We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800 Open access books available
116,000 International authors and editors
120M Downloads

154 Countries delivered to
TOP 1% Our authors are among the most cited scientists
12.2% Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
1. Introduction

Silicon industries have notched up notable achievements of computer-related technology over the past two decades, leading to rapid progression in information technology (IT). As a result of such a great improvement in IT applications, it is now not unusual to find mobile applications such as personal digital assistants, mobile phones with digital cameras, smartphones, and hand-held personal computers. These mobile applications currently require an array of single-functioned conventional memories as they are not sufficient individually in functionality, but must combine their separate functions.

For example, dynamic random access memory (DRAM) is capable of processing massive amounts of data speedily (e.g., main memory in personal computers and servers). DRAM is highly scalable (several gigabit are commonly accessible), but requires lots of power consumption even in stand-by mode (~$10^{-3}$ Ampere) because of the necessity of refreshing cycles in its operation. By contrast, static random access memory (SRAM) saves power because its stand-by current is a few micro-Ampere. The demerit of SRAM is not readily to make it high density. This is due to the fact that its unit memory element consists of four complementary metal-oxide-semiconductor (CMOS) transistors along with two conventional transistors. SRAM’s cost-benefit ratio is too high because the 6 components need much more area per unit bit memory. Data retention of both DRAM and SRAM is volatile in bit-storing nature when power goes off. In contrast to these two memories, flash memory is non-volatile. However, operation voltage during either write or erase on flash memory is too high to use the raw voltage-level of power input, $V_{cc}$ (the term of $V_{cc}$ comes from collector to collector voltage in a bipolar transistor). Thus during the write or erase operation, internal dummy operation (so called “charge pump”) are used to pump up the input power $V_{cc}$ to 5 times more than $V_{cc}$ level; this is crucial in flash memory devices due to imbalance of read and write energy. The reason why the memory needs to boost the write/erase voltage up to such a high level is that hot carriers, e.g., high energy electrons, are forced to be injected through tunnel oxide to a floating gate of the transistor structure. As a result, there are two kinds of performance restrictions for use of IT applications. Writing speed of flash memory is not fast enough of an order of several milliseconds. That

---

1This is not necessarily true because the stand-by current of SRAM begins to exceed DRAM’s in a deep sub micron scale due to involvement of high field junction.
makes the erasing speed of the device to be in the range of tenths of seconds. Another drawback of the device is endurance, which is defined as cycle times to write data in a memory cell. Generally, while write endurance of DRAM and SRAM is more than \(10^{15}\) cycle\(^2\) (\(10^{15}\) corresponds to equivalent 10 years for use), flash memory has approximately \(10^6\) cycles at most as writing endurance. In addition to flash memory, there is another non-volatile memory, so called electrically erasable and programmable read only memory (EEPROM). However, EEPROM has the same limitations in flash memory due to structural and operational similarity of the unit memory cell in flash memory.

To compensate for the aforementioned disadvantages of conventional memory devices, mobile applications in the IT world have adopted a combination of individual memories, which give several penalties such as a large volume of space to pack them all and complex time adjustments to synchronize them as well. As needs of IT technology are pushing forward to many functional requirements including much faster Internet access and far more image processing, this combining approach has a limitation to apply them to those for mobile uses. Therefore, it is strongly desirable to develop an ultimate memory solution as a single memory platform, possessing positive features of the individual memories but excluding their disadvantages. The feature of the ideal memory should have fast operation for speedy communication, high density for massive data-processing, non-volatility and low power consumption for portable applications.

Among many candidates of ideal memory devices, a memory device to use ferroelectric properties, so called ferroelectric random access memory (FRAM), was proposed and experimentally explored in terms of 512-bit memory density (Evans & Womack, 1988). This is because its functional feature is similar to that of an ideal memory. This is thanks to the bi-stable state of ferroelectrics at near ambient temperature. There are several important characteristics worth mentioning. First, since core circuitry for the memory does not require stand-by power during quiescent state and the information remains unchanged even with no power supplied, it is thus non-volatile. Second, configuration of unit memory element is similar to that either of flash or of DRAM, allowing it to potentially become cost-effective high density memory. Third, speed of ferroelectric memory could be very close to those of the conventional volatile memories such as DRAM and SRAM. This is, in practice, because repeating the polarization reversal-read and write operation, does not need boost up base voltage unlike flash memory, stemming from balance of read and write energy of the same order of magnitude (Kryder & Kim, 2009). A good example of this is that, according to literature published recently, one of the FRAMs as a non-volatile memory has attractive memory performance such as fast access time of 1.6 GB/sec, negligible stand-by current of less than 10 micro-Ampere, and low voltage operation of less than 2.0 V even in read and write action without erase operation (Shiga et al., 2009; Jung et al., 2008). Since then, there have been tremendous improvements in FRAM developments, migrating from sub-micron to nano scale in technology node. As such, this chapter is categorized into two: First demonstrates

