Chapter 2
Charge-Trap-Non-volatile Memory
and Focus on Flexible Flash Memory Devices

Konstantina Saranti and Shashi Paul

Silicon nanostructures are discussed in this chapter. First, the field of nanotechnology in relation to the properties and basic physical phenomena of nanomaterials is presented. Then, the current status of nanomaterials used for memory storage devices is reviewed. Vapour–liquid–solid method is described in details as the most common techniques to fabricate silicon nanostructures. In addition, a brief discussion on various metals for the role of the catalyst material is provided.

This chapter also familiarises the reader with the subject area of flexible electronic flash memory devices. The structure of the flash memory device, the idea behind its development and the basic operating principle are discussed. Silicon nitride is considered, for flexible flash-type memory devices, as an alternative for the dielectric material as it could solve a few of the current memory challenges. Lastly, a brief history and a critique of the current position of the flexible electronic flash memory devices are provided.

2.1 Nanotechnology and Nanomaterials

Nanoscience and nanotechnology are among the greatest emerging scientific areas in the last 50 years. The development of the next generation of electronics is partly established on the utilisation of nanomaterials. Although the origin of the use of nanomaterials goes back to the ninth century BC where nanoparticles were used as part of the surface treatment for decoration applied on ceramics [1], it is only after the early 1990s that the sophisticated equipment such as atomic force microscope [2] became available to view the nanoworld. This was one of the leading steps in
the foundations and evolution of nanotechnology. In 1986, Gerd Binnig and Heinrich Rohrer were awarded the Nobel Prize for their design of the scanning tunnelling microscope [3], which gave the fundamentals for the development of atomic force microscopy technique.

On 29th December 1959, Richard P. Feynman made a lecture to the American Physical Society titled “There’s Plenty of Room at the Bottom”. Feynman shared his vision about manipulating and controlling things on a small scale. An article was published following that talk and it is believed to have established the preliminaries of the history of nanotechnology [4].

In 1974, the term ‘nanotechnology’ was used for the first time by Norio Taniguchi [5]. As a general rule, nanostructures or nanomaterials are defined by having at least one dimension in the nanometre scale (from 1 to 100 nm). One metre is equal to one thousand million \(10^9\) of a nanometre and the prefix “nano” means dwarf in Greek. NASA suggested a more thorough description of the nanotechnology field as follows: “The creation of functional materials, devices and systems through control of matter on the nanometre length scale (1–100 nm), and exploitation of novel phenomena and properties (physical, chemical, biological) at that length scale” [6].

Nanomaterials can be classified according to the number of dimensions that have not been reduced below 100 nm (see Fig. 2.1).

- Zero-dimension (0D): The size of the materials is reduced to nanoscale in all three directions. Common examples of this category are the nanoparticles, nanoclusters and nanocrystals.
- One-dimension (1D): These nanomaterials have dimensions at nanoscale in two directions. Nanowires, nanotubes, nanofibers and nanorods are included in this category.
- Two-dimension (2D): The size of the materials is reduced to nanoscale in one direction. Examples are the nanofilms, nanocoatings and nanosheets.
- Three-dimension (3D): Materials with dimensions beyond the nanometre scale in any direction.
As soon as the novel properties of the nanomaterials were understood, their applications expanded in all the fields of science. The applications of nanoscience and nanotechnology are numerous and extend from the fields of coating, energy sources, environmental, electronics, to health and medical treatments. The commercially available nanomaterials are nanocoatings and nanocomposites, which find use in everyday products such as cosmetics, sunscreens [7], sunglasses [8], sports equipment and self-cleaning windows [9]. Nevertheless, there is still a room for many more potential applications to be discovered.

2.2 Properties of Nanomaterials

Nanomaterials offer significant differences in their properties in comparison to their bulk materials. The main two reasons are the high surface area to volume ratio and the small dimensions giving rise to quantum effects. As materials becoming smaller in size, their surface area to volume ratio starts increasing and as a result, the atoms on the surface have a more dominant behaviour than those in the interior of the crystalline lattice core. Hence, the surface atoms have more energy, which makes them more reactive, either chemically or mechanically, in interaction with other materials. This phenomenon is responsible for the catalytic properties of the nanoparticles [10]. Furthermore, the quantum confinement effect is observed for materials at the nanometre scale. The energy levels of materials are arranged in bands (conduction and valence). For insulators and semiconductors these bands are separated by a forbidden region, called bandgap (Eg), in which there is no solution of the Schrödinger’s equation (see Fig. 2.2). The quantum size effects have an impressive correlation between the bandgap and the size of the nanostructure.

Fig. 2.2 The diagram represents the energy bands in metals, semiconductors and energy levels in nanomaterials. For nanomaterials the size quantization effect is shown
Particularly, the bandgap increases as the nanostructure diameter decreases being smaller than its bulk exciton Bohr radius [11, 12]. Moreover, the quantum confinement in semiconductors leads to a ladder of discrete energy levels rather than energy bands [13] as shown in Fig. 2.2. In other words, as the dimensions of the structure become smaller, the energy levels are quantised leading to a spectrum of discrete energies. Although the quantum confinement effect has been established and explored for 0D nanomaterials, the effect has also been confirmed for 1D and 2D nanomaterials. However, different dimensionality of confinement is occurred for each case. In the case of quantum wires, quantum confinement is weakened relative to that in quantum dots [14].

Consequently, by changing the nanocrystals size, the energy gap of this material can be adjusted. This also results in changing the density of electronic energy levels. The density of states (DOS) is the number of available states per unit energy in the band structure (D(E) = dN/dE) and it is a fundamental material’s characteristic that determines many of its properties. Figure 2.3 presents the idealised distribution of states as a function of energy of 3D, 2D, 1D and 0D semiconductor structures [15].

Considering these phenomena, novel optical, magnetic, electrical, and other properties emerge at nanoscale. At this point, only a brief description of a few fundamental paradigms is presented.

The effects of quantum confinement predominantly influence the electrical properties, such as electrical conductivity and photoconductivity due to their high dependence on the energy state and DOS. Still, it is broadly reported that the shape can have a dramatic influence on the optical properties of nanostructures [16–18]. Figure 2.4a shows the fluorescence emission of (CdSe)ZnS quantum dots of various sizes. As the size of the dot gets smaller, both the optical absorption and emission of quantum dots shift to higher energies (blue) [17]. This is due to the confinement effect, in which a decrease in size leads to a larger bandgap. Moreover, Fig. 2.4b represents the absorption spectra of various sizes and shapes of gold

![Fig. 2.3 Schematic representation of the density of states as a function of energy exhibited in 3D, 2D, 1D and 0D semiconductor structures](image-url)
nanoparticles. When gold nanoparticles are enlarged a very similar absorption is observed, while other shapes such as nanorods give a significantly different absorption spectrum [18]. The combination of unique electronic and optical properties offers compelling products such as lasers and light-emitting diodes (LED) using either quantum dots or quantum wires.

Nanotubes are widely known for exhibiting superior mechanical properties. The bonding structure of nanotubes is extremely strong, enabling them to be fabricated with unique strength and toughness properties [19]. Bulk modulus of 462–546 GPa have been reported for superhard phase single-walled carbon nanotubes which exceeds even that of diamond (420 GPa for a single diamond crystal) [20]. Thus, nanotubes are ideal for lightweight construction, for example, in automotive and aerospace industries. Apart from using nanotubes as the fundamental materials, another common use is advanced filler materials in composites. Both single and multiwalled nanotubes enable enhancement to material mechanical properties for reinforced polymer fillers their fabrication can be very simple. The nanotubes and the polymer are mixed in a suitable solvent and once the solvent is evaporated a composite film is formed [21]. Hence, filling polymers with nanotubes or nanoparticles lead to improvement in their mechanical properties.

