Chapter 2
Reliability of 3D NAND Flash Memories

A. Grossi, C. Zambelli and P. Olivo

2.1 Introduction

Reliability represents one of the major antagonist towards the unstoppable technological evolution of hyperscaled NAND memories, since the correct operations must be assured throughout the entire lifetime. In particular, the ability of keeping unaltered the stored information even after a consistent number of write operations and for long times must be guaranteed.

A growth of the memory devices storage capacity without increasing the area occupation is constantly requested by the market: in order to satisfy such requirements, an increase of the memory density and of cell shrinking is mandatory. Nowadays, the transition from planar to three-dimensional architectures appears as the most viable solution for the integration of non-volatile memory cells in Tera-bit arrays. *Charge Trap* (CT) NAND memory cells are considered as one of the most promising technology for 3D integration because of a better scalability than *Floating Gate* (FG) NAND. Despite the high theoretical potentialities demonstrated by CT memories, several reliability issues affect such technology. Moreover, the transition from 2D to 3D changed the impact of the previously known reliability issues and generated new problems. Recently, in order to overcome such problems, new 3D vertical FG type NAND cell arrays have been proposed with promising performances.

A. Grossi (✉) · C. Zambelli · P. Olivo
Dipartimento di Ingegneria, Università degli Studi di Ferrara, Ferrara, Italy
e-mail: alessandro.grossi@unife.it

C. Zambelli
e-mail: cristian.zambelli@unife.it

P. Olivo
e-mail: piero.olivo@unife.it

R. Micheloni (ed.), *3D Flash Memories*, DOI 10.1007/978-94-017-7512-0_2

In this chapter the main reliability mechanisms affecting 3D NAND memories will be addressed, providing a comparison between 3D FG and 3D CT devices in terms of reliability and expected performances. Starting from an analysis of basic reliability issues related to both physical and architectural aspects affecting NAND memories, the specific physical mechanisms impacting the reliability of 2D CT NAND will be addressed. Then, a review of the main problems experimentally observed in different 3D CT cell concepts is reported. Finally, 3D FG memory concept is briefly introduced in order to understand the related reliability implications, and a comparison between 3D CT and 3D FG arrays is provided in terms of reliability and expected performances.

2.2 NAND Flash Reliability

During its lifetime a NAND Flash module undergoes a large number of Program/Erase (P/E) cycles. Every cycle involves very high electric fields applied to the tunnel oxide. The reliability of the entire memory requires that the tunnel oxide is able to correctly operate under stress conditions. It is obvious that huge efforts are to be spent to determine the right process for the tunnel oxide creation (in terms of thickness, material, growth, defectivity, interface, …) and the most effective algorithms in order to achieve a successful and reliable NAND technology.

In this section we will analyze the basic physical mechanisms related to the tunnel oxide, which affect both memory endurance and data retention. "Endurance" of a memory module is defined as the minimum number of P/E cycles that the module can withstand before leading to a failure. "Retention" is the ability of storing the information over time even when the external power supply is not applied. The tunnel oxide, which is a thin oxide, may be also responsible for other effects, such as erratic bits and over-programming, which might induce read errors.

2.2.1 Endurance

In NAND flash cells, program and erase operations rely on charge transport through thin oxides; this is accomplished via Fowler-Nordheim (FN) tunneling into/from a storage layer, which can be either a polysilicon FG [1] or an interfacial trapping layer in CT technology [2, 3]. Electron tunneling is responsible for a slow, but continuous, oxide wear out because of traps creation and interfacial damages; as a result, there might be charge trapping/detrapping into the tunneling oxide or undesired charge flowing into/from the storage layer.

As the number of P/E cycles increases, the above mentioned effects strongly impact writing operations. For instance, electron trapping reduces the tunneling efficiency so that, under constant voltage and time conditions, the charge injected into/from the storage layer decreases cycle after cycle.

To counteract "*endurance*" effects, all writing algorithms are based on a sequence of program/erase pulses, each one followed by a verify operation. This sequence proceeds until the expected amount of charge is correctly transferred into/from the storage layer. As the number of P/E cycles increases, the programming time is expected to reduce, whereas the erase time is expected to grow.

Without these write and verify algorithms (Chap. 3) it would be impossible to control the actual amount of charge transferred into/from the storage layer and Multi Level Cell (MLC) architectures would not exist [4].

Even if endurance is controlled by sophisticated (but slow and power consuming) algorithms, traps creation, charge trapping/detrapping, and interface damages still degrade the tunnel oxide. As a result, it gets really problematic to retain the stored information for extremely long times, which, at the end of the day, is a basic requirement of the non-volatile paradigm.

2.2.2 Data Retention

As mentioned in the previous section, the ability of keeping the stored information unaltered for a long time, i.e. the charge trapped into the storage layer, is mandatory for non-volatile memories. However, even with no bias applied, electron after electron, charge loss can lead to a read failure: a programmed cell can be read as erased if its threshold voltage (V_T) shifts below 0 V in case of Single-Level-Cell (SLC), or towards a lower threshold level with respect to the initial threshold voltage in case of MLC programming [5].

The higher the number of P/E cycles the worse the retention is, as it can be appreciated in Fig. 2.1, which shows how the cumulative V_T distributions of MLC programmed cells changes over time. Charge loss from the storage layer moves the V_T distributions towards lower values: the rigid shift of the cumulative V_T distributions is related to the oxide degradation and traps generation at the interface between storage layer and tunnel layer. These traps may be responsible for charge loss from the storage layer towards the silicon substrate. In fact, an empty trap, suitably positioned within the oxide, can activate *Trap Assisted Tunneling* (TAT) mechanisms characterized by a significantly higher tunnel probability with

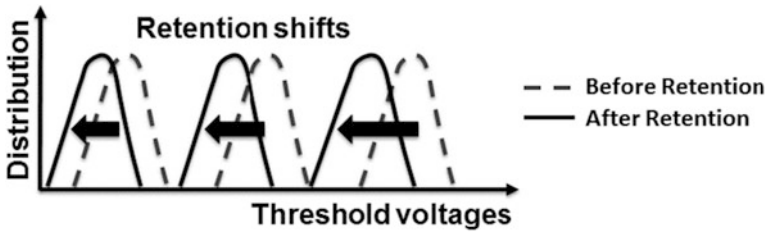

Fig. 2.1 Threshold voltage shifts induced by retention

respect to a triangular barrier unmodified by the trap presence. Moreover, an electron trapped inside the oxide during writing operations may be detrapped later on, when the cell is read or even when the cell is not addressed. As a result, the empty trap may enhance the TAT phenomenon (assuming a positive charged trap) and, in addition, it can increase the electron field at the storage layer-tunnel oxide interface, thus raising the probability of electron tunneling. It is clear that these mechanisms are strongly related to the oxide degradation and, therefore, data retention gets shorter with the number of applied writing pulses. In the MLC case, the cells programmed at higher V_T are more prone to data retention issues.

2.2.3 *Erratic Bits and Over-Programming*

The *Fowler-Nordheim* (FN) tunneling mechanism for writing and erasing data in NAND Flash has been used for several decades, demonstrating a sufficient level of reliability.

Nevertheless, it has been found that anomalous FN tunneling currents can occur in random periods of time, thus leading to significant variations of the threshold voltage after the writing operation [6] (see Fig. 2.2). This phenomenon is known as *erratic bits*.

In a NAND array, the presence of this phenomenon is detrimental for the performances of the memory as the unpredictable increase of the cell's threshold voltage may eventually induce the *over-programming* issue. As shown in Fig. 2.3, conductive cells featuring relatively large threshold voltage are erroneously read as

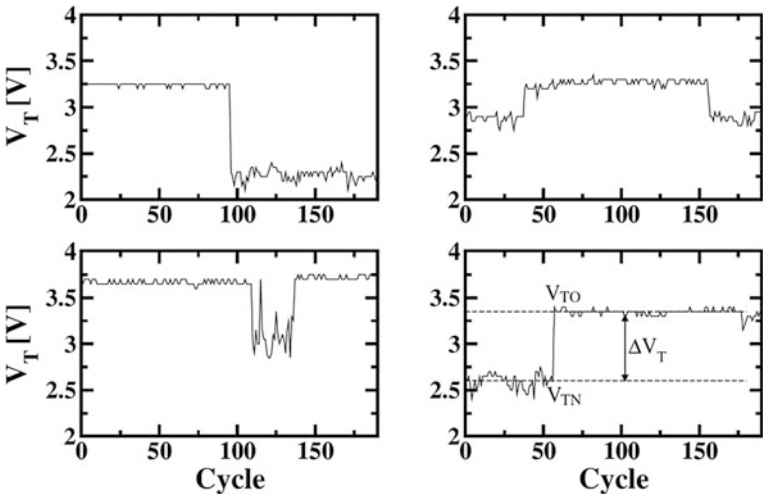

Fig. 2.2 Example of erratic behaviors in four flash cells. Cells threshold voltage V_T plotted versus the number of cycles exhibits RTN [6]

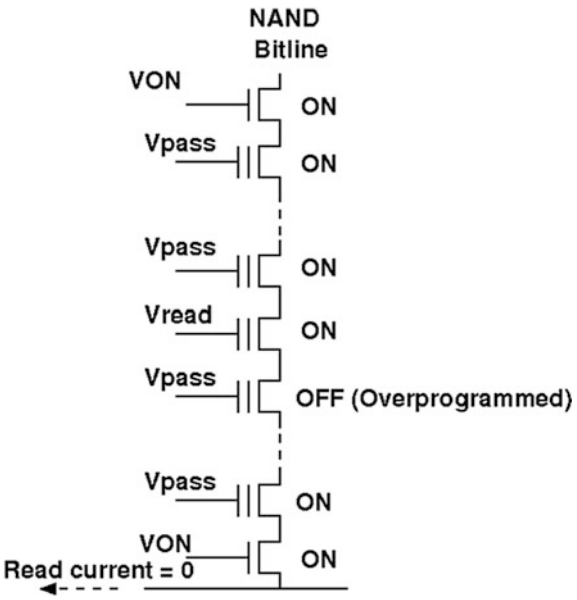

Fig. 2.3 Effect of an over-programmed cell in a NAND flash string. In normal conditions the status of the cell to be read (supposed to be ON) is correctly detected, since all other cells are driven by a V_{PASS} so that they behave as ON pass transistors. In the presence of an over-programmed cell ($V_T > V_{PASS}$), the current flow through the string is inhibited and the absence of current is attributed to a programmed status of the cell to be read, thus producing a read error [6]

OFF if over-programmed, and they can electrically isolate the NAND string. Such behavior generates read errors and consequent read throughput loss due to the additional work done by the *Error Correcting Codes* (ECC) trying to repair the failed bits.

Since erratic behaviors are intimately related to the electron tunneling mechanism, they can potentially affect all the cells of an array [6].

Anomalous tunneling has been related to the presence/absence of a cluster of positive charges in the tunnel oxide that strongly affects the FN tunneling operation. As a first approximation, erratic behaviors can, therefore, be described in terms of a two level *Random Telegraph Noise* (RTN) affecting the threshold voltage during cycling, in which the normal and the anomalous threshold voltage levels are the result of the presence of a cluster of more than 2, or less than 3, positive charges in the tunnel oxide, respectively [7, 8].

2.3 Architecture Dependent Reliability Issues

Architectural solutions for memory operations may also affect the overall reliability, by inducing errors and even cell failures [1]. The most common effects are the so called “disturbs”, that can be interpreted as the influence of an operation performed on a cell (Read or Write) on the charge content of a different cell.