\(^2\)Provided a clock frequency \(f\) of a microprocessor in an embedded system is 20 MHz (the fastest one in 2006 is about 5 MHz), reference counts necessary for cycle times per a year, is less than 1E13 in spite of considering 2% of strong data-locality in data memory. Note, the reference counts per a second is proportional to the products of \(frL\), where \(f\) is a clock frequency, \(r\) is ratio of number of cycles in read and write operation to unit cycle and \(L\) is a constant of representing data-access locality. We will discuss this more in section 3.2.
reminiscent of how memory technologies penetrate technological barriers to match the Moore’s law. Also, authors are here trying to give an insight of how silicon technology can evolve even in 20-30 nm technology node. Second is devoted to ferroelectric memories as an ultimate memory solution in many aspects such as lifetime data-retention and endurance; size effects; integration technologies; and feasibility as a fusion memory element.

2. Future memory technology

2.1 Evolution of silicon technology

2.1.1 Moore’s law

It is generally accepted that semiconductor industry will continue to expand rapidly due to steady growth of the mobile, digital consumer and entertainment markets. In addition to these, many more growth engines will appear, encompassing the automotive, information-technology, biotechnology, health, robotic and aerospace industries. The advances in silicon technology that has been the backbone of tremendous previous growth, were foreseen in 1965 when Gordon Moore published his famous prediction about the constant growth rate of chip complexity (Moore, 1965). And, in fact, it has repeatedly been shown that the number of transistors integrated into silicon chips has indeed doubled every 18 months. Increases in packing density, according to the Moore’s law, are driven by two factors: reductions in production costs and increases in chip performance. Another prominent example of the unstoppable pace of technology advancement, has been predicted (Hwang, 2006). Figure 1 shows Moore’s doubling phenomena of the number of components—the number of gate in case of CPU (central processing unit) and density in memory device.

Fig. 1. Moore’s doubling phenomena of the number of chips (Moore, 2006).

Moore’s law was predicted to stagnate to the end of the 20th century, but new sources of momentum are able to maintain or accelerate a growth trend. SoC (System-on-a-chip) integration has the potential to continue IC (integrated-circuit) cost reduction and to perpetrate growth of personal Internet products.
Despite these bright prospects, there is growing concern about whether semiconductor technology can continue to keep pace with demand when silicon technology enters the “deep nano-scale” dimension. This is because there are limits to transistor scaling, and narrowing margins in manufacturing due to ever-increasing fabrication costs tied to technical complexities (Kim & Jeong, 2005; Kim & Choi, 2006). The manufacturing cost grows because the engineering becomes more complex as transistors shrink in size. The scale is staggering, but the current generation of memory chips is 30 nm node across. This does mean that innovation is more process driven, and may require suppliers to think about what customers need and value, rather than simply pushing for ever greater density of transistors. Though most experts believe that silicon technology will maintain its leadership down to 20 nm, beyond this node a number of fundamental and application-specific obstacles will prevent further shrinkage. A common example is the inevitable occurrence of variations due to rough line edges and surfaces when pattern sizes approach atomic scales (Hwang, 2006). It is therefore the primary aim of this section to present various possible paths to overcome these obstacles and eventually to maintain the technology-scaling trend beyond 20 nm node.

As will be shown, these solutions include not only 3-D (three-dimensional) technologies but also non-silicon technologies on a molecular scale. In addition, new applications, and new growth engines for the semiconductor industry will be provided from a combination of separate technologies such as silicon-based IT with new materials (Whang et al., 2003; Wada, 2002). Therefore, this section is structured as follows: First, a review of the evolution of key silicon technologies is given. Next, we discuss scaling limits in each technology node and demonstrate practical and plausible solutions to penetrate these scaling barriers. Both DRAM and NAND flash memory are dealt with in discussion. And then, authors present prospects for the future silicon industry covering fusion technologies.

2.1.2 Evolution of silicon technology

**DRAM:** Since the first 1,024-bit DRAM was demonstrated by Intel™ in the early 1970’s, the highest available density of DRAM has doubled every 18 months. Now DRAM technology has reached 30 nm in process technology and 4 Gb in density, which will be deployed soon in the marketplace. Further, 20-nm DRAM is being developed at R&D centers around the world. DRAM technology has evolved toward meeting a need for ever-increasing demand both of data retention and of performance improvements. Increases in data retention and of performance improvements. Increases in data retention impose great challenges on DRAM technology by requiring not just a sufficient amount of capacitance in a memory cell but an extremely low level of leakage current from storage junction. As device shrinks, it has been being one of the most challenging to achieve an adequate amount of cell capacitance in DRAM. It is widely agreed that the value of cell capacitance is more than 20 fF, regardless of technology-node migration. This is because sensing signal developed from memory cells, is vulnerable to become interfered by unwanted noise factors according to its operational nature. Sensing signal, $V_s$, can be expressed by equation (1), where $C_s$ is cell capacitance at storage node; $C_{BL}$ is parasitic bit-line capacitance; $AIVC$ is cell array internal voltage; and the last term $V_{UN}$ is undesirable noise. In equation (2), the first term $I_{LEAKFBD}$ in the parenthesis is a term of charge loss due to junction leakage current, $I_{GIDL}$ gate-induced-drain-leakage (GIDL) current, $I_{NG}$ non-generic leakage current, $I_{UN}$ arising from integration imperfections (e.g., dielectric leakage current and cell-to-cell leakage current) as indicated in equation (3).
Future Memory Technology and Ferroelectric Memory as an Ultimate Memory Solution