The capability of high sensitivity, due to high surface to volume ratio, has led to the development of chemical sensors consisting of single-walled carbon nanotubes [22]. For example, gas sensors based on ZnO nanowires show a very high sensitivity to ethanol gas [23] and nanowire biosensors enable the detection of diseases such as Alzheimer’s or diabetes at an early stage [24]. Moreover, nanomaterial properties have led to improvements in drug delivery systems and cancer therapies, by using better materials to encapsulate and release the drugs [25, 26].

Fig. 2.4  a Fluorescence emission of (CdSe) ZnS quantum dots of various sizes and b absorption spectra of various sizes and shapes of gold nanoparticles. Reproduced with permission from The Royal Society of Chemistry [18]
2.3 Coulomb Blockade and Single-Electron Tunnelling

The combination of quantum tunnelling phenomena (described in Chap. 3) and the Coulomb charging energy gave rise to the function of single-electron transistor (SET) [27, 28]. By exploiting the single-electron charge transfer, electrons can tunnel, one by one, into and out of the nanostructures. This is the most studied nanoparticle-charging phenomenon and is called Coulomb blockade. There is a common electrode termed island (quantum dot), which is connected through tunnel junctions to two electrodes (the source and the drain) and through a capacitor to a gate electrode. Under the application of a constant gate voltage on the island, only one electron can transfer from source to drain via island, enabling the ability to control the transfer of individual electrons. In order to observe single-electron tunnelling phenomena or Coulomb blockade effects, two conditions must be satisfied. Firstly, the charging energy must exceed the thermal energy (at room temperature $E \gg 25.7$ meV) as described by Eq. (2.1). Secondly, the Heisenberg uncertainly, stated in Eq. (2.2), sets the lower limit of the tunnel resistance (25.813 kΩ).

$$\frac{e^2}{2C} \geq k_B T \quad (2.1)$$

$$\Delta E \Delta t \geq \frac{\hbar}{2} \quad (2.2)$$

where $\Delta t = RC$ the time to charge/discharge of an island, $\hbar$ is Planck’s constant, $e$ is the electron charge, $k_B$ is Boltzmann constant, $T$ is temperature and $C$ is the capacitance given by Eq. (2.3).

In the case of the nanoparticle, Eq. (2.3) gives the capacitance in which the nanoparticle is considered as a conducting sphere.

$$C = 8\pi\varepsilon_0\varepsilon_r r \quad (2.3)$$

where $\varepsilon_0$ is the permittivity of free space, $\varepsilon_r$ is the relative permittivity and $r$ is the radius of the sphere.

The voltage required to transfer an electron to the nanoparticle is:

$$V = \frac{e}{C} \quad (2.4)$$

The energy (in joules) required to charge the nanoparticle is:

$$E_{(j)} = \frac{e^2}{2C} \quad (2.5)$$
2.4 Current Status of Nanomaterials Used for Memory Storage—The Charge Can Trapped in Their Available Energy States

There is a plethora of applications based on the exceptional characteristics of nanomaterials and wide ranges of fields have shown to utilise them. In the scope of this chapter, only the recent applications of nanostructured materials for memory-based electronic storage devices will be covered. High-density information storage devices using nanomaterials are a fast-developing area, which competes with the conventional silicon-based materials.

Memristors (an abbreviation for memory resistor), using titanium dioxide sandwiched with platinum wires, were successfully accomplished at Hewlett Packard’s labs. Leon Chua theoretically predicted the memristor in 1971 as the fourth fundamental circuit element, the other three being resistor, capacitor and inductor [29]. Nevertheless, it was 37 years later the mathematical model and proof of existence was shown. R. Stanley et al. demonstrated that memristance arises naturally in nanoscale systems, by using metal nanowires coated with titanium dioxide and arranging them in crossbar arrays forming a memristor at each intersection. The device’s resistances change over time as a function of their existing state [30]. Recently, Chua also argued that all two-terminal non-volatile memory devices based on resistance switching phenomena are memristors, in spite of their material or physical mechanism [31]. It is believed that the development of high memory density memristors could lead to the substitution of the commonly used dynamic random-access memory (DRAM). Commercial availability of the memristor is estimated for 2018 [32].

It is widely known that magnetic materials have been efficiently used through the years as digital data storage. The evolution of nanomaterials and the need for miniaturisation of the memory storage units in the electronic devices led to the exploitation of magnetic nanomaterials. IBM researchers developed a new type of magnetic memory called racetrack memory in which magnetic nanowires made of an alloy of iron and nickel are utilised [33]. Racetrack memory store bits of data in the form of magnetised regions or domain in nanowires on silicon substrate. The nanowires can be arranged horizontally or vertically on the silicon chip. The racetrack memory is based on the ability to use an electric current to move magnetic domains along the nanowire and can read the bits as they go past the reading element [34]. As in all magnetic memories, magnetism plays the key role in the physical operating mechanism. A working prototype with all the necessary components on the single chip was presented at the International Electronic Devices Meeting in Washington, in December 2011 [35].

Researchers at Rice University demonstrated that resistive switches and memories could be built solely from silicon dioxide. A two-terminal resistive switch with 5 nm diameter silicon dioxide nanowires sandwiched between semiconducting sheets of polycrystalline silicon for top and bottom electrodes were studied to create memory devices by changing the resistance of the nanowire at that location. Silicon dioxide is
proven to be the source for the formation of silicon nanocrystals, which can be switched between conductive and non-conductive state without damaging material’s properties [36]. The simplicity of the structure, the greater compatibility (fully CMOS compatible) and 3D capability are very promising characteristics for the production of high-density memories using this approach. Moreover, only a few years later, a highly transparent memory with indium tin oxide or graphene as the electrodes and silicon dioxide as the active material was fabricated by the same research group [37].

Another approach is to store information on magnetic nanoparticles called nanodots or nanoscale magnets. Each grain is individually magnetised in its own direction and the working principle is based on switching their magnet states and reading them [38]. Research is undergoing into reducing these structures to their nanoscale size limit in which they can retain their magnetism and in consequence the memory effect. To accomplish this, highly orientated nickel magnetic nanoparticles have been studied [39]. The structures are self-assembled giving the advantage of precise arrangement with minimum energy consumption and maximum stability production. Bits are stored by controlling the orientation of each nanodot composed of magnetic nickel alloy. Developing the technique they are claiming a terabit of data could be stored on a chip of area $6 \times 6 \text{mm}$ [40].

A nanoparticle encapsulated within a multiwall carbon nanotube has been observed to function as an electromechanical memory device. In particular, an iron nanoparticle inside a carbon nanotube can be controllably positioned and creates the binary states 0 or 1. Researchers believe that via an electrical write signal, the shuttle can move reversibly and have a thermodynamic stability that can exceed one billion years [41].

### 2.5 Synthesis of Silicon Nanowires

Silicon nanowires are 1D nanostructures and their use has already been established in different applications such as sensors [24], high-performance field effect transistors [42], solar cells [43] and memory devices [44]. There are many different methods to fabricate silicon nanomaterials. In general, they can be categorised into two main groups: the bottom-up approaches (i.e. laser ablation, chemical vapour deposition and thermal evaporation) and the top-down approaches (i.e. nanolithography, chemical and reactive-ion etching). In the former method, by the effective combination of the fundamental blocks the desired characteristics are achieved providing a great advantage of potential self-assembly ability and no need for patterning. In the latter case, the additional material is removed forming the final device or structure. Although nanolithography techniques can design well-patterned areas, resulting in well-oriented nanostructures, they are limited due to high cost and low yield.