Read disturbs are the most frequent source of disturbs in NAND architectures. This kind of disturb may occur when reading many times the same cell without any

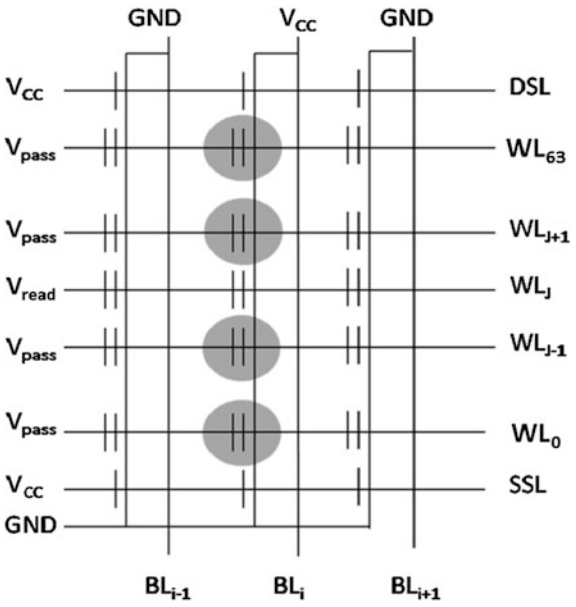

Fig. 2.4 Representation of read disturb in a NAND flash array. The cells potentially affected by read disturb are marked in *gray*

erase operation. All the cells belonging to the same string of the cell to be read must be ON, independently of their stored charge. The relatively high V_{PASS} applied to the control gate may induce a charge increase, especially if the read operation is repeated many times. These cells suffer a positive shift of their threshold voltage, which may lead to read errors. Figure 2.4 shows the typical read disturb configuration. As a matter of example, a 64 cell string is considered in the following, however the reported considerations can be extended to longer strings as well.

The probability of suffering from read disturb increases with the number of P/E cycles (i.e. towards the end of the memory useful lifetime) and it is higher in damaged cells. Read disturb does not cause permanent oxide damages: it can be reset by a simple erase operation.

In case of MLC programming, cells with lower V_T are slightly more vulnerable to shift than cells with higher V_T (see Fig. 2.5). In fact, the lower the threshold voltage the higher the voltage difference ($V_{PASS} - V_T$) across the tunnel oxide, which translates in a higher tunneling current. For cells in the erased state (ER) we observe a systematic shift of the cell's V_{TH} to the right (i.e. to higher values). The shift for P1 and P2 is much lower, since the read disturb effect becomes less prominent as V_T increases. For cells in P3, on the contrary, the average V_T shifts to the left. This is mainly due to charge loss (retention), which outweighs read disturb.

Two other important types of disturbs arise during the write operation: *Pass disturb* and *Program disturb*, which are shown in Fig. 2.6 (left and right), respectively. The former is similar to the read disturb and affects cells belonging to the same string of a cell to be programmed. With respect to the read disturb, the Pass disturb is characterized by a higher V_{PASS} voltage applied to cells that are not

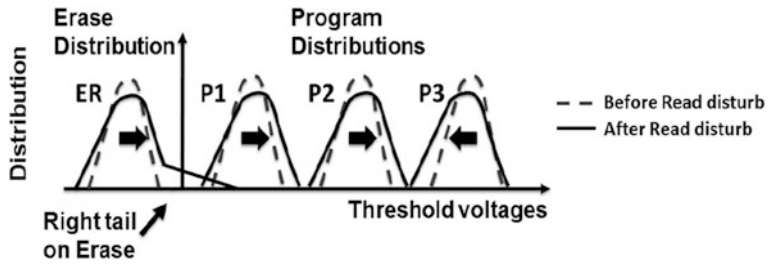

Fig. 2.5 Threshold voltage shift induced by read disturb

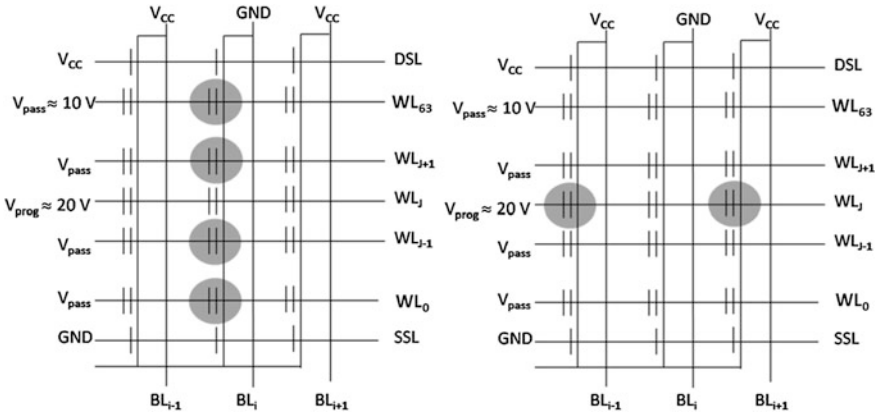

Fig. 2.6 Representation of pass disturb (*left*) and program disturb (*right*) in a NAND flash array. The cells potentially affected by disturbs are marked in *gray*

to be programmed (thus enhancing the electric field applied to the tunnel oxides and the probability of undesired charge transfer). On the other hand, the pass disturb can be repeated for a limited number of times (i.e. the number of cells in the string minus 1). In fact, when a string (block) has been fully programmed, an erase operation must be necessarily performed before any other reprogram.

The Program disturb, on the contrary, affects cells that are not to be programmed and belong to the same wordline of those that are to be programmed. In this case, again, no cumulative effects are present.

Edge Wordline disturbs affect the cells belonging to the first and last wordline, which connect the cell strings to the string selectors [9]. This disturb is due to a difference between the V_T of the cells belonging to WL_0 and WL_{63} with respect to the average V_T of all other cells. Such difference can be ascribed to three effects:

- different potentials at their terminals with respect to the other cells depending on the specific WL selected for programming;
- a different cell geometry due to the fact that these cells are located between a cell and a transistor (differently from cells belonging to WL_1–WL_{62}), therefore with a different field underneath their channels and a modified programming dynamics;

- a different cell lithography, especially when extremely scaled technology nodes are considered;
- the presence of a large *Gate Induced Drain Leakage* (GIDL) [10] current generated at the drain edge of DSL/SSL transistors due to their drain potential raised by channel boosting: such a field can efficiently trigger electron-hole pair generations followed by an acceleration of the electrons toward the channel of WL_0 and WL_{63} cells. These electrons can be injected into the floating gate of these cells, thus provoking an undesired increase of their threshold voltages.

In order to overcome such problem, the most common solution is to introduce two or more dummy WL before WL_0 and after WL_{63}, shielding the Edge Wordline Disturb: in such a way, the difference between the edge cells and all other cells in terms of potential at their terminals and cell geometry is minimized.

2.4 2D Charge Trap: Basics

The basic concept of a CT NAND memory cell consists of a metal oxide semiconductor device where the FG is replaced by an insulating charge trapping layer. Such storage layer, typically made of silicon nitride, is isolated by means of a tunnel oxide and a blocking oxide as sketched in Fig. 2.7 where the FG cell structure is reported for comparison. The tunnel oxide plays a basic role for the control of the device threshold voltage, whose value represents, from a physical point of view, the stored information. The blocking oxide prevents electrons from passing to/from the control gate. Electrons transferred into the storage layer give a threshold voltage variation. In quiescent conditions, thanks to the two oxides, the stored charge is supposed not to leak away, thus granting the nonvolatile paradigm fulfillment. Oxides are available in different materials depending on the *Back-End-Of-Line* (BEOL) process. The most common materials are: pure silicon dioxide (SiO_2) for blocking oxides, and either SiO_2 or a barrier engineered stack of Oxide-Nitride-Oxide (SiO_2-Si_3N_4-SiO_2) for tunnel oxides. A 2D planar Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) cell is used as an example in this section [3].

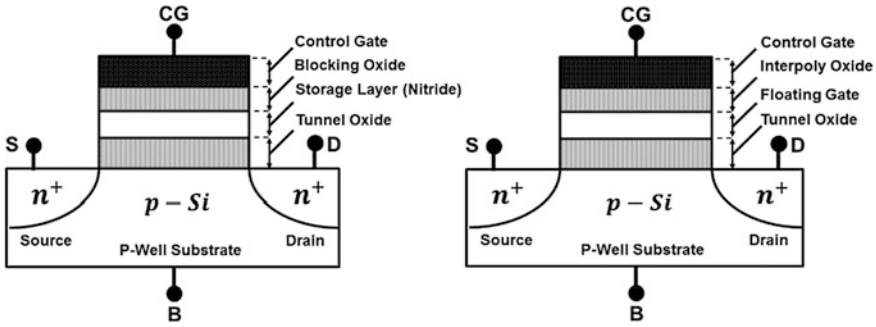

Fig. 2.7 *Left* Example of a charge trap device. *Right* Example of floating gate device

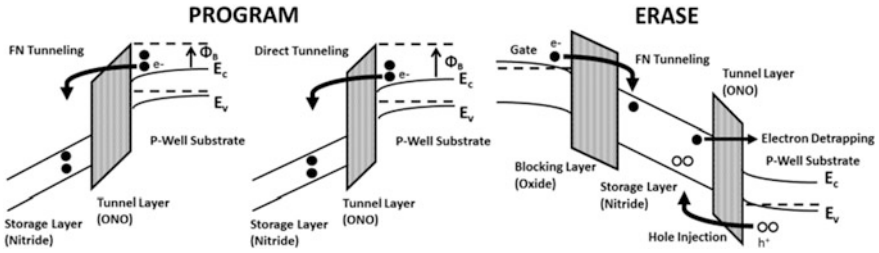

Fig. 2.8 Band diagrams of tunneling mechanisms in planar SONOS CT cell during programming (*left*) and erase (*right*). The two different conditions triggering FN or DT are sketched for programming [3]

High electric fields applied to the tunnel oxide allow electron transfer across the thin insulator to the storage layer. The physical mechanism used for injecting electrons into the storage layer depends on the applied electric field and oxide barrier thickness. In case of high electric fields and large oxide barriers, injection mainly occurs through FN tunneling, whereas in case of low electric field and thin oxide barrier, electrons mainly transfer through *Direct Tunneling* (DT): in this case there is a higher read margin window but retention is worse [3]. In CT cells electron tunneling involves the MOS channel/substrate and it requires appropriate biasing of control gate and bulk terminals (see Fig. 2.8), while drain and source are left floating. Erase operation occurs either through electron detrapping from the storage layer or hole injection from the substrate into the storage layer; at the same time, such operation causes an electron injection from the control gate to the storage layer through FN tunneling, and this is the reason for the well-known "erase saturation" problem [11]. The results of charge separation experiments [12] demonstrate that both electron detrapping and holes injection mechanisms contribute to the erase of a previously programmed CT device: electron detrapping dominates the first part of the transient, whereas hole injection prevails after the removal of the trapped electron charge due to electron emission.

2.5 2D Charge Trap: Reliability Issues

Despite the huge potential, several reliability issues affect CT memories, especially endurance and retention.

