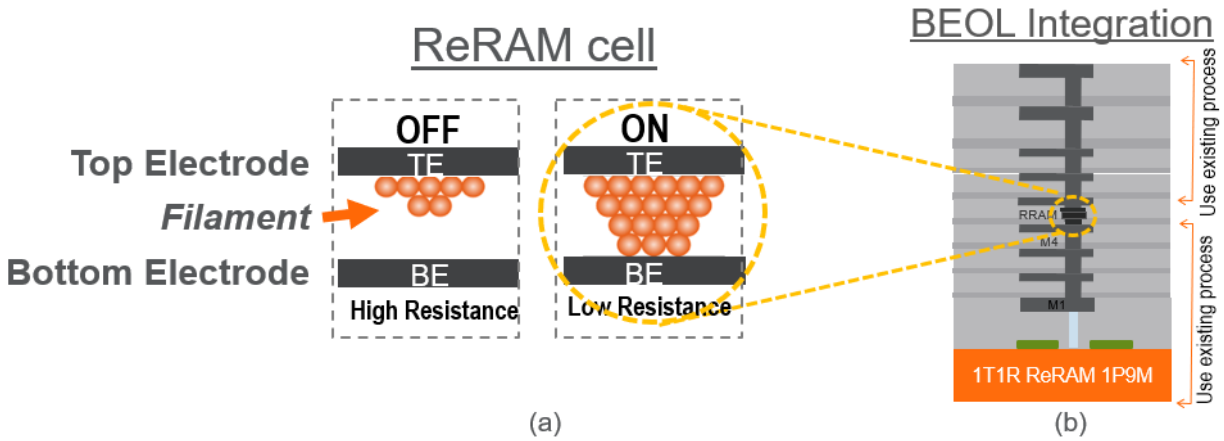

Figure 6. (a) ReRAM structure depicting ON and OFF states; (b) how the cell integrates in-between layers of metal in a 9-layer CMOS process

The dimensions of the filament in a ReRAM are on the order of a nanometer in both width and length, so it is quite small in comparison to other critical baseline CMOS process features, which, in a 22nm process for example, would be on the order of tens of nanometers. The small physical dimensions of the ReRAM coupled with even smaller atomic level differences in the structure depending on the data state stored in it make it virtually impossible to "read" the value of a cell even with advanced high-resolution imaging techniques like Transmission Electron Microscopy (TEM).

As Figure 6 (b) illustrates, the ReRAM structure is fabricated in-between layers of metal, so not only is it physically isolated from malicious attacks from the top or bottom sides of silicon, it is also shielded from basic imaging techniques like Scanning Electron Microscopy (SEM).

Power Analysis

Unlike charge-based NVM technologies, ReRAM does not require a mass erase of cells (generally a block or a page) prior to a write operation on the targeted cells, typically a byte or a few bytes. Additionally, the write voltages in ReRAM (typically in the 3V range) are lower compared to write voltages in charge-based memories (typically 10V+). This combination of technology features makes for far reduced write power in ReRAM compared to charge-based NVMs. Read operations also require lower voltages in ReRAM technology compared to charge-based ones, and coupled with lower data access latency, the combination of lower voltage and lower latency significantly lowers energy requirements for read access to a unit of memory. Various aspects of, and conclusions based on, ReRAM technology, including power compared to other prevalent NVM technologies can be found in this Springer article.

One of the most common forms of side-channel attacks targeting malicious access to secret NVM data in secure applications is Power Analysis. It involves analyzing power consumption measurements from the system while it performs operations using secret keys or other secure data. Since the power consumed by ReRAM is much lower compared to charge-based NVM technologies, **it is much harder to infer valuable data states and transitions through simple or even differential power analysis, making it far more secure against motivated hackers.**

Data Destruction using Laser Fault Injection

Unlike previous sections where malicious hackers target gathering sensitive data held in NVM, some motivated hackers may seek to destroy critical data altogether.

This article presents a laser-based fault injection attack on the Flash memory of a microcontroller in which an attacker can set individual bits of the words fetched from Flash memory in a very predictable manner. Such attacks leverage floating-gate transistor behavior in the presence of a laser spot illuminating specific areas of the Flash cell. As before, **this technique does not apply to ReRAM since its fundamental mechanism for data storage is non-charge based.**

In conclusion, not only is ReRAM resistant to threats targeting inference of sensitive data stored in it through a variety of means including use of invasive and state-of-the-art tools, but it is also far superior in mitigating malicious attacks to destroy its contents using lasers for fault injection.

Validation through independent third party

To provide an experimental means of validating the extent of security enhancement provided by ReRAM, CrossBar engaged with a contracted third party, MicroNet Solutions, Inc., and gave them the challenging task of inferring the contents of a ReRAM array using any equipment available to them, including means of sophisticated semiconductor de-processing, electron imaging, focused-ion-beam milling and power analysis tools. They were free to use invasive and even physically destructive techniques to infer the contents of the ReRAM, but they were unable to do so per details in this press release.

Conclusion

Incumbent NVM technologies, like Flash, are charge-based in nature. Since data states are stored in the form of charge on an electrical node, such technologies are prone to retention loss as the stored charge is dissipated in time. Retention loss can also be triggered by radiation from terrestrial and extraterrestrial sources. Consuming more operating power than ReRAM causes them to be more prone to common side-channel attacks based on power analysis targeting sensitive data stored in the NVM. Direct charge measurement and photon emission analysis are also means of inferring data held in charge-based NVMs. Finally, techniques like laser fault injection can provide malicious parties the means to destroy critical content stored in the NVM.

ReRAM technology is based on a simple structure that is also simpler to manufacture than incumbent NVM technologies. There is no charge associated with a data state, and values are stored by way of modulating resistance between two electrodes by altering the physical dimensions of electrically conducting material between them. Without the need for stored charge, and leveraging the benefits of a physically small, simple structure buried in-between metal layers in a silicon process, ReRAM offers superior data retention at temperatures expected in human environments, outstanding protection against radiation and inherent protection against means to steal sensitive data, including mitigation against malicious data destruction.