To the best of the author’s knowledge, Treuting and Arnord have published the first paper on silicon wire growth in 1957 [45]. At that time, the term whisker was used. Seven years later Wagner and Ellis presented the well-known vapour–liquid–
solid (VLS) growth mechanism of silicon whiskers [46], which was a stepping-stone for progress in the research field of the synthesis of silicon nanostructures.

VLS is the most common growth mechanism applied to synthesise variety of nanomaterials through chemical vapour deposition (CVD), thermal evaporation oxide-assisted growth (OAG), laser ablation, solution-phase synthesis and molecular beam epitaxy (MBE) [47] techniques. It is worth mentioning that each technique has advantages and disadvantages. However, depending on the required application each technique can be applied effectively. The amazing feature of the VLS mechanism is that it works well over a range of material sizes from hundreds of micrometres to a few nanometres.

For the CVD process the precursor gas used is usually silane (SiH₄) or silicon tetrachloride (SiCl₄). It is a bottom-up synthesis method that allows the epitaxial growth of the silicon nanostructures. The use of metal seeds as catalyst material is essential and the main constraint of the mechanism, because contamination is possible due to metal diffusion into the grown structures, which can affect the nanostructures’ properties. With the appropriate choice of catalyst and deposition conditions, well-defined nanomaterials can be grown. For lower substrate temperatures and wider range of metal catalysts, plasma-enhanced chemical vapour deposition (PECVD) method is used. The temperature for the CVD process ranges from 500 to 800 °C setting a limit for the use of temperature-sensitive substrates as well as several metal catalysts. The plasma-enhanced CVD process provides high ion energy allowing the creation of reactive gas species at lower temperatures ranging from 100 to 300 °C. Hence, if a low-temperature synthesis is required the PECVD method is preferred.

### 2.5.1 Vapour–Liquid–Solid Growth Mechanism

As stated previously, Wagner and Ellis described the formation of single crystal silicon whiskers by VLS mechanism in 1964. This was the leading step for the future of the synthesis of silicon nanowires. The observed structures were silicon crystal whiskers with liquid droplets at the tip. Gold (Au) was used as the catalyst material and a mixture of SiCl₄ and H₂ for the process [46]. Until today, these factors (Au, SiCl₄ and H₂) are the most favourable for the VLS growth of nanomaterials. The part of the metal seed as catalyst is essential for the VLS growth. Initially the substrate was heated up. The vapour Si (available from the source gas SiCl₄) was decomposed on the surface and diffused into the metal. Therefore, a liquid Si–metal alloy was formed and once it became supersaturated the nanowire growth started. As the name states there are three phases involved in this process; vapour, liquid and solid. These three phases can be fully understood by a phase diagram. Since Au metal was first exploited and still is used for many applications the Au–Si (gold–silicon) phase diagram is used as an example here to explain the VLS growth mechanism. With a simple eutectic phase diagram as shown in Fig. 2.5
Fig. 2.5 Binary phase diagram of a Au–Si system and schematic illustrations of the VLS growth mechanism involved at the states of the binary system

the state of the mixture of two elements (Au and Si) can be described as a function of temperature and composition.

A eutectic phase diagram usually results when the melting point of the two components is not vastly different and the solid solubility is limited. The alloy formed by the two elements is liquid above and along the liquidus lines and solid below the solidus line [48]. For all the other regions, there is a coexistence of two phases, liquid and solid, with a certain composition and temperature. The liquidus line gives the composition of the liquid alloy as a function of temperature. The lowest melting point of the Au–Si alloy has a eutectic point \( T = 370 \, ^\circ\text{C} \) where Au is 83 at. % and Si 17 at. % [49].

Ideally, for the nanowire synthesis, we are moving along the green arrow in Fig. 2.5. Initially, the metal catalyst is at the solid phase (if the process temperature is below its melting point). (i) Precursor gas transports to the catalyst particle. (ii) The surface of the metal catalyst cracks by the precursor gas and silicon atoms are diffused into the droplet forming an alloy of Au–Si, however solid Au is also still present. (iii) While moving along the line, the Si concentration is increased up to a point where the entire droplet is a liquid Au–Si alloy and supersaturation is reached. (iv) After supersaturating, the element (Si) precipitates in a solid phase and the nanowire growth starts. The binary phase diagram of the two elements describes this process very clearly. Typically, the nanowire growth by the VLS mechanism occurs as long as the temperature is above the eutectic point. Theoretically, the growth of the nanowires will suspend if there is no source material and/or the catalyst material is not efficient.

2.5.2 Choice of Catalyst Material for Silicon Nanowires Growth

For the VLS growth method the most critical part is the choice of the metal catalyst. Since the use of a catalyst is essential for the process, the best candidate to each
particular application is extremely important and should be carefully chosen. The quality and morphology of the obtained nanostructures depend on the catalyst material properties as well as the growth conditions.

Up to now, silicon nanowires have been successfully grown by several catalyst materials including Al [50–52], Ag [46, 53, 54], Ni [46, 50, 53, 54], Cd, Bi, Fe, In [52, 55], Zn [50], Pd [46, 54], Ga [55], Au [46, 50, 53–57], Cu [46, 50, 53, 54] and Pt [46, 50, 53]. Starting from the Wagner and Ellis publication [46], the nanomaterials grown by Au metal are undoubtedly the most reviewed mainly because of the high chemical stability of Au, its non-toxicity and availability [47]. Moreover, another important parameter is the Au–Si low eutectic point (370 °C). Nevertheless, the impurity levels of Au are close to the middle of the silicon bandgap (see Fig. 2.6) causing degradation of the optical and electrical properties of the silicon nanowires and as a result malfunction of the devices [58]. Hence, a lot of effort is focused on the synthesis of Au-free catalysed silicon nanowires.

The main factors for choosing the catalyst material aiming to synthesise silicon nanowires are the eutectic point of the metal–Si alloy, the substrate temperature, the melting point of the metal as well as the impurity levels of the catalyst in the Si bandgap. In Fig. 2.7 the impurity levels of various metal catalysts as a function of the minimum temperature required for the VLS silicon wire growth are presented. The levels are shown in respect to the centre of the Si bandgap, which has the value of 1.12 eV. Above the midgap, the solid symbols mark the donor levels while the open symbols the acceptor levels. Correspondingly, below the midgap the solid symbols mark the acceptor levels and the open symbols the donor levels [47].

The metals with impurity level energies close to the conduction or valence band will cause doping. In particular, Ga, Al and In have acceptor levels close to the valence band will cause p-type doping. Similarly, Li, Sb, Te and Bi can cause n-type doping. The materials that are close to middle of the bandgap such as Au, Zn, Cu, Fe, Cr, Pd or Co should not be the first choice if deep impurity levels are not desired [59]. However, for charge storage deep level energy states can be ideal candidates for a long retention. The drawback of using Al metal is its high oxygen sensitivity. Cd has a relatively high vapour pressure, which means the metal may also evaporate during the growth process rather than having a catalytic effect on the growth. In that case, even a few atoms can compromise the performance of the device. Ga is also unfavourable because it has a very low melting and eutectic point (30 °C) [60]. It melts and tends to migrate on the sample’s surface, making the size and position control of the catalyst very difficult [61]. Ag has a high eutectic

![Fig. 2.6](image.png)
temperature (845 °C); therefore it is not compatible with temperature-sensitive substrates and low-temperature processes [62]. Hence, Sn, Pt, Ni, Pd and Tl could be possible options of catalyst material for silicon nanostructures growth via VLS mechanism if high doping are to be avoided [47, 59].