\[ V_S = \frac{AVC}{2} \cdot \frac{C_S}{C_{BL} + C_S} - V_{UN} \]  

\[ V_{UN} = \frac{1}{C_{BL} + C_S}(I_{LEAK}t_{REF} + C_{BL}V_N + Q_I) \]  

\[ I_{LEAK} = I_{JUNC} + I_{GIDL} + I_{NG} \]

All of those loss factors are constituents of data-retention time, so called refresh time, \( t_{REF} \). \( V_N \) is noise voltage due both to noise coupling and to mis-matches of threshold voltage and conductance of sense amplifiers. Another source of charge loss, \( Q_I \) in Eq. (2) has to be considered when DRAM is exposed to irradiations such as \( \alpha \)-particle and cosmic rays. These undesirable components are very difficult to attenuate and become dominant as device dimensions are smaller. To maintain almost non-scalable requirement of cell capacitance of more than 20 fF/cell, dielectric material of cell capacitors have continuously evolved into high-\( \kappa \) dielectric materials and at the same time their structures have been pursued actively for novel ones (Lee et al., 2003a; Kim et al., 2004a). This is due to the fact that cell capacitor area decreases by a factor of \( 1/k \sim 1/k^2 \) as technology scales, where \( k \) denotes a scaling factor, where \( k > 1 \) (See Denard et al., 1974). In general, when designing device to a smaller dimension, the device is scaled by a transformation in three variables: dimension, voltage, and doping concentration. Firstly, all the linear dimensions are reduced by a unit-less scaling constant \( k \), e.g., \( t_{OX}' = t_{OX}/k \). Such reduction includes not only vertical dimensions such as thickness of gate oxide and junction depth, but also horizontal dimensions, for example, channel length \( L \) and width \( W \). Secondly, voltage applied to the new device has to be reduced by the same factor, e.g., \( V_{CC}' = V_{CC}/k \). Lastly, doping concentration, \( N_A \) is to be increased, e.g., \( N_A' = kN_A \).

In practice, DRAM’s capacitor has begun with a stacked 2-D (two-dimensional) structure, integrated under bit-line in process architecture until the mid 1990’s. Since then, DRAM has changed in structure to have an integration scheme of cell-capacitors placed over bit-line (COB) though there was an attempt to use trench-type capacitors, which are buried deeply in silicon substrate. In the 1990’s, dielectric material of the cell capacitors has adopted silicon-based dielectrics, SiO\(_2\)/Si\(_3\)N\(_4\), whose dielectric constant lies in between 3.9 and 7.0. With these relatively low-\( \kappa \) dielectrics, a cell capacitor has headed for expanding its area as much as possible. Thus, its structure has been transformed in substantially complex ways, from a simple stack to a hemi-spherical-silicon-grain (HSG) stack, to a HSG cylinder until the late 1990’s. The advent of high-\( \kappa \) dielectrics since the beginning of 21\(^{st}\) century has brought a new era of building the cell capacitors. Table 1 compares fundamental material properties of high-\( \kappa \) candidates with those of conventional low-\( \kappa \) dielectrics. These high-\( \kappa \) dielectrics have allowed us to form the cell capacitors into simpler one-cylinder-stack (OCS) than those in low-\( \kappa \) dielectrics due to relatively higher dielectric constant. Provided high-\( \kappa \) dielectric material utilizes, increase in cell capacitance will be achieved simply by increase in height of a cylinder. Such an increase in height gives rise to skyrocketing of aspect ratio of

\(^4\)A scaling factor of capacitance \( C = \varepsilon A/t \) is supposed to be \( 1/k \) in a 2-D stack structure, but since capacitor thickness \( t \) is not a contraint factor any more in a 3-D structure, capacitance can be written in \( 1/k^2 \).