2.5.1 Endurance Degradation

The band diagram depicted in Fig. 2.9 describes oxide degradation mechanisms for blocking and tunnel layers. During programming operations (left), electron

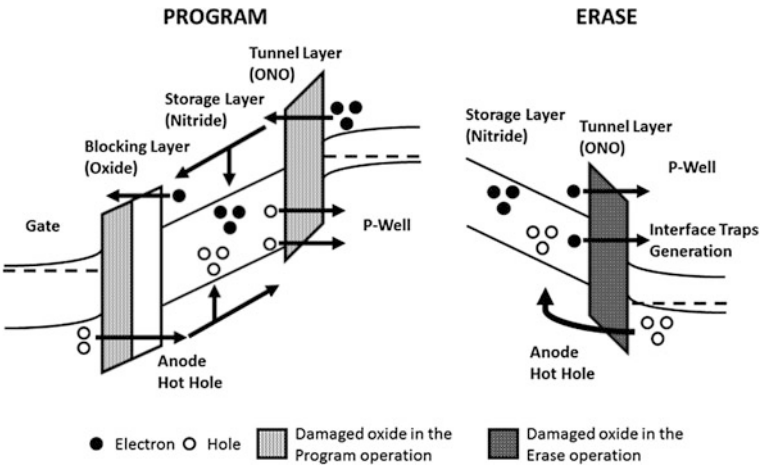

Fig. 2.9 Band diagram sketch of charge transport and trapping/detrapping during program (*left*) and erase (*right*) in planar SONOS CT cell [3]

injection occurs through either FN or DT, damaging the Tunnel Layer; damages to the Blocking Layer are caused by *Anode Hot Hole Injection* (AHHI). Moreover, electrons and holes going through blocking layer and tunnel layer from the storage layer contribute in a marginal, but not negligible, way to oxide degradation. During erasing (right), the hot hole injection from the substrate generates interface traps at the oxide/nitride interface, causing several damages to both storage and tunnel layers, as well as electrons transfer through the tunnel layer [11]. The generation of such interface traps between oxide and nitride interface is the main cause of endurance degradation: in programmed cells, electrons sitting in shallow traps can easily escape via oxide damages induced by cycling, resulting in a charge reduction that may cause read errors.

2.5.2 Data Retention

Data retention is one of the major issues of CT cells, especially at high temperature. Charge loss mechanisms of CT cells has been deeply investigated [13], identifying two main discharging paths: the first is related to thermal excitation of trapped carriers, the second one is due to direct tunneling through the thin tunnel oxide.

The charge loss processes are schematically depicted in Fig. 2.10. For each electron trapped inside the silicon nitride, two discharge mechanisms have to be considered. The first one is the direct *Trap-to-Band* (TB) tunneling from the storage layer traps to the conduction band of the substrate or of the gate; the second one is the thermal emission from traps to the conduction band of the storage layer.

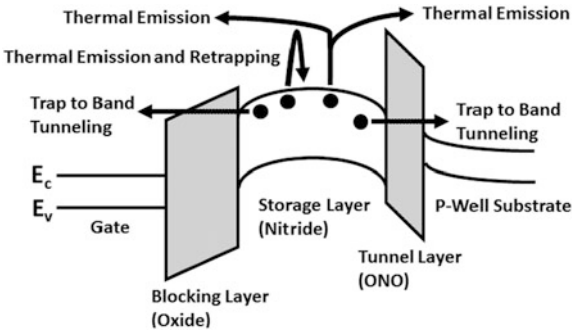

Fig. 2.10 Mechanisms involved in the discharging of programmed planar SONOS CT cell: trap-to-band tunneling through the tunnel layer, trap-to-band tunneling through the blocking layer, thermal emission above the oxide barriers, thermal emission and subsequent retrapping [13]

When thermal emission is considered, the charge loss is the result of two subsequent steps: the emission process and the escape of the electrons towards the bulk and gate electrodes. After emission, retrapping is also possible: the tunneling rate through the oxide barrier of the electrons emitted in the storage layer conduction band could be comparable with the emission and the recapture rates. Here we consider a simplified model where electrons leave the ONO layer only if their energy is higher than the lowest between the tunnel oxide and the top oxide barriers. Consequently, the tunneling of thermally excited carriers towards the bulk and the control gate at energies lower than the oxide conduction band are neglected, assuming that carriers with such an energy are recaptured in the same traps.

In additions, a fast initial charge loss has been observed on a small percentage of cells [14] (see Fig. 2.11). This V_T transient phenomenon has been attributed to the dielectric relaxation effect in the high-k layer, to charge trapping/detrapping, or to mobile charges in the blocking layer [14]. Such mechanism, denoted as *fast detrapping*, is mainly related to electrons trapped in shallow traps which have lower stability than electrons in deep traps; they can easily escape via oxide damages within 1 s after programming.

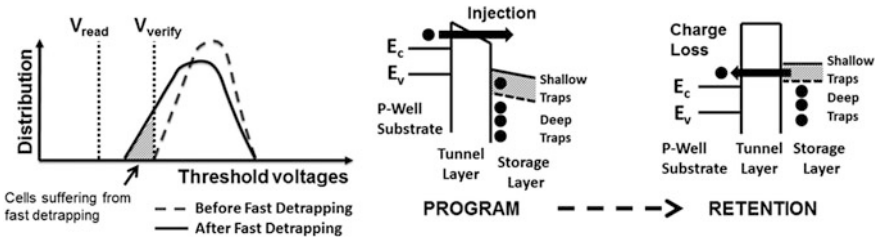

Fig. 2.11 Threshold voltage shift induced by fast detrapping (*left*). Band diagram sketch of fast detrapping effect (*right*) [14]

The same effect is observed after erase too: since the threshold voltage after program/erase does not immediately settle to the final value, there is a wrong estimation of the error bits during the verification step; of course, there is a dependency from the time interval between program/erase and read operations. Waiting for the final V_{TH} would significantly increase the total program/erase time and, of course, this is not acceptable. The transient threshold voltage shift after erase is due to hole re-distribution in the charge trap layer [15].

2.5.3 *Threshold Voltage Shift During Sensing*

Sensing the cell's threshold voltage during retention has been identified as one of the main reliability issues of CT cells [16]. This V_T decrease can be understood within the process of the temperature-activated charge transport through the blocking layer. The charge loss can be minimized when V_T sensing time is decreased down to microseconds. Moreover, blocking oxides engineered by adding a thin SiO_2 layer at the trapping layer/blocking oxide interface exhibit significant suppression of charge loss. Experimental data show that, for identically programmed devices, charge loss rate significantly increases when the V_T sensing operation is repeated more frequently. Furthermore, a similar amount of charge loss is observed when the cumulative sensing time is the same (same numbers of sensing measurements—dashed line in Fig. 2.12). These results indicate that charge loss might be strongly affected by the V_T sensing operation as well as retention time. The charge loss dependency from the V_T sensing time was evaluated by varying the sensing time from microseconds to few seconds. Shorter V_T sensing time is seen to minimize the initial charge loss and significantly reduce charge loss rate.

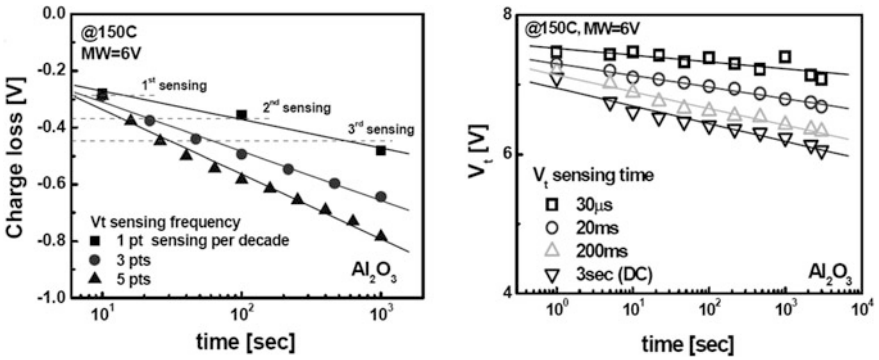

Fig. 2.12 *Left* Retention charge loss as measured by the V_T shift from the programmed V_T value (read window MW = 6 V) for different DC V_T sensing frequencies (V_T sensing time = 3 s). *Right* Programmed V_T dependency on retention time (retention charge loss) for various V_T sensing times [16]

2.6 From 2D to 3D Charge Trap NAND

Three-dimensional architectures appear today as the most viable solutions for the integration of non-volatile memory cells in Tera-bit arrays [17–19]. Two different approaches are possible in order to obtain 3D NAND devices (see Fig. 2.13): the first and simplest is to build the cell on a thin polysilicon substrate as usually done in 2D planar arrays, and stack more levels (Chap. 3) [20]. The second and most interesting approach, defined as vertical channel, is to build a CT cell with a cylindrical channel [17]. Both architectures have a physically large cell size (indicated with the feature process size $F_{process}$), mainly due to a channel width wider than planar devices, although offering a smaller equivalent area occupation due to the stacking of multiple tiers [21, 22]. The former solution does not offer any advantage over conventional planar CT cells in terms of P/E and retention, whereas the latter allows improving the cells programming performance, compared to planar devices, thanks to the shape of the CT cell, also known as *Gate-All-Around* (GAA) [23, 24]. Nevertheless, 3D NAND memories face new reliability issues because of the cylindrical shape and the multi-layer stacking. To understand these new reliability problems, basic concepts of 3D NAND cells are briefly introduced in this section; for more details about each single 3D architecture please refer to Chaps. 4, 5, 6, and 7. In the following sections, the reliability issues affecting 3D devices will be discussed by reviewing the main problems experimentally observed in different 3D NAND arrays [20, 21].

The cross section diagram along a single WL plane of the 3D vertical channel NAND reported in Fig. 2.13 (left) is shown in Fig. 2.14. It is worth highlighting that a *String-Select-Line* (SSL) group is equivalent to a 2D planar flash memory cell array. The 3D vertical flash memory has a nitride layer inside an *Oxide-Nitride-Oxide* (ONO) stack, which acts as a CT layer along the circumference of the thin poly-silicon vertical channel. Please note that each CT cell in this 3D vertical NAND memory is surrounded by the metal gates [24].

The GAA-CT cell is considered one of the most promising solution for 3D integration [25]. This is due to the curvature effect that relaxes the erase saturation problem: the electric field in the blocking oxide is lower than the one in the tunnel

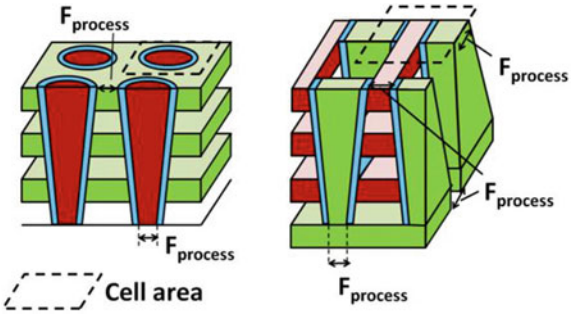

Fig. 2.13 3D NAND vertical (*left*) and horizontal channel (*right*) architectures with the corresponding feature process size $F_{process}$ [21]

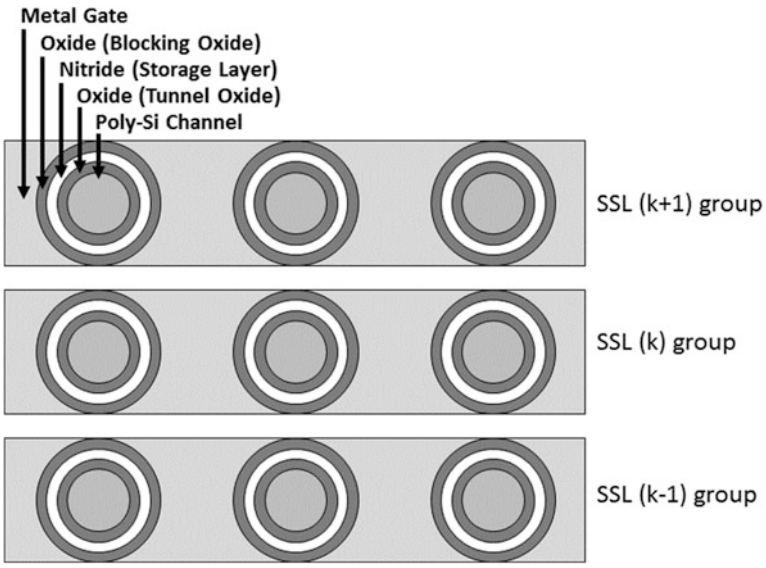

Fig. 2.14 Horizontal section of 3D vertical flash memory cell array [24]

oxide, allowing to increase the electrons detrapping from nitride traps towards the substrate. Moreover, thanks to the reduction of corner and fringing field effects during program, erase and read, GAA-CT cells allow more uniform trapped charge distributions in the storage layer and provide, in turn, steeper incremental step pulse programming (ISPP) transients than planar cells [26].