2.5.3 In and Sn Catalyst Materials for Silicon Nanowires Growth

For the purpose of this chapter, a review of the synthesis of silicon nanowires (SiNWs) using In and Sn catalyst material was conducted. In and Sn are very promising candidates as catalysts for the VLS growth of silicon nanostructures, especially at low temperatures. Moreover, the different fabrication methods in which In and Sn were used to synthesise SiNWs are presented. Although many researchers have attempted to synthesise SiNWs using In and Sn, there are only a few papers demonstrating a successful growth. These will be discussed below.

In and Sn metals share many common characteristics such as similar binary phase diagrams. They both form a single eutectic point with nearly zero concentration of silicon and without any silicide phases. In–Si eutectic temperature (156.63 °C) is almost indistinguishable from In melting point (156 °C) [63].

![Fig. 2.7](image_url) The impurity levels of various metal catalysts as a function of the minimum temperature required for VLS silicon wire growth. Silicon bandgap has the value of 1.12 eV and the levels are shown in respect to the centre of the bandgap [47]. The impurity states can also be exploited in charge storage memory
Similarly, Sn–Si eutectic point (231.9 °C) is almost the same as Sn melting point (231.93 °C) [64]. The low eutectic point is one of the main features of these materials that enable the formation of silicon nanostructures at very low temperatures. In addition, the low silicon solubility and silicon concentration at the eutectic temperature are the main parameters of the liquid alloy droplet that can determine the nucleus size for the silicon growth and crystallisation.

Although much research has been conducted for SiNWs growth using alternative metal catalysts, only a very few preliminary studies have been reported for In- and Sn-catalysed synthesis. These are presented in Tables 2.1 and 2.2 giving respectively their corresponding literature references. Different methods to fabricate SiNWs by In-seeded process have been reported such as PECVD [65–67], high boiling point organic solvent [68], hydrogen radical assisted [69], electron beam evaporation [70] and in-plane solid-liquid-solid [71]. As shown in Table 2.1, thermally evaporated In layer, indium tin oxide (ITO) substrate, indium oxide

<table>
<thead>
<tr>
<th>In catalyst layer</th>
<th>Synthesis process</th>
<th>Silicon nanowires growth details</th>
<th>Ref. year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thermally evaporated In pads</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, PECVD process for deposition of a-Si:H layer and annealing in H$_2$ ambient at 550 °C for 1 h</td>
<td>Self-aligned, in-plane SiNWs on Si (100) wafers</td>
<td>[66] 2014</td>
</tr>
<tr>
<td>Thermally evaporated In</td>
<td>High boiling point organic solvent as the growth medium</td>
<td>SiNWs 1–5 μm length and around 42 nm diameter</td>
<td>[68] 2012</td>
</tr>
<tr>
<td>ITO layer</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 500 °C</td>
<td>SiNWs 1.8 μm length and 30 nm diameter</td>
<td>[67] 2012</td>
</tr>
<tr>
<td>Thermally evaporated In</td>
<td>Electron beam evaporation at 300 °C</td>
<td>SiNWs with 600 nm length</td>
<td>[70] 2012</td>
</tr>
<tr>
<td>Electron beam deposition of In</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 500–600 °C</td>
<td>SiNWs with 150 nm length from 5 nm In layer</td>
<td>[65] 2010</td>
</tr>
<tr>
<td>ITO substrate</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, in-plane solid-liquid-solid (IPSLS) at 350–500 °C</td>
<td>Lateral growth</td>
<td>[71] 2009</td>
</tr>
<tr>
<td>In$_2$O$_3$</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by microwave plasma-assisted CVD (MPCVD) at 200–400 °C</td>
<td>SiNWs with 1 μm length</td>
<td>[72, 73] 2008</td>
</tr>
<tr>
<td>In$_2$O$_3$</td>
<td>Hydrogen radical-assisted method at 400 °C</td>
<td>SiNWs with up to 10 μm length and 10–150 nm diameter</td>
<td>[69] 2008</td>
</tr>
<tr>
<td>Electrodeposited In nanoparticles</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 500–600 °C</td>
<td>SiNWs with around 30 nm diameter</td>
<td>[52] 2007</td>
</tr>
</tbody>
</table>
(In$_2$O$_3$) layer and electron beam deposition of In are some of the routes to obtain the initial In layer. Afterwards, in most cases, H$_2$ plasma treatment is used to form the catalyst droplets that will initiate the VLS growth.

Synthesis of Sn-catalysed SiNWs has been achieved with the methods presented in Table 2.2; among them are PECVD [74–77], supercritical fluid–liquid–solid (SFLS) [78], electron beam evaporation [79], hydrogen radical assisted [80, 81] and recently magnetron sputtering [82] method. Respectively, tin dioxide (SnO$_2$)

<table>
<thead>
<tr>
<th>Sn catalyst layer</th>
<th>Synthesis process</th>
<th>Silicon nanowires growth details</th>
<th>Ref. year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Magnetron sputtering of Sn</td>
<td>Magnetron sputtering at 250–400 °C</td>
<td>SiNWs with 10–40 μm length</td>
<td>[82] 2015</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>Hot wire chemical vapour processing (HWCVP) at 300–400 °C</td>
<td>SiNWs with 3.5 μm length and 350 nm diameter</td>
<td>[85] 2013</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>Electron beam evaporation at 300–350 °C</td>
<td>SiNWs with 500 nm length</td>
<td>[79] 2013</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>High boiling point organic solvent as the growth medium</td>
<td>SiNWs with 70 nm diameter</td>
<td>[84] 2013</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>Catalytic thermal CVD at 500 °C</td>
<td>Crystalline amorphous core-shell structure</td>
<td>[86] 2012</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 600 °C</td>
<td>SiNWs with 1–3 μm length</td>
<td>[83, 87] 2012</td>
</tr>
<tr>
<td>Electron beam deposition of Sn</td>
<td>Electron cyclotron resonance chemical vapour deposition (ECRCVD) at 380 °C</td>
<td>SiNWs with up to 1 μm length and 200 nm diameter</td>
<td>[88] 2012</td>
</tr>
<tr>
<td>Sn (HMDS)$_2$</td>
<td>Supercritical fluid-liquid-solid (SFLS)</td>
<td>SiNWs with 10–100 μm length</td>
<td>[78] 2012</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 300–400 °C</td>
<td>Crystalline core of 10 nm diameter and polycrystalline shell 60 nm thick</td>
<td>[77] 2011</td>
</tr>
<tr>
<td>Thermally evaporated Sn</td>
<td>Hydrogen radical-assisted method at 400 °C</td>
<td>SiNWs with up to 2 μm length and 50–200 nm diameter</td>
<td>[80, 81] 2009</td>
</tr>
<tr>
<td>SnO$_2$ substrate</td>
<td>H$_2$ plasma treatment for catalyst droplet formation, VLS mechanism by PECVD at 300–600 °C</td>
<td>Length growth rate of 1 nm s$^{-1}$</td>
<td>[74–76] 2008, 2010–2011</td>
</tr>
</tbody>
</table>
substrates, Sn(HMDS)$_2$ and evaporated Sn layer were required to produce the Sn catalyst droplets with the assistance of H$_2$ plasma treatment.