\(^5\)So called capacitor-under-bit-line (CUB) in integration architecture.
cell capacitors when technology scales, together with dramatic decrease in footprint. In typical, an aspect ratio of cell capacitors ranges from 6 to 9 until 100 nm technology node. A higher aspect ratio has brought another obstacle in building cell capacitors robust: mechanical instability of OCS structures. As a result, many smart engineers in silicon industries has introduced a novel capacitor structure, supporter-added OCS such as mesh type cell capacitors, which can increase the cell capacitor height with desired mechanical stability (Kim et al., 2004a). Taking into account the recent advances of the cell-capacitor technology, the aspect ratio reaches 35 to 45, which is far beyond those of the world tallest skyscrapers, ranging from 8.6 to 10.0.

<table>
<thead>
<tr>
<th>Materials</th>
<th>Dielectric constant (κ)</th>
<th>Band gap $E_G$ (eV)</th>
<th>Crystal Structure(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SiO$_2$</td>
<td>3.9</td>
<td>8.9</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Si$_3$N$_4$</td>
<td>7.0</td>
<td>5.1</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Al$_2$O$_3$</td>
<td>9.0</td>
<td>8.7</td>
<td>Amorphous</td>
</tr>
<tr>
<td>Y$_2$O$_3$</td>
<td>15</td>
<td>5.6</td>
<td>Cubic</td>
</tr>
<tr>
<td>La$_2$O$_3$</td>
<td>30</td>
<td>4.3</td>
<td>Hexagonal, cubic</td>
</tr>
<tr>
<td>Ta$_2$O$_5$</td>
<td>26</td>
<td>4.5</td>
<td>Orthorhombic</td>
</tr>
<tr>
<td>TiO$_2$</td>
<td>80</td>
<td>3.5</td>
<td>Tetragonal, rutile, anatase</td>
</tr>
<tr>
<td>HfO$_2$</td>
<td>25</td>
<td>5.7</td>
<td>Monoclinic, tetragonal, cubic</td>
</tr>
<tr>
<td>ZrO$_2$</td>
<td>25</td>
<td>7.8</td>
<td>Monoclinic, tetragonal, cubic</td>
</tr>
</tbody>
</table>

Table 1. Comparison of material properties of high-$\kappa$ dielectric candidates with those of conventional low-$\kappa$ dielectrics (Wilk et al., 2001).

In the meanwhile, charging and discharging properties of cell capacitors depend strongly on performance of cell array transistors (CATs). On-current of the CAT plays a critical role in its charging behaviors while off-leakage current of the CAT is a decisive factor to determine their discharging characteristics. On the one hand, on-current ($I_{on}$) needs to be at least greater than several $10^{-6}$ Ampere to achieve reasonable read and write speed. On the other, off-leakage current ($I_{off}$) has to satisfy a level of $10^{-16}$ Ampere to minimize charge loss just after charging up the cell capacitors to ensure adequate sensing-signal margin as indicated in Eq. (2). Despite continuation of technology migration, the ratio of $I_{on}$/$I_{off}$ has remained constant to $10^{10}$ approximately. CAT’s technology has evolved to meet this requirement.

$$I_{ON} = \mu_{eff} \cdot C_{OX} \cdot \frac{W}{L_{eff}} \cdot \frac{\left( V_{GS} - V_{th} \right)^2}{2}$$

where $\mu_{eff}$ is effective mobility for electrons, $C_{OX}$ is capacitance of gate oxide, $W$ is width of transistor’s active dimension, and $L_{eff}$ is effective channel length.

At first, from the structural point of view, 2-D planar-type CAT (PCAT) has been moved to 3-D CAT. The reason why 3-D CAT has been adopted is to relieve data retention time. In 100-nm technology node, $I_{off}$ of the PCATs does not ensure a specific level of off-leakage current requirement (less than $10^{-15}$ A) due to high-field junction. The high electric field is caused by high-doping concentration near the channel region to block short-channel-effect (SCE). Under such a SCE circumstance, a transistor does not, in general, work any longer, by a way of punch-through between source and drain when its channel length becomes shorter. As denoted in Eq. (2) and (3), off-leakage currents $I_{LEAK}$ are closely related to data retention.
Future Memory Technology and Ferroelectric Memory as an Ultimate Memory Solution