Figure 2.15 (left) shows a comparison between the GAA-CT cell (solid) and the planar CT cell (dashed), given the same thickness of gate dielectrics, in case of programming operation at V_G = 12 V with neutral nitride. The energy band profile

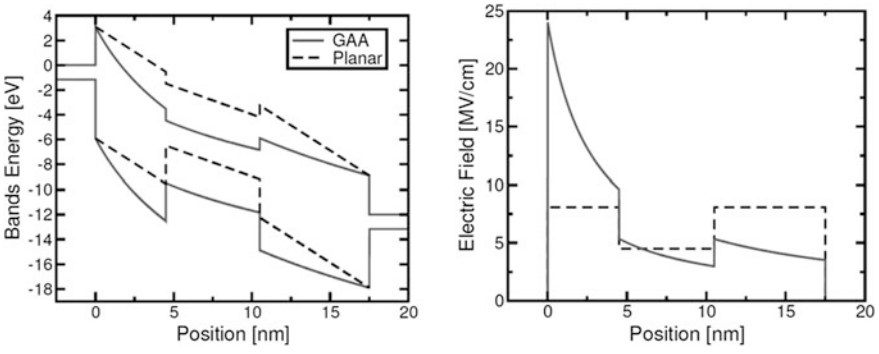

Fig. 2.15 Comparison between a GAA-CT cell and a planar CT cell having the same thickness of the gate dielectrics in terms of energy-band profile (*left*) and electric field (*right*) during program, for V_G = 12 V and neutral nitride [27]

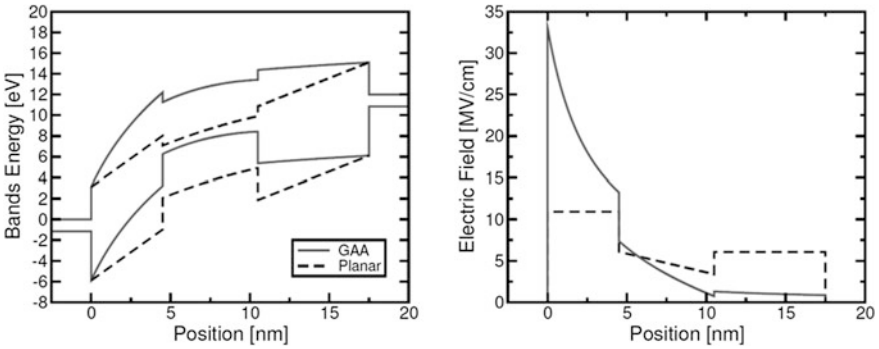

Fig. 2.16 Comparison between a GAA-CT cell and a planar CT cell having the same thickness of the gate dielectrics in terms of energy-band profile (*left*) and electric field (*right*) during erase, for $V_G = -12$ V and neutral nitride [27]

clearly shows a reduction of the thickness of the energy barrier which prevents electron tunneling from the substrate to the nitride. Compared to the planar case, the electric field inside GAA-CT dielectrics is not constant, showing a maximum value at the substrate/tunnel oxide interface. This maximum value is about three times larger than the electric field in the tunnel oxide of the planar device: of course, this is a strong improvement for the programming dynamics [27]. In addition to that, the electric field in the GAA-CT blocking oxide is lower than in the planar case, thus resulting in a reduced electron leakage from the nitride to the gate during programming (see Fig. 2.15 right).

A comparison between GAA-CT and planar cell during erase at $V_G = -12$ V is shown in Fig. 2.16. As for positive V_G, in the case of GAA-CT the electric field reaches a maximum at the substrate/tunnel oxide interface, and it is quite larger than the electric field present in the planar device (right). This behavior enhances the hole tunneling current from the substrate to the nitride during erase. In addition, the lower electric field at the gate/blocking oxide interface prevents electron injection from the gate, thus relieving the erase saturation issues [27].

2.7 3D Charge Trap: Reliability Issues

Even if the transition from 2D to 3D devices can leverage the advantages of the ring shape stack of the memory cell, all the reliability issues affecting planar devices (i.e. endurance, retention and read disturbs) are still there. On top of that, there are new reliability challenges due to vertical charge loss (i.e. through Top/Bottom oxides) and lateral charge migration (i.e. towards spacers). In case of 3D CT arrays the causes are ascribed to either uneven electrical field distribution in *Bottom Oxide* (BTO), equivalent to the blocking oxide in 2D CT, and *Top Oxide* (TPO), equivalent to the tunnel oxide in 2D CT, due to the cylindrical geometry (see Fig. 2.17).

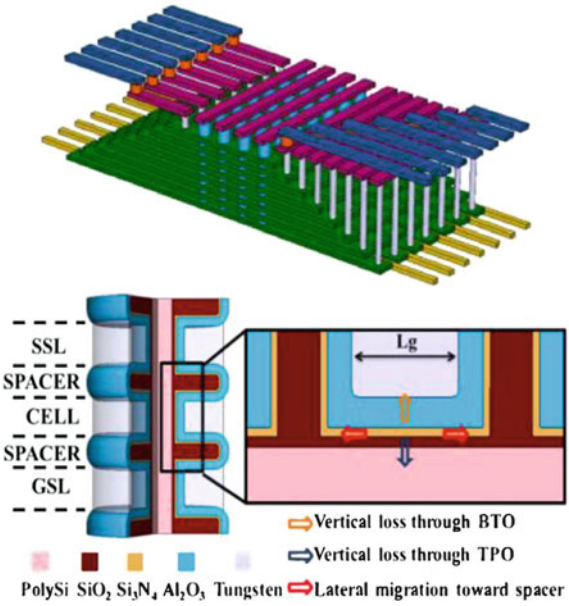

Fig. 2.17 *Top* bird's eye view of 3D CT memory. *Bottom* 3D CT-NAND string with one cell and two select transistors and schematic diagram of the charge loss paths [28]

As a result, charge loss in 3D memories is worse than what we have seen with planar devices, and this is considered the most critical reliability issue for high density and high reliability 3D integration [28]. In this section, physical mechanisms related to the vertical structure of the memory devices, such as vertical charge loss and lateral charge migration, will be described.

2.7.1 *Vertical Charge Loss Through Top and Bottom Oxides*

A constraint for the 3D vertical arrays, not present in planar devices, is that the charge-trap layer cannot be easily interrupted between layers. This fact creates additional leakage paths for the charge, from each cell's active area towards other cells on the same string, as schematically illustrated in Fig. 2.18. In addition to vertical charge loss through Top and Bottom oxides (along Y axis), lateral charge leakage (along X axis) represents an extra source of retention loss for cells in 3D vertical arrays, which should be carefully considered for the reliability assessment of the technology [29].

The involved physical mechanisms accounting for charge distribution evolution during time, along X and Y axes, are illustrated in Fig. 2.19. Charge transport in the conduction band of the charge trapping layer is described based on the drift-diffusion transport scheme. The interaction between free carriers and trapped carriers is governed by the carrier capture phenomenon calculated by *Shockley-Read-Hall*

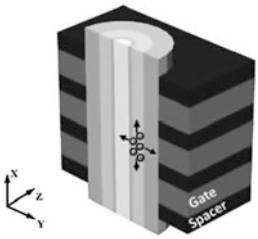

Fig. 2.18 3D CT-NAND structure and charge loss along X and Y axes [29]

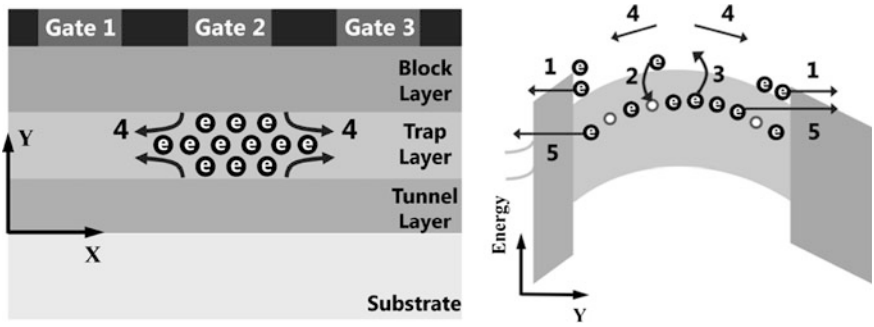

Fig. 2.19 Dominant physical mechanisms along X (*left*) and Y (*right*) axes: 1-DT/FN Tunneling, 2 and 3-carrier capture and emission, 4-drift and diffusion transport, 5-TB tunneling [27]

(SRH) theory, and carrier emission contributed by thermal and Poole-Frenkel effects. Besides, Band-to-Trap (BT) tunneling and Trap-to-Band emission should be taken into account as additional charge capture and loss mechanisms.

Figure 2.20 shows the simulated *Remaining Charge Percentage* (RCP), defined as the percentage of the initial charge which remains inside the storage layer, during

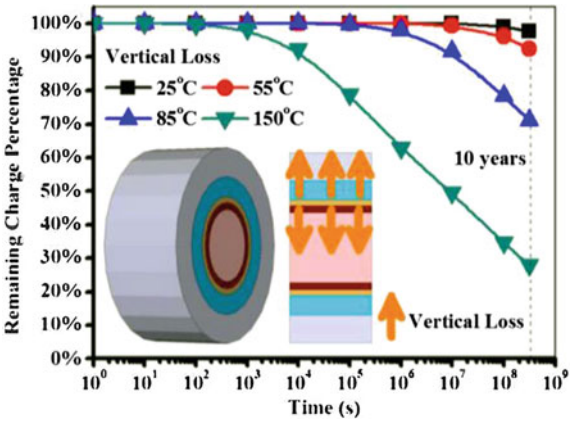

Fig. 2.20 Simulated vertical loss transients. The inset shows the schematic diagram of device structure and the cross-sectional schematic of charge loss paths [28]

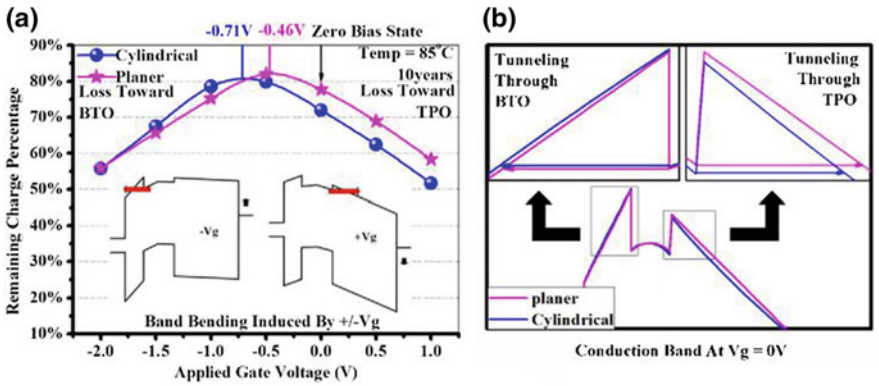

Fig. 2.21 **a** RCP of cylindrical and planar devices, with insets showing the schematic band diagram under positive and negative bias. **b** Conduction band diagram of cylindrical and planar devices at $V_G = 0$ V [28]

vertical charge loss transients [28]. The charge loss characteristics strongly depends on the temperature, which can be explained by enhanced Poole Frenkel emission from trap to conduction band at elevated temperature.