From the literature, it is evident that In- and Sn-catalysed SiNWs can be fabricated by the same manner [61]. Moreover, the annealing with the hydrogen plasma is the key aspect of the successful synthesis [65]. H$_2$ plasma treatment has two benefits; it (i) assists in the formation of catalyst in nanodroplets sizes and (ii) can hinder the oxidation of the catalyst layer enabling their catalytic activity. One of the main differences between the In- and Sn-catalysed SiNWs is the impurity levels. Sn does not form any deep centres in the silicon bandgap while in the case of SiNWs obtained by the In catalyst the impurity levels can cause n-type doping (see Fig. 2.7). Hence, based on the aforementioned literature review, Sn could be the best candidate for synthesising undoped SiNWs at low temperature.

The alternative to Au metal-seeded synthesis of silicon nanostructures is very new and only a few applications utilising these structures have been reported. So far, Sn-catalysed SiNWs have been used for solar cell applications showing efficiency as high as 4.9% [83]. Furthermore, the electrochemical performance of In- and Sn-catalysed SiNWs has been studied for testing the suitability of the obtained nanowires as anode material for Li ion batteries and Sn-catalysed SiNWs have been demonstrated to exhibit capacities greater than 1000 mAh g$^{-1}$ after 50 charge/discharge cycles [68, 84]. To the best of the author’s knowledge, neither In- nor Sn-catalysed SiNWs have been utilised as the storage element of any electronic memory devices.

### 2.6 Flexible Electronic Flash Memory Devices

#### 2.6.1 Introduction to Flash Memory Technology

In the last two decades, electronic flash memory technology has taken on a starring role in the semiconductor industry. Electronic memory is used in the majority of everyday electronic devices such as mobile phones, tablets, personal computers, digital cameras, global positioning system and other equipment. The size of all these devices has been shrunk enormously over the years [89, 90]. This progress is partly in accordance with Moore’s law. In 1965 Gordon Moore wrote a paper entitled “Cramming more components onto integrated circuits” in which he noted that “the complexity for minimum component costs has increased at a rate of roughly a factor of two per year”. His estimation has been verified through the years and it is widely known as Moore’s law [91]. Specifically, in 1971 there were chips with 2,300 transistors while in 2006 the number increased to 300 million transistors [92]. However, a similar exponential growth has not been noted during the last
couple of years, compelling researchers to examine alternative approaches to address the limits of further scaling down.

Recently, the International Technology Roadmap for Semiconductors (ITRS) community released an article named “More-than-Moore White Paper” in which it is pointed out that for further miniaturisation a different approach might be required. In particular, the conditions that made “More Moore” possible were analysed and then the new “More-than-Moore” (MtM) trend was identified and examined as the new feasible and desirable approach [93]. Figure 2.8 shows the ideas for beyond complementary metal-oxide-semiconductor (CMOS) technology presented by the ITRS community.

Up to now, it has been accepted [94, 95] that memories are chosen based on the application they are to be utilised for. A new approach to scale the memory cell size down without compromising its performance is of great importance.

The flash memory market is currently following two paths: (i) evolutionary path, which involves a continuous improvement of the current products and technologies and (ii) disruptive path, where entire new storage mechanisms and technologies are offered. The former path is mostly followed in which silicon nanocrystals

---

**Fig. 2.8** Generic representation of trends beyond the common CMOS technology, like “More Moore” for miniaturisation and “More-than-Moore” for functional diversification approaches that were identified and presented by the ITRS community [93]
memories, high dielectric constant (high-k) materials and fin flash memories were established [96].

CMOS memories can be categorised into two major categories: the volatile memories and the non-volatile memories. In the case of the volatile memories, the information is lost when the power is turned off, and in non-volatile, the information is retained even when the power is switched off. Figure 2.9 shows many types of semiconductor memories and flash memory that fall into the class of non-volatile types. The non-volatile flash memory market contains the NAND and NOR technologies. The NAND flash memory cell is fundamentally the same as the NOR flash memory cell. The difference is in the array organisation and the operation principle. The NAND flash memory utilises the Fowler–Nordheim tunnelling (described in Sect. 3.3.1.) for programming operation while the NOR flash memory uses the channel-hot-electron injection process (described in Sect. 3.3.2.). Both NAND and NOR flash memory cells are erased by the Fowler–Nordheim tunnelling process [97]. The advantage of the NAND over the NOR flash memory cell is the higher density of cells in the former and hence a higher capacity data storage as well as faster write and erase operations. Moreover, the NOR memory cells is almost double the size of the NAND flash cell and therefore more expensive [98].

Flash memory devices are the future for any type of portable application since they are non-volatile. However, they are slow and usually a combination of more than one type of memory is required in order to create a mobile device [95, 96].

Among the non-volatile memories, flash has been proved to be the most suitable one for portable applications, the main reasons are:

- Flash memory has a very simple structure requiring only one transistor, thus making it easier to achieve the highest chip density in comparison to other non-volatile memories [98].
- Compatibility with the current CMOS technologies. The materials and the fabrication process of the flash memory are compatible with conventional CMOS processes [96].
They can be programmed and erased many times (more than 100,000 times) while at the same time they offer the smallest cell size [98].

Flash memories support the multi-bit per cell storage property. Two-bits per cell has already been established, whereas the four-bits per cell was recently commercialised by Spansion MirrorBit Quad Technology [99].

Nevertheless, electronic flash memories do have drawbacks. They require high programming and erasing voltages and their program/erase speed is slow in contrast to the rest of the non-volatile memories.

CMOS floating gate is the dominant architecture for the flash memory devices. For NOR and NAND architecture further scaling looks impossible [96, 100]. There are many challenges, such as the scaling of the tunnelling dielectric layer thickness. The tunnelling dielectric layer (usually an oxide) should be thin to allow the charges to tunnel through (for the programming and erasing processes) but, at the same time, must be thick enough to ensure data retention by preventing charge tunnelling [97]. Defects in the oxide may cause failure of the device and are among the most important factor that cause reliability concerns, which can lead to high operating voltages and high leakage currents. Another issue is the limit of the channel length. A minimum channel length has to be maintained, which impedes the further scaling of the cell. A technological challenge is additionally faced when lithography is used, as there is a minimum feature size depending on the wavelength of the light source. Finally, another difficulty in the continuous scaling down is the high process temperatures and the high thermal energy generated. Due to the reduction of the dimensions, higher operating voltages are required [96] that can cause undesirable device behaviour. The total amount of energy during a process termed ‘thermal budget’ should be as low as possible. A short time process (a few seconds) can fulfil this requirement since the total energy is temperature-time dependent.