123

time. Generically this $I_{\text{LEAK}}$ arises from sub-threshold current and gate-induced drain leakage (GIDL) current of cell array transistors along with junction leakage current from storage node. As $L_{\text{eff}}$ is scaled down, the increased doping concentration against the SCE strengthens electric field across storage node junction. This increase in junction-leakage current results in degrading the data retention time (Kim et al., 1998). The degradation of data retention time becomes significant below 100 nm node due to rapid increase in junction electric field again (Kim & Jeong, 2005). This issue since the mid 2000’s has been overcome by introducing 3-D cell transistors, where the junction electric-field can be greatly reduced due to lightly doped channel. One example of these newly developed structures is RCAT (Recess Channel Array Transistor) structure whose channel detours around a part of silicon substrate so that the elongated channel can be embodied in the array transistor (Kim et al., 2003). Also, the RCAT structure gives us another benefit, which lessens threshold voltage ($V_{\text{th}}$) due to lower doping concentration. Thereby, not only does DRAM’s core circuitry operate at lower voltage but also CAT’s on-current increases, as denoted in Eq. (4). Note that, according to the Moore’s law, $V_{\text{cc}}$ must be scaled down for power save. This trend has continued to come to 60 nm technology node. However, beyond 60 nm of technology node, on-current requirement has not been satisfied with such a RCAT approach alone. Thus, further innovations since 50 nm node have been pursued in a way of a negative word-line (NWL) scheme in DRAM core circuitry. The NWL scheme compared with a conventional ground-word-line (GWL) scheme, allows us $V_{\text{th}}$ reduction further, which means more on-current. However, another adverse effect on the CAT can occur as a result of the NWL. Since CAT’s gate potential goes more negative during holding data stored at the storage junction, from which GIDL current increases as a function of gate-storage voltage, level of which is as high as that of gate potential compared with the conventional GWL. Many device engineers have given much effort to tackle this problem and finally have figured it out by technological implementation, for instance, mitigation of electric field exerted locally in the region overlapped between source/drain and gate in the RCAT. In pursuit of purpose, gate oxide needs to be different in thickness.

Provided that the oxide thickness in the overlapped region is thicker than that in the channel, unwanted GIDL current will decrease in proportion to electric field of the overlapped zone in the storage-node to gate (Lee et al., 2008; Jung et al., 2009). According to our calculation, one can extend this NWL-based RCATs down to 40 nm node with minor modifications (Jung et al., 2009). In 30 nm technology node, it becomes extremely difficult to achieve the successful $I_{\text{on}}/I_{\text{off}}$ ratio. A report has shown that a body-tied FinFET (fin field-effect-transistor) as a cell array transistor seems to be very promising due to its superb performances: excellent immunity against the SCE; high trans-conductance; and small sub-threshold leakage (Lee et al., 2004). For example, it allows us to have not only lower $V_{\text{th}}$ but lower sub-threshold swing due to a fin-gate structure, providing more width for on-current and wrapping the gate for $V_{\text{th}}$ and sub-threshold swing down. It is believed that the body-tied FinFET leads DRAM technology to be extendable down to 30 nm node. In off-leakage current, CAT’s gate material has been being transformed to metal gate of higher work function (4.2~4.9 eV) instead of n+ poly-silicon gate. The lower $V_{\text{th}}$ coming from higher work function provides us with lower channel doping. This leads to lower junction electric field and results in lower off-leakage.

Since a level of dc (direct current) bias at unselected word-lines is negative, sub-threshold leakage current of a cell transistor becomes extremely low because its channel has never chance to be on-set of inversion, leading to keeping a reasonable level of off-leakage current despite low $V_{\text{th}}$. 

Since a level of dc (direct current) bias at unselected word-lines is negative, sub-threshold leakage current of a cell transistor becomes extremely low because its channel has never chance to be on-set of inversion, leading to keeping a reasonable level of off-leakage current despite low $V_{\text{th}}$. 

www.intechopen.com
current. Figure 2 shows how DRAM’s CAT structure has evolved during the past decade. Beyond 30 nm of technology node, a novel structure must be suggested for continuing the successful \( \frac{I_{on}}{I_{off}} \) ratio. Among many structures, a vertical channel CAT (VCAT) is one of the good candidates (Yoon et al., 2006). This is because it can plausibly permit us to access an ideal transistor. A VCAT has a surrounding gate buried in silicon substrate (Kim, 2010). Bit-line connected to its data node runs buried under silicon substrate, too. With such a burying architecture, VCAT-base DRAM is expected to provide minimum size of lateral dimension per unit memory element as indicated in the inset of Fig. 2.

Fig. 2. Maximum electric field, \( E_{\text{MAX}} \) as a function of channel-doping concentration in various CAT’s structures. As CATs evolves, the doping concentration decreases \( E_{\text{MAX}} \), denoted in red in \( E \)-field strength of simulation structures as shown in the inset. An example of DRAM architecture based on VCAT is also shown in the inset.

\textbf{NAND flash memory:} NAND flash memory has the smallest cell size among silicon-memory devices commercially available due to its simple one transistor configuration per one bit and a serial connection of multiple cells in a string. Because of this, NAND flash has carved out a huge market for itself, as was expected since it first appeared in the mid 1980’s. The need for NAND flash memory will continue to surge due to the recent resurgence of demand for mobile products such as smart phones and smart pads. With the rise of the mobile era, NAND flash has pushed toward ever-higher density, along with improving programming throughput. As a consequence, the memory has evolved toward an ever-smaller cell size in two ways: by increasing string size and by developing two bits per cell, while at the same time, increasing page depth. Now, current NAND flash memory reaches 30 nm node in process technology and 32 Gb in density, mass production of which has blossomed since the late 2000’s. In addition, NAND technology beyond 30 nm is now under development at R&D centers across the world. Alongside the recent development of two-bit-per-cell technology, introduction to multi-bit cells should greatly accelerate this trend.
There are several prerequisite requirements to meet in terms of cell operation. Its cells must satisfy write and read constraints. First is programming disturbance. To program a cell, it is necessary to apply a certain amount of electric field across between floating gate and channel of the cell so that a sufficient amount of Fowler-Nordheim (FN) tunneling electrons can be injected into the floating gate.