In order to distinguish the role of charge loss through BTO and charge loss through TPO, the RCP under different gate bias is given in Fig. 2.21 (left); in addition, the RCP under different stresses of planar device with identical structure parameters is also plotted as reference. As shown in the energy band diagram, the charge loss through BTO occurs under negative bias while charge loss through TPO dominates with positive bias. Compared to the planar device, charge loss towards TPO is higher and charge loss from BTO is lower in cylindrical devices. This can be explained by the conduction band diagram of Fig. 2.21 (right). Due to the uneven distribution of the electric field in BTO and TPO, the conduction band graphs of the cylindrical device are not straight any longer, but convex in BTO and concave in TPO. As a result, charge loss through BTO is slightly reduced [28].

2.7.2 *Lateral Migration Towards Spacers*

Lateral charge migration towards the spacer (i.e. the region between each layer) region (along X axis, referring to Fig. 2.18) is another key path for the charge loss due to the difficulty of cutting the CT layer between memory cells. Different shapes of the CT layer exhibit different lateral migration performances. To analyze this shape dependency, *Bit Cost Scalable* (BiCS)-type structures and *Terabit Cell Array Transistor* (TCAT)-type structures have been studied [28]; more details about both arrays can be found in Chap. 4. In order to focus on lateral migration performance, the cells were programmed to the same threshold voltage (6 V). Lateral migration is accelerated by temperature (Fig. 2.22). The considerable threshold voltage loss

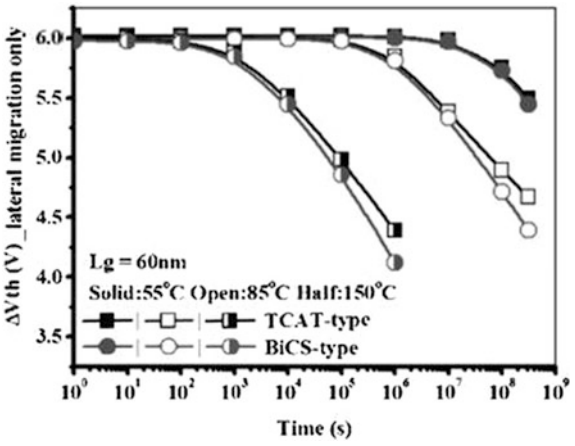

Fig. 2.22 Comparison between TCAT-type and BiCS-type devices [28]

caused by the lateral migration indicates that carriers in the nitride significantly migrate laterally. Figure 2.23a shows the distribution of trapped carriers versus elapsed time in TCAT-type and BiCS-type devices: lateral migration of trapped charges can clearly be observed as time proceeds. TCAT-type devices show better retention characteristics. Shape dependency of lateral migration can be explained by the lateral charge profile evolution shown in Fig. 2.23b: the corner of TCAT-type devices suppresses the migration of trapped charges, which can be described in device-level simulations not only by the higher charge density along the channel direction, but also by the sharp peak at the corner of the CT layer.

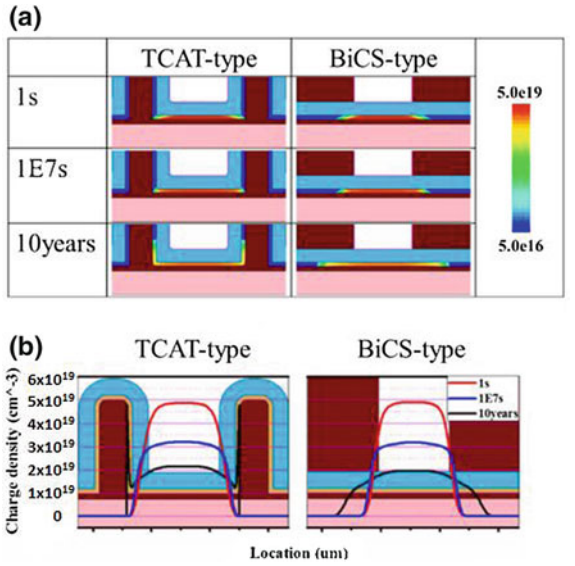

Fig. 2.23 **a** Simulated trapped charge distribution at different elapsed time and T = 85 °C. **b** Simulated lateral charge profile evolution (cutline at *middle* of CT layer along the channel) [28]

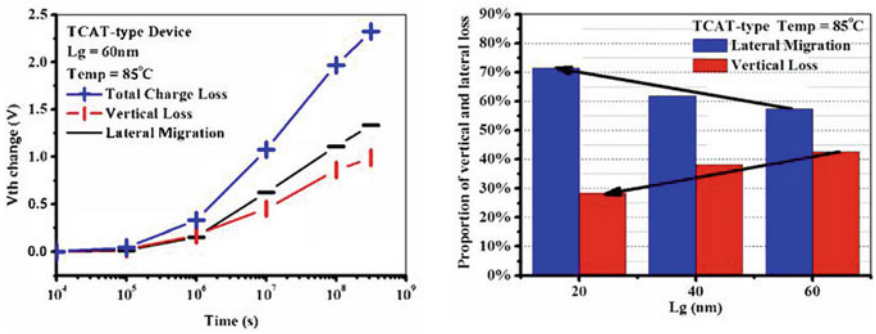

Fig. 2.24 Split of threshold voltage change by vertical loss and lateral migration versus time (*left*) and versus channel length (right) for TCAT devices [28]

Based on the above discussion, the impact of vertical loss and lateral migration in retention has been compared for TCAT and summarized in Fig. 2.24 (left). It can be seen that lateral migration is the dominant charge loss mechanism. The charge loss behavior with different channel length is shown in Fig. 2.24 (right). As the channel length reduces, lateral migration accounts for a larger percentage of charge loss, which indicates that lateral migration should be a more critical issue than vertical loss in high-density and high-reliability design of 3D GAA-CT memories.

2.7.3 *Transient V_T Shift*

Transient V_T shift after erase, previously described for 2D cells (Sect. 2.5.2), is observed on 3D devices too. In the GAA-CT cell, a smaller diameter of silicon nanowire shows a better erase efficiency due to the electric field concentration effect on the tunnel oxide (Fig. 2.25, left). However, the GAA-CT device also shows a transient V_T shift after erase: the amount of V_T shift is related to the amount of read window (defined as the voltage difference between the 'programmed' and 'erased' states in case of SLC architectures, or between two adjacent levels in case of MLC architectures) and the V_T shift in GAA-CT is well correlated to that in planar CT devices (Fig. 2.25, right), thus implying the same mechanism. It is worthwhile noting that the transient V_T shift can be reduced by scaling the channel length (L_G) and the diameter of the silicon nanowire (W_{NW}) in GAA-CT, probably due to a compensation effect by charge crowding and lateral charge spreading. As shown in Fig. 2.26, when the diameter of nanowire goes below 6 nm, the tendency of the drain current I_D to increase with time disappears. Therefore, 3D GAA-CT device with small nanowire diameter shows advantages in fast erase operation [15].

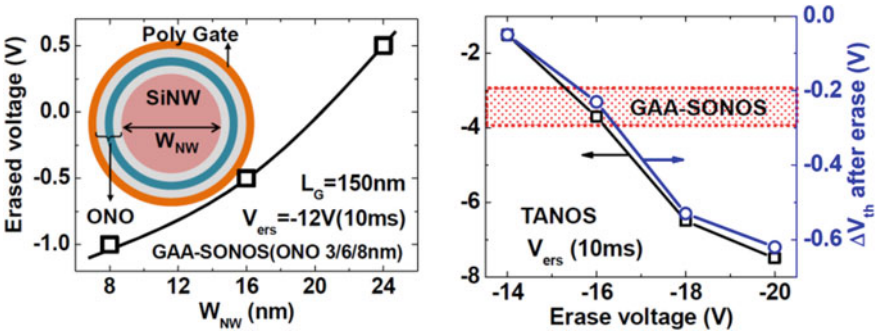

Fig. 2.25 *Left* Erased saturation voltage versus W_{NW} in GAA-CT device. A smaller W_{NW} shows a better erase efficiency due to the electric field concentration effect on the tunnel oxide. *Right* Correlation of V_T window and V_T shift after erase operation. The V_T shift in GAA-CT is well correlated to that in planar CT devices, implying the same mechanism for the transient V_T shift [15]

2.7.4 Program and Pass Disturbs

All kinds of 3D NAND suffer from two program disturbs. Besides the traditional program and pass disturbs affecting 2D architectures (see Fig. 2.6), in 3D NAND also the disturbs related to the vertical structure are to be taken into account.

2.7.5 Vertical Hole Design Limitations

Since scaling and design of 3D NAND are completely different from planar NAND, and with different implications on the memory reliability, new methodologies are

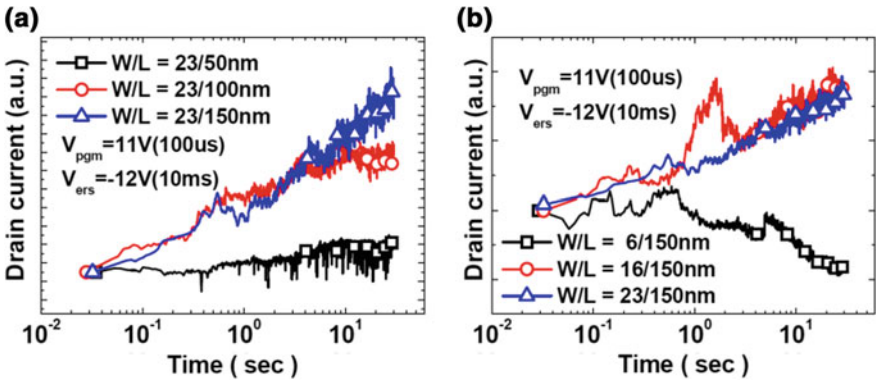

Fig. 2.26 Transient I_D of GAA-CT devices with different L_G (**a**) and W_{NW} (**b**). Transient V_T shift becomes smaller when L_G and the W_{NW} are scaled. I_D fluctuations in GAA-CT may be due to single electron effects or random telegraph noise [15]

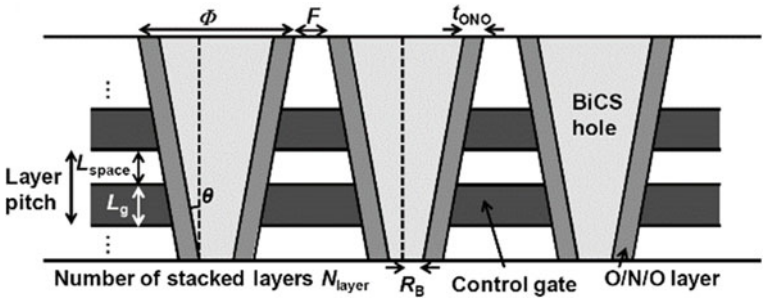

Fig. 2.27 Cross sectional view of 3D NAND [30]

required. One of the problems of 3D NAND is the reduced cells density per single memory layer. As a matter of example, Fig. 2.27 shows the simplified cross sectional view of the Bit-Cost Scalable (BiCS) type 3D NAND. The BiCS hole must be filled with ONO film (~20 nm) and silicon channel. Since it is not possible to aggressively scale the ONO film because of read window and reliability, the diameter of the BiCS hole is not so scalable. Therefore, the number of stacked layers (N_{layer}) should be increased to compensate this drawback. Moreover, as shown in Fig. 2.27, there is an additional limitation due to the finite taper angle θ in the BiCS hole: the memory cell at the top of the stack is always larger than the cell at the bottom. In other words, once the bottom of the stack reaches the minimum cell's size of a specific technology, then the area can't be shrunk anymore. On the other hand, since the minimal line and space lithography pattern are not required for the control gate (CG) formation in 3D NAND, CG length L_g and spacing L_{space} can be independently chosen. This design flexibility is allowed for 3D NAND. Therefore, suitable device design for 3D NAND can be explored in terms of L_g and L_{space}. Programming and disturbance characteristics are evaluated for the memory operation, as shown in next sections [30].