In order to overcome these difficulties, new materials and device architectures are being explored. For example, charge-trapping devices, such as silicon-oxide-nitride-oxide-silicon (SONOS), were presented as a potential solution. SONOS devices have a simpler structure than the flash memory devices due to exploiting the intrinsic defects of the silicon nitride that can trap charges [97]. Another suggestion is the use of metal or silicon nanocrystals for the floating gate that was first introduced in 1995 by Tiwari [101]. High-k materials are mainly considered for the dielectric layer because a material with higher dielectric constant (k) can be physically thicker (reduces the leakage current) while having the potential to increase the gate capacitance and therefore improve the device performance [102]. Three-dimensional (3D) approach hopes to solve the obstacle of the minimum value of the channel length. It offers the prospect of a multilevel cell where the memory arrays are vertically stacked [96]. On the other hand, totally new storage mechanisms are studied, such as the ferroelectric and phase change memories [103, 104]. However, the cost of these is higher than of the flash memory and there are still several doubts about their reliability.
2.7 Flash Memory Structure

The basic flash memory cell consists of one transistor combined with a floating gate. In particular, it is a conventional metal-oxide-semiconductor field-effect-transistor (MOSFET), as can be seen in Fig. 2.10a, with an extra gate between the tunnel and the control (or gate) oxide, named the floating gate. The floating gate is the memory storage component and it is isolated between the insulator layers in order to trap charges. By adding the floating gate, the device properties change and result in a flash memory cell. In the literature, sometimes referred to as a floating gate based memory or three terminal memory cell, because the memory cell has three terminals: a source, a drain and a gate [98, 100, 105].

As part of this work, the structure shown in Fig. 2.10b was revised to meet the requirements of a flexible flash memory cell. It is however very similar to the basic flash memory cell structure and holds the same operating principle. Yet, for enabling the utilisation of the flexible substrate, the rigid silicon substrate was replaced. A suitable architecture is shown in Fig. 2.11. The structure is an inverted conventional flash cell and a high-quality semiconductor layer is used for the transistor area.

![Fig. 2.10 Schematic illustration of a the conventional MOSFET structure and b the floating gate transistor or flash memory cell](image1)

![Fig. 2.11 Schematic illustration of a flexible flash memory cell structure](image2)
2.8 Working Mechanisms of the Flash Memory Cell

As previously mentioned, the two main mechanisms for programing and erasing a flash memory cell are the channel-hot-electron injection and the Fowler–Nordheim tunnelling. Both mechanisms trap electrons inside the floating gate, whereas the latter removes the electrons from the gate by tunnelling processes. Trapped electrons shift the threshold voltage of the cell. This shift can be observed during the read process between the source and drain. By defining two different states (for shifted and non-shifted threshold voltage) we are able to store one bit of information (1 or 0) inside the cell.

In order to explain the working mechanism of the flash memory, the quantum tunnelling phenomena and the probability of an electron to tunnel through a barrier is discussed. Quantum tunnelling is a phenomenon that can happen through a thin potential barrier and cannot be explained by classical physics. For a finite potential barrier of height $V_0$, width $L$ and for particles with kinetic energy smaller than the barrier height ($E < V_0$), there is a small non-zero probability of particles tunnelling through the barrier. In that case, the time-independent Schrödinger equation for a particle moving in one dimension is the following:

$$\frac{-\hbar^2}{2m} \frac{\partial^2 \Psi(x)}{\partial x^2} = (E - V_0) \Psi(x)$$

where $\hbar$ is Planck constant divided by $2\pi$, $m$ is the mass of the particle and $\Psi$ is the wave function of the system.

Moreover, the transmission probability or the transmission coefficient for the potential barrier is

$$T_{E < V_0}(E) = \frac{4k^2 \gamma^2}{V_0^2 \sin \hbar^2 (\gamma L) + 4k^2 \gamma^2}$$

where $k = \sqrt{2m(E - V_0)/\hbar^2}$ and $\gamma = \sqrt{2m(V_0 - E)/\hbar^2}$.

For a barrier with a very large width ($L \gg$) and a barrier height greater than the energy of the particle ($V_0 \gg E$), the hyperbolic $\sin (\gamma L)$ term can be simplified to $\approx \frac{1}{2} \exp (\gamma L)$. In this case, (2.7) gives the following approximation:

$$T(E) \approx \frac{16E}{V_0} \left(1 - \frac{E}{V_0}\right) \exp(-2\gamma L) \approx \exp(-2\gamma L)$$

Therefore, the probability of a particle tunnelling through the barrier depends exponentially on the thickness of the barrier $L$, the energy difference ($V_0 - E$) and the mass of the particle (2.8) [106, 107].
2.8.1 Fowler–Nordheim Tunnelling

It has been established that the charging and discharging of the floating gate of the flash memory devices happens through the tunnelling phenomena. The potential barrier of each device segment is shown in Fig. 2.12 to explain the working mechanism. Initially, there is no applied voltage and all the energy levels are flat (Fig. 2.12a). Once voltage is applied, the energy levels shift. In the case of a high-applied electric field, the conduction band of the tunnel oxide is triangular in nature (Fig. 2.12b, c) and the Fowler–Nordheim tunnelling occurs [108]. The magnitude of the applied voltage is very critical since the higher the gate voltage the greater the probability of tunnelling. At higher electric fields, the potential becomes steeper and the energy barrier width reduces.

Wentzel–Kramers–Brillouin (WKB) [28], the approximation is the most adopted approach for calculating the transmission coefficient of variable potential barriers.

![Figure 2.12](image.png)

**Fig. 2.12** Energy band diagrams showing the potential barrier of **a** each section of the flash memory, **b** the Fowler–Nordheim tunnelling/programming process and **c** the Fowler–Nordheim for the erase process.
for the time-independent one-dimensional Schrödinger equation. For $E < V_0$, the
tunnelling region, the triangular shaped barrier has a potential of
$V = E_F + \Phi - eEx$, $E_F$ is the Fermi energy and the tunnel probability is given by
(2.9).

$$T = \exp \left[ -\frac{4\sqrt{2m^*} \Phi^{3/2}}{3h} \frac{eE}{3q} \right]$$

where $E$ is the applied electric field, $m^*$ is the effective mass of the electron and $\Phi$ is
the injection barrier height at the interface. The electron barrier at the Si/SiO$_2$
interface is about 3.15 eV and is higher than that of the silicon nitride (2.1 eV) [89]
which means that a higher applied voltage is required in order to program the
memory cell, when using the silicon oxide as a tunnelling layer. Moreover, the
Fowler-Nordheim tunnel current density is given by (2.10) [109, 110].

$$J_{FN} = C_{FN}E^2 \exp \left[ -\frac{\beta}{E} \right]$$

where $C_{FN} = \frac{q^2m^*}{8\pi\hbar\phi m}$, $\beta = \frac{4\sqrt{2m^*} \Phi^{3/2}}{3hq}$, $h$ is Planck constant, $m$ is the mass of a free
electron, $q$ is the charge of single electron, $m^*$ is the effective mass of the electron
and $\Phi$ is the injection barrier height of the interface. Hence the current density $J_{FN}$
is exponential dependent on the applied field. Usually a plot of $\ln(J/E^2)$ versus $1/E$
is plotted which should yield a straight line [111, 112].

### 2.8.2 Channel-Hot-Electron Injection Process

During programming a NOR flash memory cell, channel-hot-electron injection
process occurs. Energy higher than the energy barrier (for example 3.15 eV for
silicon oxide tunnel layer) is required for an electron to surmount the tunnel oxide
barrier and be injected into the floating gate. Under positive drain bias, the lateral
channel electric field given by the source-drain potential $V_{DS}$, accelerates the
electrons from the source to the drain side and are said to become ‘hot’ (have high
kinetic energy). Those electrons can ‘jump’ from the channel region into the gate.
However, only a few hot electrons can successfully gain sufficient energy to
overcome the energy barrier and be injected into the floating gate [98, 100].