\[
\frac{1}{t_{OX}} \cdot \gamma \cdot V_{PGM} \geq 10 \text{ MeV}, \quad \gamma = \frac{C_{CS}}{C_{TUNNEL} + C_{CS}}
\]

where \(t_{OX}\) is thickness of tunnel oxide; \(\gamma\) is a coupling ratio; \(V_{PGM}\) is programming voltage; \(C_{CS}\) is capacitance between control gate and storage media; and \(C_{TUNNEL}\) is capacitance of tunnel oxide. Figure 3 illustrates (a) a schematic diagram of NAND cell arrays and (b) their programming conditions. During the programming, there are two types of unselected cells that tolerate unwanted programming: One type is cells connected to the same bit-line of the selected cell. And the other is cells connected to the same word-line. The former suffers so called \(V_{PASS}\)-stress cells while the latter endures so called \(V_{PGM}\)-stress cells as follows:

\[
V_{PASS\ stress} = \frac{1}{t_{OX}} \cdot \gamma \cdot V_{PASS} \leq 1 \text{ few MeV},
\]

\[
V_{PGM\ stress} = \left(1 + \frac{C_{D}}{\gamma \cdot C_{TUNNEL}}\right)^{-1} \leq 1 \text{ few MeV},
\]

where \(V_{PASS}\) stress is voltage applied to the unselected cells which share the same bit-line of the programming cell; \(V_{PGM}\) stress is voltage applied to the unselected cells which share the same word-line; and \(C_{D}\) is depletion capacitance of silicon substrate (See Fig. 3b). The \(V_{PASS}\)-stress and the \(V_{PGM}\)-stress are, in general, so small that neither electron injection into the unselected cells nor ejection from those is allowed in programming, respectively. Thus, \(V_{PASS}\) window is determined by allowable both \(V_{PASS}\)-stress and \(V_{PGM}\)-stress. However, the \(V_{PASS}\) window will be narrow when scaling down because of increase in depletion capacitance \((C_{D})\) as denoted in Eq. (7). Thus, as technology scales, adequate \(V_{PASS}\) window has to be satisfied. Next, in read operation, read voltage of a floating gate has to be higher than the highest threshold voltage of a cell string in order to pass read current through the string on which 32 cells are connected in series (in case of Fig. 3a). In similar to programming disturbance, read disturbance might occur in the unselected cells on the same string, and thus together with appropriate pass voltage, it is believed not only to choose tunnel oxide but to regulate its thickness in integration as well.

\[
\gamma \cdot V_{READ} \geq V_{TH}
\]

As a rule of thumb, an adequate value of the coupling ratio in read lies in the range of 0.5 ~ 0.6, and reasonable thickness of the tunnel oxide is about 80 Å. Last but not least, one of the fundamental limitations of the NAND flash stems from the number of stored charge because the available number of storage electrons decreases rapidly with technology scaling. Provided that the voltage difference between the nearest states in a 2-level cell is less than 1 V, threshold voltage shifts due to charge loss will be restricted to less than 0.5 V, which puts the limitation on charge loss tolerance as follows,

\[
\Delta Q \leq C_{CS} \cdot \Delta V_{TH} = \sim 0.1Q,
\]
In case of the floating gate, \( C_{CS} \) is \( C_{ONO} \) of capacitance of oxide-nitride-oxide. Therefore, at most 10% of charge loss is tolerable, which means that less than 10 electrons are only allowed to be lost over a 10 year period.

Fig. 3. (a) Schematic diagrams of memory cell arrays and (b) their programming conditions.

In technology evolution, flash memory since the late 1990’s has continued to migrate technology node to 70 nm until the beginning of 2000’s, based on a floating gate (FG) (Keeney, 2001; Yim et al., 2003). Due to an unprecedented growing pace of flash-memory demand for use in mobile applications, higher-grade memory in packing density has been driven by burgeoning of multi bits per a cell since the mid 2000’s (Park et al., 2004; Byeon et al., 2005).