2.7.5.1 V_T Shift Induced by Stored Electrons During Programming

Given a specific electron density in the nitride layer, 3D and 2D cells can be compared in terms of the resulting V_T shift of the programmed cell, as shown in Fig. 2.28 [30]. L_g and L_{space} should not be too small and too large, respectively. The smallest shift is observed at L_g = 10 nm and L_{space} = 50 nm. When L_{space} is large, the spacing region determines the V_T of the cell. Therefore, the effect of the stored electrons on V_T relatively decreases for large L_{space} and the corresponding read window decreases. Large L_g shows high V_T shift because the potential of the center region of the channel in the target cell is mainly controlled by the stored electrons.

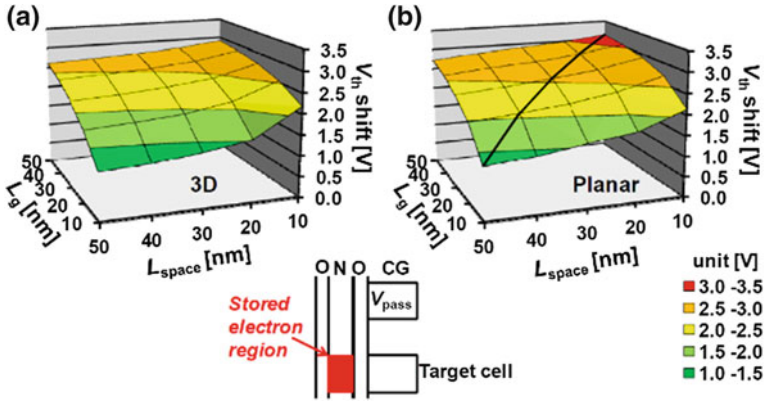

Fig. 2.28 V_T shift of the programmed cell (corresponding to read window) as a function of L_g and L_{space} when electrons (electron density: 1×10^{19} cm^{-3}) are stored in 3D NAND (**a**) and planar NAND (**b**) [30]

2.7.5.2 V_T Shift Induced by Neighboring Cells

V_T shift induced by the electrons stored in the neighboring cell was investigated as shown in Fig. 2.29 [30]. As L_g and L_{space} decrease below 20–30 nm, V_T shift due to the neighboring cell drastically increases in both 3D and planar NAND. When L_{space} is small, stored electrons couple with the channel of the target cell. This is severe when L_g is small because stored electrons potential can affect the entire channel region of the target cell. Moreover, at small L_g and L_{space}, the V_T shift degradation in 3D NAND is higher than in planar NAND. The substrate-channel coupling in planar NAND reduces the V_T shift by because of the coupling between the channel and stored electrons.

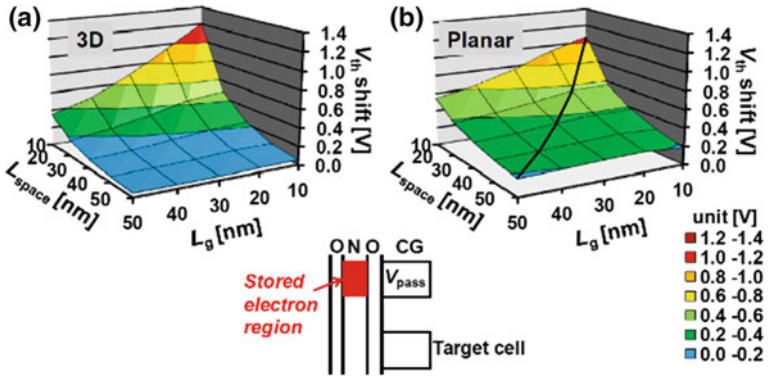

Fig. 2.29 V_{th} shift as a function of L_g and L_{space} when electrons (electron density: 1×10^{19} cm^{-3}) are stored in the neighboring cell in 3D NAND (**a**) and planar NAND (**b**) [30]

2.7.5.3 Electric Field in the Tunnel Oxide During Programming

Evaluations of the electric field in the tunnel oxide during the programming has been performed [30]: Fig. 2.30 plots the electric field of the tunnel oxide in the channel (NAND string) direction. The figure shows that the electric field spreads out in the lateral direction (fringing electric field exists). If L_g is small, the tunnel oxide electric field cannot concentrate at the center of the CG of the programmed cell. Therefore, the electric field in the tunnel oxide of the programmed cell (E_{ox_pgm}) decreases at small L_g. If L_{space} is small, then the electric field penetrates into the neighboring cells; however, if L_g is large, the penetration is only at the edge of the cell. Therefore, the electric field at the center of the CG remains low for the neighboring cells (low E_{ox_ngb}).

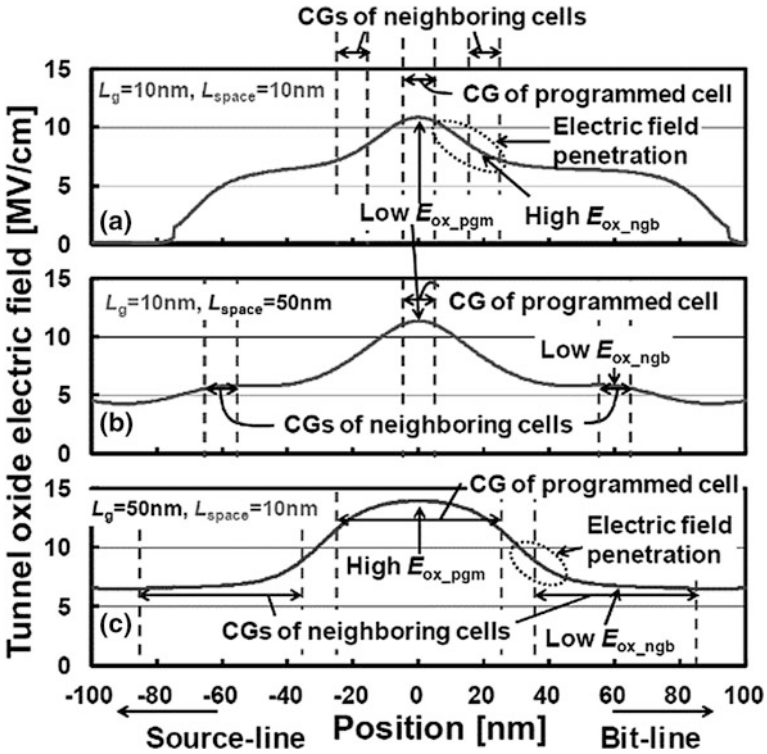

Fig. 2.30 Electric field of the tunnel oxide in the channel direction. 3D NAND string with: L_g = 10 nm and L_{space} = 10 nm (**a**), L_g = 10 nm and L_{space} = 50 nm (**b**), L_g = 50 nm and L_{space} = 10 nm (**c**) [30]

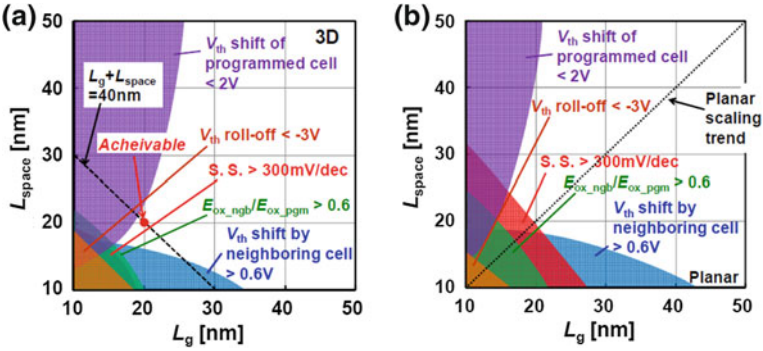

Fig. 2.31 L_g and L_{space} design window for 3D NAND (**a**) and planar NAND (**b**) [30]

2.7.5.4 Design Window of L_G and L_{SPACE}

Figure 2.31 shows the design window of L_g and L_{space} for 3D (a) and planar (b) NAND [30]. The criteria for unacceptable regions (shaded regions in Fig. 2.31) are assumed as follows; V_T roll-off < −3 V, Subthreshold Slope (S. S.) > 300 mV/dec, V_T shift < 2 V in the programmed cell, V_T shift > 0.6 V by the neighboring cell and $E_{ox\_ngb}/E_{ox\_pgm} > 0.6$. $L_g = L_{space} = 20$ nm (layer pitch of 40 nm) is achievable in 3D NAND in terms of the electrical characteristics. Same L_g and L_{space} are preferable to cope with the tradeoff between the large V_T shift for the programmed cell and the small V_T shift induced by the neighboring cell. For further improvements, the diameter of the BiCS hole should be decreased. Table 2.1 summarizes the comparison between 3D and 2D: 3D NAND achieves very good on-current (I_{on}), S.S. and low program voltage (V_{pgm}) compared to planar NAND. Slight degradations in V_T roll-off and V_T shift caused by the stored electrons in the neighboring cell are observed only at the small L_g and L_{space} region.

2.8 3D CT Versus State-of-the-Art 2D FG

In this section, a comparison between a particular 3D CT NAND (denoted as *Stacked Memory Array Transistor*—SMArT [31], depicted if Fig. 2.32) and state-of-the-art 2D FG is reported, in terms of both performances and reliability. A detailed description of 2D FG reliability problems is provided in [1].

V_T distribution widths of MLC 3D CT cells are ∼30 % smaller compared to the 2y-nm FG, because of their interference free nature (Fig. 2.33, left). The widening of cells V_T distributions during cycling is compared in Fig. 2.33 (right), where

Table 2.1 Summary of 3D NAND cells [30]

	I_{on}	V_{th} roll-off	S.S.	V_{th} shift (programmed cell)	V_{th} shift (neighboring cell)	Tunnel oxide electric field (E_{ox_ngb}/ E_{ox_pgm})	V_{pgm}
Planar NAND	Poor	Fair	Poor	Fair	Fair	Fair	20 V
3D NAND	Very good	Poor at small L_g, L_{space}	Very good	Fair	Good at large L_g and L_{space}, poor at small L_g and L_{space}	Fair	17 V
Preferable scaling parameter	L_{space}	–	–	L_{space}	L_g	–	L_{space}

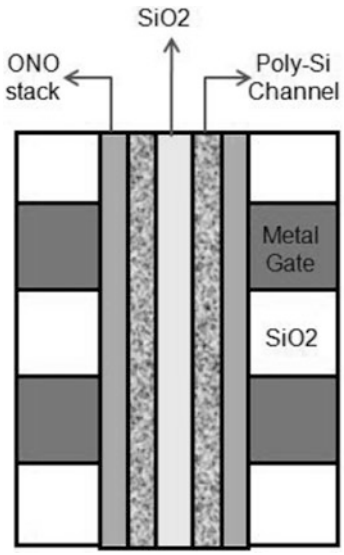

Fig. 2.32 SMArT cell schematic [31]

SMArT cells show no widening up to 5k cycles, whereas FG cells start broadening from 3k [31].