In general, the channel-hot-electron injection process takes place when the
electrons have energy higher than the potential barrier in order to be injected into
the floating gate. A schematic illustration is shown in Fig. 2.13. This process is not
preferable because the ‘hot’ electrons have a high impact on the reliability per-
formance of the device and can cause degradation of the cell [113].
2.8.3 Reading Operation

The presence of charges in the floating gate give rise to a threshold voltage shift which is monitored as shown in Fig. 2.14. For charging the floating gate, it is necessary to apply an outer gate voltage. For discharging the floating gate, an opposite in polarity but equal in magnitude to the charging voltage should be applied. The stored information in the floating gate is recalled from the cell during the read process.

![Fig. 2.13](image)

Fig. 2.13 A schematic illustration using the energy band diagrams of a flash memory device to describe the channel-hot-electron process

![Fig. 2.14](image)

Fig. 2.14 The behaviour of a flash memory device during the reading process. The threshold voltage shift and the formation of the two states (charged or programmed state with bit value = 0 and uncharged or erased state with bit value = 1) are demonstrated based on the I–V characteristics
When charges are stored, a conductive channel is formed, the channel current is ON and a shift in I–V characteristics of the device is observed, known as memory window. The two states are illustrated in Fig. 2.14; the programmed state with a bit value of zero and the erased state with a bit value of one. During the reading process, the states can be identified using the I–V characteristics of the device. For example, for reading a bit, a read voltage ($V_R$) on the control gate is applied and the drain current ($I_D$) of the floating gate transistor is measured. In the case shown in Fig. 2.14, at the voltage ($V_R$), the read current shows a positive value, indicating state ‘1’. When state ‘0’ is formed, the drain current value is zero for the same voltage $V_R$ [98].

2.9 Choice of the Dielectric Material: Silicon Nitride

One of the main limitations for the continuous reduction of the flash memory cell dimensions is the thickness of the tunnel oxide layer. The role of the tunnel oxide is to allow the charge to transfer to and from the floating gate quickly under the influence of an electric field (see section Fowler–Nordheim tunnelling). A low electric field is preferred in order to have a low power operation and minimum undesirable effects for the rest of the memory cell. At the same time, the tunnel oxide needs to play its role as an insulator. During the read operation, the charges should not leak out fulfilling the requirement for at least 10 years retention time [98]. Moreover, the stressing effect on the layer that occurs due to the repeatedly write/erase operations will introduce defects—oxide traps in the layer.

Silicon oxide is widely used because it is easy to fabricate and forms a good interface with silicon. The silicon oxide thickness currently used, satisfactory for the tunnelling process is 7–9 nm, and it has not been possible to reduce it further. A thicker dielectric layer with higher dielectric constant is possible to replace the thin silicon oxide. Hence, alternative materials are considered.

Silicon nitride is used in a wide range of applications such as thin film transistors, oxidation masks and solar cells. Its most common utilisation is in the CMOS industry as a dielectric layer for electrical isolation, mainly due to its mechanical properties and chemical inertness [114]. Silicon nitride films can be produced by different deposition methods such as CVD, PECVD and reactive sputtering. However, they are generally deposited at high temperatures (700–800 °C) by both high-temperature chemical vapour deposition (HT-CVD) techniques and low-pressure chemical vapour deposition (LPCVD), which provides highly stoichiometric films with very low hydrogen content [115, 116].

The main benefits for using silicon nitride films for flexible memory are:

- High-quality silicon nitride layers can be fabricated at low temperatures (200–400 °C) by PECVD. In general, for plastic and flexible substrates, the fabrication temperature must not exceed 350 °C. For such applications, the PECVD is highly suitable, although films tend to be non-stoichiometric and contain appreciable amounts of hydrogen [116].
• Silicon nitride films can be fabricated with commercialised processes. Apart from compatibility issues, their chemical stability at high temperature processes is also very advantageous [117].

• Due to its high dielectric constant, a thicker layer is possible (fulfilling the requirement for 10 years retention time). The (electrical) equivalent oxide thickness (EOT) is an important parameter that indicates the essential thickness of the silicon oxide in order to have the same effect as with the high-k material being used and is calculated by (2.11) [118].

\[
EOT = \left( \frac{k_{\text{SiO}_2}}{k_{\text{high}-k}} \right) t_{\text{high}-k}
\]  

(2.11)

where, \(k_{\text{SiO}_2}\) is the dielectric permittivity or dielectric constant of the silicon oxide (3.9), \(k_{\text{high}-k}\) is the dielectric permittivity of the high-k material and \(t_{\text{high}-k}\) is the physical thickness of the high-k material. Hence, approximately 15.4 nm of silicon nitride layer (\(k_{\text{silicon nitride}} = 7.5\)) would produce the same effect as 8 nm of silicon oxide layer.

• Silicon nitride is also preferred due to its lower injection barrier (2.1 eV) in comparison to silicon oxide (3.15 eV) for the tunnelling process [89]. This will enhance the programming speed (in the case of the thin layer) as well as the retention time (in the case of the thick layer).

The only major drawback of using silicon nitride is that it can trap charges. However, the controlled deposition process and optimisation of the deposition conditions can offer a very low charge trapped density. Another challenge is the fabrication at low temperature. Deposition of silicon nitride dielectric films at low temperatures is more likely to produce high concentrations of hydrogen in the films, which lead to poor electrical properties and an increase in porosity of the material that results in oxidation over time [102].

2.10 A Brief History of Flexible Flash Memory Devices

It has been almost 50 years since the first floating gate memory device was presented using an insulated gate field effect transistor in 1967 by D. Kahng and Sze [119]. A simple sandwich structure of metal (1)/insulator (1)/metal (2)/insulator (2)/metal (3) was demonstrated effectively as a memory device with the metal (2) being the isolated floating gate component. More than a decade later, in the 1980s, Masuoka working for Toshiba made a breakthrough by inventing the flash memory [120]. Following that, it took only a couple of years for the first flash chip memory to become commercially available from Intel. Since then, fast evolution has taken place, with flash memory being the dominant memory for portable applications and with memory devices on flexible substrate paving a path for new futuristic products.
Until now, the demand for flash memory devices has been growing rapidly. Since the flash memory device is effectively a transistor with an additional isolated floating gate, these two parts have been evolving aiming to accomplish a fully flexible memory cell. Many types of transistors were fabricated before the memory idea came into function. The dominant materials for thin film transistors (TFTs) are the hydrogenated amorphous silicon (a-Si:H), the polycrystalline silicon (poly-Si) and the organic semiconductors [121]. All these materials have been used in flash memory devices and up to now only organic semiconductors have shown a great potential for the application with flexible substrates. It is well known that the high-performance transistors (high mobility and chemical stability) are the poly-Si transistors. However, the methods to obtain flash memories with poly-Si transistors are utilising high temperatures such as solid-phase crystallisation (SPC), excimer laser annealing (ELA) and sequential lateral solidification (SLS) [122–124]. In order to overcome this difficulty, low-temperature poly-Si thin film transistors (LTPS TFTs) [125] have gained great interest and they are excellent candidates for low-temperature non-volatile memories (NVM) with poly-Si TFTs. Many research groups have reported low-temperature NVM poly-Si TFTs with (i) trapping silicon nitride layer [126], (ii) embedding nickel–nanocrystals (Ni–NCs) into the silicon nitride layer for the trapping layer [127, 128], (iii) In$_2$O$_3$ nanodots embedded in polyimide insulating layer [129] or (iv) using biotechnology [130]. However, in all these cases, the maximum deposition temperature of the whole process exceeded the plastic temperature limit (350 °C), making incompatible with plastic substrates.