Now that multi-level cell (MLC) technology means a wide range of \( V_{PASS} \) window in Eq. (5) to (7), it is essential to increase the coupling ratio as shown in Eq. (5). In addition, to overcome stringent barrier of charge-loss tolerance is simply to increase storage charge. This can also be achieved by increasing \( C_{CS} \) as indicated in Eq. (9). However, thickness scaling for high \( C_{CS} \) may not be easy in case of \( C_{ONO} \) (e.g., a floating gate, here). This is because 60-nm flash memory has already reached 13 nm of equivalent oxide thickness (EOT\(^7\)), which is believed to be a critical limit in thickness, for allowable charge loss–ONO thickness \( \sim 14.5 \) nm (Park et al., 2004). It has been reported that \( C_{CS} \) can be increased by replacing top-blocking oxide into new high-\( \kappa \) dielectric of \( Al_2O_3 \) instead. This provides us with strengthening electric field across the tunnel oxide and at the same time with lessening electric field across the blocking oxide in program and erase. Also, fast erase can be possible even at thicker tunnel oxide of over 30 Å where direct-tunneling hole current could be reduced significantly and thus such a structure gives robust data-retention characteristics (Lee et al., 2005).

Meanwhile, from the scaling point of view, flash memories have faced a serious problem since 50 nm of technology node: Cell-to-cell separation becomes so close each other that influence between adjacent cells cannot be ruled out. This is often posed not only by physical aspects of

\(^7\)EOT indicates how thick a silicon oxide film needs to have the same effect as a different dielectric being used.
cell structures but also by certain aspects of its performance. To circumvent cell-to-cell interference, width of a floating gate tends to be more aggressively squeezed than space between floating gates (See Fig. 3b). This seems to result in a high aspect ratio of a gate stack. Such a high aspect ratio can provoke fabrication difficulty of memory cells due to its mechanical instability. And stored charge (e.g., electron) in a floating gate can redistribute easily in operational conditions, leading to vulnerability of poor data retention. Since the interference originates from another type of coupling between floating gates (FGs), it is desirable to find innovative structures, where charge storage media do not have a form of continuum of charge like the floating gate style but have a discrete sort such as charge traps (CTs) in a nitride layer. The typical examples are non-volatile memories with non-floating gate, for example, SONOS (silicon-oxide-nitride-oxide-silicon) (Mori et al., 1991), SANOS (silicon-alumina-nitride-oxide-silicon) (Lee et al., 2005), TANOS (TaN-alumina-nitride-oxide-silicon) (Shin et al., 2006) or nano-crystal dots (Tiwari et al., 1995; Nakajima et al., 1998). Recently, 32 Gb flash memory has been reported, in particular, in 40 nm of technology node (Park et al., 2006). They have pioneered a novel structure with a high-$\kappa$ dielectric of $\text{Al}_2\text{O}_3$ as the top oxide and TaN as a top electrode. With this approach, they can achieve several essential properties for NAND flash memory: reasonable programming/erasing characteristics, an adequate $V_{\text{PASS}}$ window for multi-bit operation and robust reliability. It is noteworthy that a TANOS structure has much better mechanical stability than that of an FG-type cell because of the far lower stack in height. Interference among TANOS cells hardly occurs due to nature of the charge trap mechanism—$\text{SiN}$ (silicon nitride) traps act as point charges. This is the biggest advantage in CT-NAND flash memory. To scale NAND flash further down, we may need another cell technology. A FinFET could be a very promising candidate because it can increase storage electrons effectively by a way of expanding channel width of cell transistors, similar to 3-D CATs in DRAM. In this pursuit, a research group has successfully developed flash memory with a TANOS structure based on a 3-D, body-tied FinFET (Lee et al., 2006), where they can obtain excellent performance of NAND-flash cells with robust reliability. If there are much higher $\kappa$ dielectrics than $\text{Al}_2\text{O}_3$, then we can further scale down the FinFET CT-NAND flash memory.

![Fig. 4. (a) A schematic diagram of 3-D, body-tied FinFET NAND cells and (b) comparisons of the 3-D cells with 2-D, planar cells in threshold-voltage shift as a function of programmed threshold voltage, measured after suffering 5k program/erase cycles and a bake at 200 °C for 2 hours (Lee et al., 2006).](www.intechopen.com)
Figure 4 represents (a) a schematic diagram of 3-D, body-tied FinFET NAND cells and (b) comparisons of the 3-D cells with 2-D ones in threshold-voltage shift as a function of programmed threshold voltage, measured after suffering 5k program/erase cycles and a bake at 200 °C for 2 hours. The threshold-voltage delay has been improved to 0.32 V in 3-D NAND cells, compared with 2-D NAND ones.

2.2 Prospects of silicon technology

As well aware that the era of 2-D, planar-based shrink technology is coming to an end, semiconductor institutes have seen enormous hurdles to overcome in order to keep up with the Moore’s doubling pace and thus to meet the requirements of highly demanding applications in mobile gadgetry. They have attempted to tackle those barriers by smart and versatile approaches of 3-D technology in integration hierarchy. One strand of the responses is to modify structures of elementary constituents such as DRAM’s CATs, its storage capacitors and storage transistors of flash memory to 3-D ones from the 2-D. A second thread revisits these modifications to a higher level of integration: memory stacking. And another move is to upgrade this into a system in a way of fusing of each device in functionality by utilizing smart CMOS technology, e.g., through-silicon-via (TSV).