On the other hand, 3D CT have worst retention performances as shown in Fig. 2.34, where the post cycling V_T shifts at high temperature are compared. In 3D CT cells the V_T shifts are so large that the distributions are no longer separated [31].

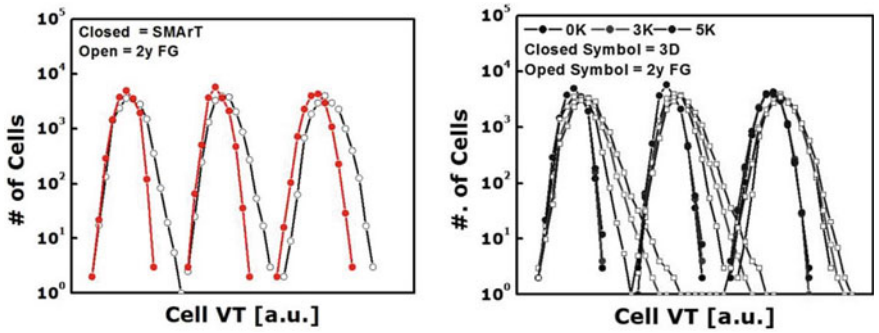

Fig. 2.33 *Left* Comparison of cells V_T distributions of 2y node FG and SMArT cells. *Right* Comparison of V_T widening during program-erase cycling [31]

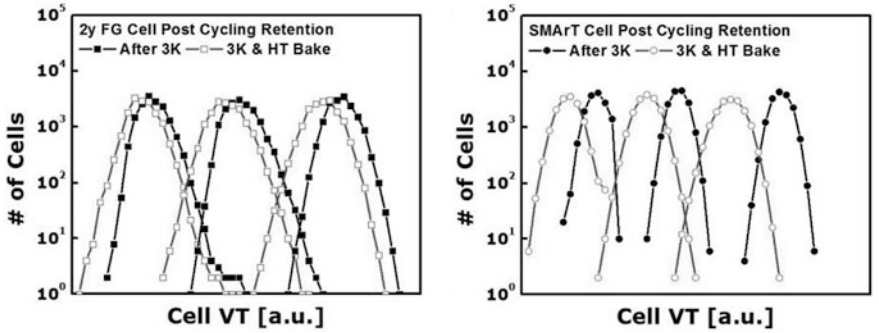

Fig. 2.34 Cells V_T distributions of HT retention after cycling of 2y node FG (*left*) and SMArT cells (*right*) [31]

2.9 3D-FG Nand

Recently, 3D vertical FG type NAND cell arrays have been proposed to overcome the retention and overall reliability issues of 3D CT NAND cell arrays [32–36]. In this section, an overview of the proposed 3D FG cells and their main reliability problems is reported. Figure 2.35 shows the bird's-eye view of published 3D vertical FG type NAND cell's structures: *Extended Sidewall Control Gate* (ESCG) [32], *Dual Control-gate with Surrounding Floating-gate* (DC-SF) [33, 34], and *Separated-Sidewall Control Gate* (S-SCG) cells [35].

ESCG and DC-SF cells suffer from interference and disturbance problems when integrated into an array due to the direct coupling effect of neighboring cells in the same string. S-SCG overcomes such a problem, strongly reducing interference and disturbance effects. In the S-SCG structure, the Source/Drain (S/D) region can be implemented by electrically inverting the pillar surface, and high CG coupling capacitance can be achieved. S-SCG structure allows obtaining highly reliable

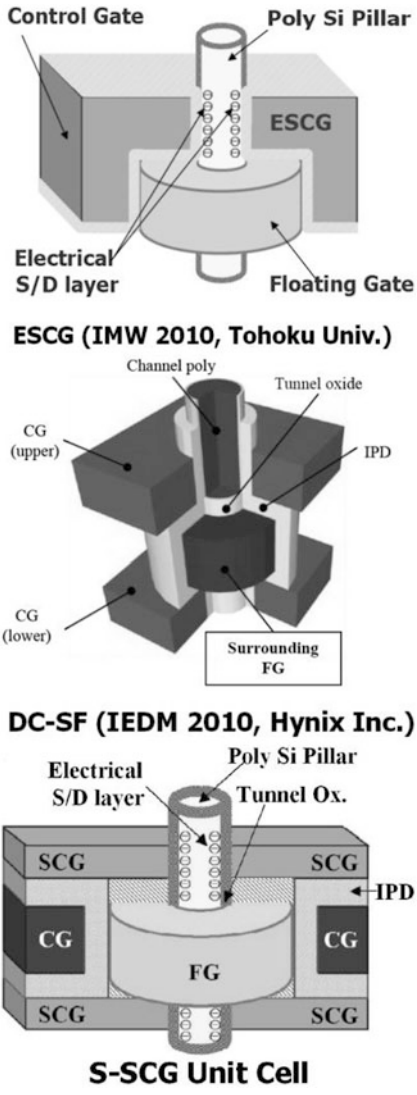

Fig. 2.35 Bird's eye views of the most recent 3-D vertical FG type NAND cell schemes: ESCG (*top*), DC-SF (*middle*) and S-SCG (*bottom*) [35]

MLC operation, high speed P/E operation and good read current margin. More details about these architectures can be found in Chap. 5.

2.9.1 DC-SF Interference and Retention Results

The interference between a programmed cell and an adjacent cell in case of 3D-FG NAND array with DC-SF cells was studied as shown in Fig. 2.36 (left) [33].

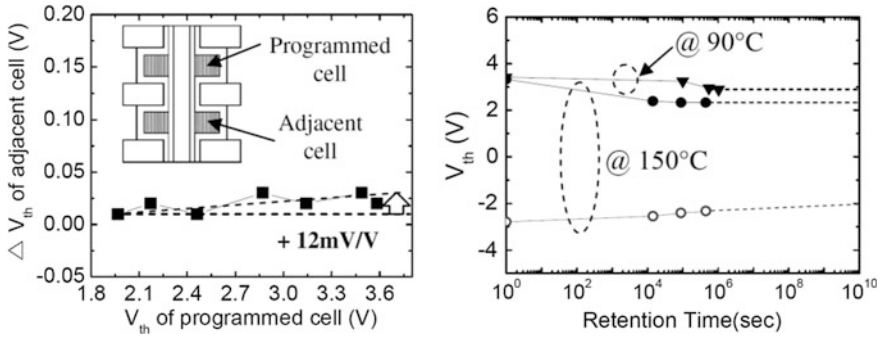

Fig. 2.36 *Left* FG-FG interference characteristics (V_T variation in the adjacent cell as a function of V_T in the programmed cell). Very small FG-FG coupling value of 12 mV is obtained. *Right* Data retention characteristics of the DC-SF NAND flash cells [33]

Negligible V_T shift of adjacent cell (12 mV) was obtained as the V_T of programmed cell is increased by 3.6 V, suggesting that each CG acts as shielding layer in a string. The data retention characteristics of DC-SF cells at two different temperatures (90 and 150 °C) are reported in Fig. 2.36 (right), showing the increase of charge loss with temperature. In case of high temperature condition, a program and erase V_T shift due to charge-loss of 0.9 and 0.2 V after a retention time of 126 h is shown, respectively.

2.9.2 S-SCG Interference Results

In 3D FG NAND arrays with S-SCG cells two critical interference coupling paths exist: the former is the indirect coupling path while the latter is direct. Figure 2.37 shows the interference effects as a function of the S-SCG initial V_T. The S-SCG structure can sufficiently suppress the indirect interference effect; however, the direct coupling from neighboring FG to the channel of S-SCG remains a very serious problem. ESCG and DC-SF cells have remarkable interference problems by this direct coupling effect as well, which directly influences the parasitic transistor below the S-SCG. To suppress this direct coupling effect, S-SCG cell applies the SCG voltage to control the parasitic transistor [35].

2.9.3 S-SCG Performance and Reliability Advantages

The S-SCG cell strongly reduces both the interference effect and the disturbance problem with good performance, and it has good potentials of highly reliable MLC operation. Moreover, lower operation voltages than conventional 3D CT is

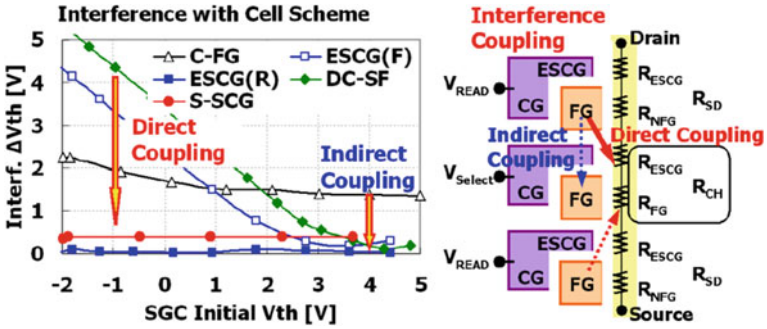

Fig. 2.37 Interference effect in 3-D vertical FG NAND cells (*left*) and cross sectional view of interference coupling paths of the conventional ESCG cell array (*right*) [35]

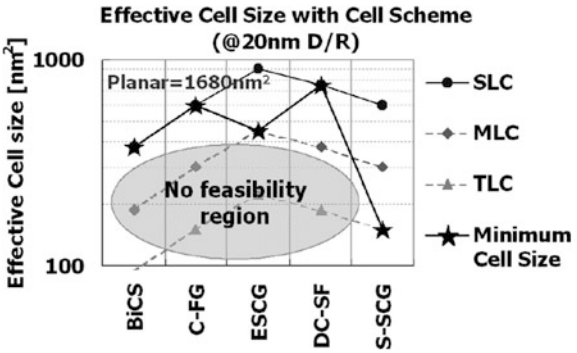

Fig. 2.38 The effective cell size of S-SCG cell in comparison with other 3-D vertical NAND cells at 20 nm technology [35]

required: this implies that cell operation is more effective because of higher coupling ratio. The vertical cell height is decreased compared to that of the conventional FG cell by using a predeposited SCG layer. Figure 2.38 shows the effective cell size with 3D vertical NAND cell schemes at 20 nm technology. Although the

Table 2.2 3D NAND cells comparison summary [35]

D/R=20nm Technology (Aspect Ratio=32)			CT type	FG type			
			BiCS	C-FG	ESCG	DC-SF	S-SCG
Cell Perfromances	CG C/R		-	NG	G	VG	VG
	Reliability		NG	G	G	G	G
	Disturbance		G	G	NG	NG	VG
	Interference		G	NG	G	NG	VG
MLC feasibility (bit/cell)			Not Easy (1bit)	Not Easy (1bit)	Normal (2bit)	Not Easy (1bit)	Easy (3bit)
Minimum IPD		nm	-	-	12	12	7
Minimum Gate Elctrode		nm	-	-	6	6	6
Cell height		nm	40	40	60	50	40
# of stacked cell		ea	16	16	11	13	16

G=Good, NG=Not Good, VG=Very Good

cell size of the proposed S-SCG cell is larger than that of CT type NAND cell by about 60 %, the possibility of implementing MLC operation allows obtaining lower bit costs. Less than half of the bit cost can be achieved by implementing TLC operation to the proposed S-SCG cell. Finally, we show the MLC feasibility and the number of stacked cells in comparison with conventional 3D cells in Table 2.2 [35].