Since process temperature is a vital concern, organic materials were thoroughly studied to replace the inorganic semiconductors. The polymers can be deposited at low temperature and cost using spin coating, roll-to-roll and inkjet printing [131]. Research in organic TFTs (OTFTs) has had a rapid increase through the years and a carrier mobility as high as 3 cm$^2$/Vs has been exhibited which is similar or even better than that reported for a-Si:H TFTs [132]. Hence, OTFTs were actively investigated for the transistor part of the non-volatile memories (ONVM-TFTs). Pentacene organic semiconductor thin film is one of the most dominant materials [133–135]. Another promising approach is the use of transparent oxide semiconductors for the channel of thin film transistors such as indium gallium zinc oxide (IGZO) [136]. The use of these in a ferroelectric transistor memory on a flexible foil has also been demonstrated recently [137]. Nevertheless, organic materials have some vital drawbacks, such as poor electrical performance, light sensitivity and material degradation caused by the environment.

One of the earliest reports on the storage function of TFTs using the floating gate concept was reported as early as 1975 in which cadmium selenide (CdSe) thin film transistor and aluminium thin layer (or clusters) for the floating gate were used [138]. However, due to the rapid evolution of silicon technology, semiconductor chalcogenide compounds did not progress as quickly. The reassessment came years later in order to challenge the established silicon industry and propose alternative materials such as nanoparticles or nanowires of chalcogenide compounds [139].

Hence, for further scaling down nanostructures were considered for their potential integration into electronic devices, either for the semiconductor channel
area of the transistor or for the floating gate component of the memory cell. The first breakthrough took place in 1995 when Tiwari first proposed the silicon nanocrystals non-volatile memory device, in which the nanocrystals served as discrete nodes for the storage of charges [101].

2.11 Review of the Current Status of Flexible Flash Memories

As previously stated, after the establishment of flexible thin film transistors, some of them were studied in order to develop into flexible flash memory devices. An interesting study by Dr. Young [140] on the fabrication of poly-Si TFTs and memory devices on glass using a silicon nitride layer with a high trapping state density, reported a high field-effect mobility (170 cm²/Vs) as well as a threshold voltage of 4.2 V. The process temperature did not exceed 300 °C, but the poly-Si was produced by the crystallisation of amorphous silicon films using a scanned excimer laser annealing process.

To date, the most capable candidates for the flexible flash have been demonstrated to be the organic semiconductor materials that take advantage of the OTFTs for the transistor based memory. Flexible organic pentacene transistor non-volatile memories based on high-dielectric layers [141] or embedding self-assembled gold nanoparticles by solution process [135] are already described in literature. However, the memory window and the retention time are not high enough to fulfill the great demand of high performance flexible flash memory devices. For flexible NVM-OTFT only Ha et al. reported a carrier mobility (4.25 m²/Vs) using Al₂O₃ dielectric layers by solution-based combustion process and gold nanoparticles by contact printing process [142].

Flexible ferroelectric memory seems to be another popular approach in which the gate dielectric is replaced by a ferroelectric insulator [137, 143–146]. Among these, Breemen et al. reported a retention time of more than 12 days using indium-gallium-zinc-oxide (InGaZnO) as the semiconductor layer [137] while with amorphous InGaZnO a TFT mobility of 8 cm²/Vs was measured [143]. Moreover, flexible ferroelectric memory has been demonstrated on a polyethylene naphthalate (PEN) substrate in a NAND-like architecture [146].

Recently, all-solution processed organic transistor flexible memory devices were described [147–149]. The flexible organic transistor showed a very low mobility (0.08 cm²/Vs) and the organic flexible memory had a lower data retention time (10⁵ s). Yet, good endurance and bending stability were reported [149]. Higher mobility (0.8 cm²/Vs) and on/off ratio of 10⁴ were achieved for diode-load inverters [147]. It is worth mentioning that in the above cases the entire fabrication process was performed by using solution-based techniques.

A new idea is the use of graphene materials, in particular reduced graphene oxide as the charge-trapping layer, on flexible memory devices based on organic
transistor [150]. In addition, C_{60} molecular floating gate layers were successfully demonstrated as the charge-trapping layer with OTFTs for flexible flash memories [151]. All of these approaches have not been fully explored yet and it is expected that researchers will push these forward.

One step further is the use of paper substrates. Martins et al. reported a write-erase and read paper memory transistor with carrier mobility as high as 40 cm^2/Vs using an n-type memory field effect transistor [152]. The most recent article uses Fuji Xerox printer paper for the substrate without any treatment. An average carrier mobility value of 0.297 cm^2/Vs, on/off ratio >10^4 and a retention time of 10^4 s for 25 organic transistors was reported [153].

As a final point, for achieving a fully flexible flash memory some issues should be addressed. First of all, to make the structure flexible, all components must be able to bend to some degree without losing their functionality. In addition, materials with characteristics such as high carrier mobility, low cost, chemical stability and low-temperature fabrication processes should be considered. The reliability of the memory device is also a vital factor. It is possible to overcome these challenges if the device is prefabricated and then transferred onto the plastic substrates. However, that adds one more step to the fabrication process and may significantly increase the cost. The polymer substrates can also be high-temperature substrates like polyimide or polyethersulphone. Still, the overall temperature limit of the full preparation of the flexible flash memory device should be considered.

2.12 Summary

To sum up, nanotechnology and nanomaterials constitute a recent rapid developing area. Nanomaterials are used either to study fundamental quantum phenomena or to fabricate novel and improved performing devices. The state-of-the-art electronic memory devices utilising nanostructures were briefly introduced in this chapter. Side by side, the critical function of the metal catalyst in the growth dynamics of the silicon nanostructures via VLS mechanism was discussed. The energy states of these nanostructures (including the energy traps formed by catalyst materials) can be used for trapping the charge for memory devices.

Flash memory cells have been established as the dominant types of the non-volatile semiconductor memories mainly for portable applications. Here, an overview of flash memory and specifically flexible flash memory devices has been presented. The structure, the basic principles behind their working mechanism, the challenges being faced as well as the previous work done in the field of the flexible flash memory devices have been discussed in this chapter. For the fabrication of a fully flexible flash memory, a combination of the required characteristics of the materials as well as a low thermal budget process for the main steps such as the fabrication of the memory element (floating gate) should be taken into consideration.
References

48. V. Raghavan, Materials science and engineering: a first course, 5th edn. (PHI Learning Pvt Ltd, Delhi, 2013)
59. S.M. Sze, K.K. Ng, Physics of Semiconductor Devices (Wiley, USA 2007)
66. L. Yu, M. Xu, J. Xu et al., In-Plane epitaxial growth of silicon nanowires and junction formation on Si(100) substrates. Nano Lett. 14(11), 6469–6474 (2014)
89. J. Brewer, M. Gill, Nonvolatile memory technologies with emphasis on flash: A comprehensive guide to understanding and using NVM devices. (Hoboken and Wiley, 2008)
123. Y. Kuo, Thin Film Transistors: Materials and Processes (Kluwer Academic, 2004)


149. C. Kim, J.M. Song, J.S. Lee et al., All-solution-processed nonvolatile flexible nano-floating gate memory devices. Nanotechnology 25(1) (2014)


Charge-Trapping Non-Volatile Memories
Volume 2--Emerging Materials and Structures
Dimitrakis, P. (Ed.)
2017, V, 211 p. 170 illus., 117 illus. in color., Hardcover
ISBN: 978-3-319-48703-8