2.10 3D-CT Versus 3D-FG

In this section a final comparison between 3D CT and 3D FG in terms of performance and reliability is reported. For the comparison of the structures, vertical schematic of the conventional 3D CT structure and 3D FG cell are shown in Fig. 2.39 [34]. Unlike planar CT device, the CT nitride layer in a string of the conventional 3D CT is continuously connected from top to bottom CGs along the channel side, and it acts as a charge spreading path, which is an unavoidable problem of 3D CT cell. As a result, this causes degradation of data retention characteristics and poor distribution of cell state. In 3D FG cells, on the contrary, the FG is completely isolated by the tunnel oxide and the Inter Poly Dielectric (IPD). This approach allows obtaining a significantly reliable structure, able to contain charges without any problem related to leakage paths [34].

In order to compare the cell size, DC-SF as 3D FG and BiCS/TCAT as 3D CT are considered. The effective cell size is estimated and plotted in Fig. 2.40 as a function of the number of stacked cells. Even if the physical size of DC-SF cell is assumed to be 54 % larger than that of conventional BiCS/TCAT, DC-SF allows fabricating 1 Tb arrays with 3 bit/cell and 64 stacked cells or 2 Tb arrays with 3 bit/cell and 128 stacked cells, thanks to the small FG-FG interference.

Moreover, 3D FG ensure reliable retention characteristics and lower operation voltage than that of conventional 3D CT [34]. As a result, the 3D FG structure allows highly enhanced device performance for 3D NAND flash memory compared to 3D CT. In the following, a summary of 3D FG advantages and disadvantages compared to 3D CT is reported [37]:

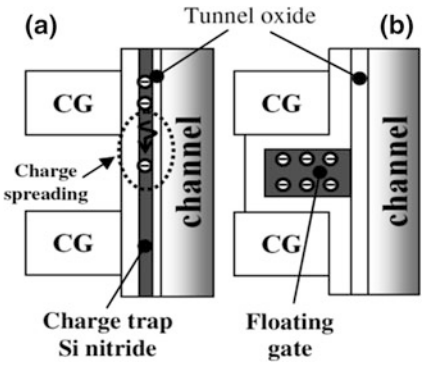

Fig. 2.39 Comparison of 3D NAND flash cell structures (**a**) CT cell (BiCS) (**b**) 3D-FG cell [34]

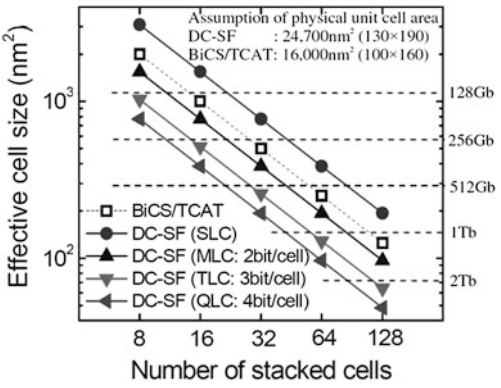

Fig. 2.40 Effective cell sizes for various DC-SF NAND flash structures [34]

Advantages of 3D FG:

- Lower charge spreading resulting in less read errors, and consequently less ECC intervention, in particular in multilevel architectures;
- better data retention because of a more stable charge into the storage layer;
- direct connection between channel poly and p-well, allowing bulk erase (no GIDL).

Disadvantages of 3D FG:

- Larger cell size;
- larger 3D pillar due to the presence of the floating gate, and hence lower scalability;
- floating gate coupling effect: even if the S-SCG structure can sufficiently suppress the indirect interference effect, direct coupling remains a very serious problem that strongly reduces the programming speed.

Even if 3D FG shows a relevant number of reliability advantages compared to 3D CT, the higher scalability still makes 3D CT the most attractive solution for the integration in hyper-scaled arrays. Moreover, it must be pointed out that enhanced programming algorithms and error correction techniques allows mitigating the previously described reliability issues.

References

1. R. Micheloni et al., *Inside NAND Flash Memories* (Springer, 2010)
2. C. Lee et al., Multi-level NAND flash memory with 63 nm-node TANOS (Si-Oxide-SiN-Al_2O_3-TaN) cell structure, in *VLSI Symposium Technical Digest* (2006), pp. 21–22
3. G. Van Den Bosch, Physics and reliability of 2D and 3D SONOS devices, in *IEEE International Memory Workshop (IMW)*, Tutorial, 18–21 May 2014

4. A. Grossi et al., Bit error rate analysis in charge trapping memories for SSD applications, in *IEEE International Reliability Physics Symposium (IRPS)*, June 2014, pp. MY.7.1–MY.7.5
5. Y. Cai et al., Threshold voltage distribution in MLC NAND flash memory: characterization, analysis, and modeling, in *Design, Automation Test in Europe Conference Exhibition (DATE)*, Mar 2013, pp. 1285–1290
6. A. Chimenton et al., A statistical model of erratic behaviors in NAND flash memory arrays. IEEE Trans. Electron Devices **58**, 3707–3711 (2011)
7. T. Ong et al., Erratic erase in ETOX TM flash memory array, in *Proceedings of VLSI Symposium Technical*, pp. 83–84 (1993)
8. C. Dunn et al., Flash EPROM disturb mechanisms, in *Proceedings of IEEE International Reliability Physics Symposium (IRPS)*, Apr 1994, pp. 299–308
9. C. Zambelli et al., Analysis of edge wordline disturb in multimegabit charge trapping flash NAND arrays, in *IEEE International Reliability Physics Symposium (IRPS)*, 10–14 Apr 2011, pp. MY.4.1–MY.4.5
10. J. Lee, C. Lee, M. Lee, H. Kim, K. Park, W. Lee, A new programming disturbance phenomenon in NAND flash memory by source/drain hot electrons generated by GIDL current, in *Proceedings of the NVSM Workshop* (2006), pp. 31–33
11. A. Arreghini et al., Experimental extraction of the charge centroid and of the charge type in the P/E operations of the SONOS memory cells, in *IEDM 2006 Technical Digest* (2006), pp. 499–502
12. L. Vandelli et al., Role of holes and electrons during erase of TANOS memories: evidences for dipole formation and its impact on reliability, in *IEEE International Reliability Physics Symposium (IRPS)*, 2–6 May 2010, pp. 731–737
13. A. Arreghini et al., Characterization and modeling of long term retention in SONOS non volatile memories, in *Solid State Device Research Conference, 37th European ESSDERC*, 11–13 Sept 2007, pp. 406–409
14. C.-P. Chen et al., Study of fast initial charge loss and it's impact on the programmed states VT distribution of charge-trapping NAND flash, in *IEEE International Electron Devices Meeting (IEDM)* (2010), pp. 5.6.1–5.6.4
15. J.K. Park et al., Origin of transient Vth shift after erase and its impact on 2D/3D structure charge trap flash memory cell operations, in *IEEE International Electron Devices Meeting (IEDM)*, 10–13 Dec 2012, pp. 2.4.1–2.4.4
16. H. Park et al., Charge loss in TANOS devices caused by Vt sensing measurements during retention, in *IEEE International Memory Workshop (IMW)*, 16–19 May 2010, pp. 1–2
17. H. Tanaka et al., Bit cost scalable technology with punch and plug process for ultra high density flash memory, in *IEEE Symposium on VLSI Technology*, 12–14 June 2007, pp. 14–15
18. J. Jang et al., Vertical cell array using TCAT (terabit cell array transistor) technology for ultra high density NAND flash memory, in *IEEE Symposium on VLSI Technology*, 16–18 June 2009, pp. 192–193
19. S.J. Whang et al., Novel 3-dimensional dual control-gate with surrounding floating-gate (DC-SF) NAND flash cell for 1 Tb file storage application, in *IEEE International Electron Devices Meeting (IEDM)*, 6–8 Dec 2010, pp. 29.7.1–29.7.4
20. W. Kim et al., Multi-layered vertical gate NAND flash overcoming stacking limit for terabit density storage, in *IEEE Symposium on VLSI Technology*, 16–18 June 2009, pp. 188–189
21. A. Goda et al., Scaling directions for 2D and 3D NAND cells, in *IEEE International Electron Devices Meeting (IEDM)*, 10–13 Dec 2012, pp. 2.1.1–2.1.4
22. H.T. Lue et al., 3D vertical gate NAND device and architecture, in *IEEE International Memory Workshop (IMW)*, Tutorial, 18–21 May 2014
23. Y. Fukuzumi et al., Optimal integration and characteristics of vertical array devices for ultra-high density, bit-cost scalable flash memory, in *IEDM Technical Digest* (2007), pp. 449–452
24. Y. Kim et al., Coding scheme for 3D vertical flash memory, in *IEEE International Conference on Communications (ICC)*, 8–12 June 2015

25. E. Nowak et al., In-depth analysis of 3D silicon nanowire SONOS memory characteristics by TCAD simulations, in *IEEE International Memory Workshop (IMW)*, 16–19 May 2010, pp. 1–4
26. H.-T. Lue et al., Understanding STI edge fringing field effect on the scaling of charge-trapping (CT) NAND flash and modeling of incremental step pulse programming (ISPP), in *IEDM Technical Digest* (2009), pp. 839–842
27. S.M. Amoroso et al., Semi-analytical model for the transient operation of gate-all-around charge-trap memories. IEEE Trans. Electron Devices **58**(9), 3116–3123 (2011)
28. X. Li et al., Investigation of charge loss mechanisms in 3D TANOS cylindrical junction-less charge trapping memory, in *IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)*, 28–31 Oct 2014, pp. 1–3
29. Z. Lun et al., Investigation of retention behavior for 3D charge trapping NAND flash memory by 2D self-consistent simulation, in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 9–11 Sept 2014, pp. 141–144
30. Y. Yanagihara et al., Control gate length, spacing and stacked layer number design for 3D-stackable NAND flash memory, in *IEEE International Memory Workshop (IMW)*, 20–23 May 2012, pp. 1–4
31. E.-S. Choi et al., Device considerations for high density and highly reliable 3D NAND flash cell in near future, in *IEEE International Electron Devices Meeting (IEDM)*, 10–13 Dec 2012, pp. 9.4.1–9.4.4
32. M.K. Seo et al., The 3-dimensional vertical FG NAND flash memory cell arrays with the novel electrical S/D technique using the extended sidewall control gate (ESCG), in *IEEE International Memory Workshop (IMW)*, May 2010, pp. 146–149
33. S. Aritome et al., Advanced DC-SF cell technology for 3-D NAND flash. IEEE Trans. Electron Devices **60**(4), 1327–1333 (2013)
34. S.J. Whang et al., Novel 3-dimensional dual control-gate with surrounding floating-gate (DC-SF) NAND flash cell for 1Tb file storage application, in *IEEE International Electron Devices Meeting (IEDM)*, 6–8 Dec 2010, pp. 29.7.1–29.7.4
35. M.K. Seo et al., A novel 3-D vertical FG NAND flash memory cell arrays using the separated sidewall control gate (S-SCG) for highly reliable MLC operation, in *IEEE International Memory Workshop (IMW)*, 22–25 May 2011, pp. 1–4
36. K. Parat et al., A floating gate based 3D NAND technology with CMOS under array in *IEEE International Electron Devices Meeting (IEDM)*, 7–9 Dec 2015
37. B. Prince, 3D vertical NAND flash revolutionary or evolutionary, in *IEEE International Memory Workshop*, Tutorial, 17–20 May 2015

http://www.springer.com/978-94-017-7510-6

3D Flash Memories
Micheloni, R. (Ed.)
2016, XXII, 380 p. 396 illus., 304 illus. in color.,
Hardcover
ISBN: 978-94-017-7510-6