2.2.1 Elementary level of 3-D approach

When working with silicon devices, a transistor’s key parameters must take into account: on-current; off-leakage current; the number of electrons contained in each transistor; or the number of transistors integrated. All of these factors are very important, but not equally important in functional features of silicon devices. For instance, for memory devices, off-leakage current is regarded as a more important factor and thus memory technologies tend to be developed with a greater emphasis on reducing off-leakage current. For logic, transistor delay is the single most important parameter, not just to indicate chip performance but to measure a level of excellence in device technology as well. This transistor delay is related closely to transistor’s on-current state. And with 2-D planar technology in logic, one can continue to reduce transistor’s channel length down to 40 nm. However, at less than 30 nm, the transistor begins to deviate in spite of a much relaxed off-current requirement. This is because of non-scalable physical parameters such as mobility, sub-threshold swing and parasitic resistance. To resolve these critical issues, two attempts have been examined. One is to enhance carrier mobility by using mobility-enhancement techniques such as strained silicon (Daembkes et al., 1986), SiGe/Ge channel (Ghani et al., 2003), or an ultra thin body of silicon (Hisamoto et al., 1989), where carrier scattering is suppressed effectively. Another approach is to reduce channel resistance by widening transistor’s width. In this case, it appears very promising to use different channel structures such as tri-gate (Chau et al., 2002) or multi channel (Lee et al., 2003b). We have witnessed that, with 3-D FinFETs in memory devices, this attempt is very efficient for extending incumbent shrink technology down to 30 nm of technology node. As silicon technology scales down further, the two will eventually be merged into one single solution for an optimum level of gate control. With this type of structure, one will arrive at nearly ideal transistor performance such as being virtually free from the SCE, sufficient on-current and suppressed off-leakage current. Figure 5 shows (a) evolution trends of logic transistors in terms of EOT: A sharp decrease in EOT trend appears due to lack of gate controllability in 2-D planar structures despite high-κ dielectrics. By contrast, those in 3-D, multi-gate structures
are expected to have the same trend of EOT as those with conventional SiON dielectrics. This suggests that 3-D structures seem to become essential even with high-κ materials. It is thus believed that developing a 3-D transistor with either a multi-gate or an gate-all-around structure (Colinge et al., 1990) is quite feasible if one can extend 2-D planar technology to 3-D. This is because the channel length is no longer restricted by lateral dimension. Figure 5 also shows (b) a cross-sectional TEM (transmission-electron-micrograph) image of one of the 3-D, multi-gate transistors and (c) its Ion-Ioff characteristics are compared with those of 2-D planar structures.

Fig. 5. (a) Equivalent-oxide-thickness (EOT) scaling trends (Kim, 2010) are shown in reciprocal scale. Due to the difficulty in controlling the SCE, a sharp decrease in EOT trend is inevitable for the coming nodes. However, the historical trend can be reverted back in the case of 3-D, multi-gate transistors. (b) A cross-sectional TEM image of a 3-D, multi-gate and (c) its Ion-Ioff characteristics are compared with those of the planar (Lee et al., 2003b).

2.2.2 3-D stacking of memory cells
New silicon technology based on 3-D integration has drawn much attention because it seems to be regarded as one of the practical solutions. Though the concept of 3-D integration was first proposed in the early 1980’s (Kawamura et al., 1983; Akasaka & Nishimura, 1986), it has never been thoroughly investigated or verified until now, as neither silicon devices approached their limits at those times nor high-quality silicon crystal was ready for fabrication. Recent advances both in selective epitaxial silicon growth at low temperature (Neudeck et al., 2000) and in high quality layer-transferring technology with high-precision processing (Kim et al., 2004b), can bring major new momentum to the silicon industry via 3-D integration technology. The simplicity of memory architecture consisting of memory array, control logic and periphery logic, makes it relatively easy to stack one-memory cell array over another. This will ultimately lead to multiple stack designs of many different memories. Recently, one of the memory manufacturers has started to implement 3-D integration technology with SRAM to reduce large cell-size (Jung et al., 2004). Figure 6
shows (a) a cross-sectional TEM image of 3-D stacking SRAM (Left) and its schematic diagram (Right) (Jung et al., 2004): Since transistors stacked onto a given area do not need to isolate p-well to n-well, SRAM-cell size of 84 $F^2$ is being reduced to the extremely small cell size of 25 $F^2$. Encouraged by this successful approach, stacked flash memory has also been pursued. Figure 6 also represents (b) 3-D stacking NAND flash memory (Jung et al., 2006): This suggests great potential of 3-D memory stacking for large-scale use with 3-D flash-cell technology, which will spur further growth in high-density applications. Beyond 20 nm node, we believe that the most plausible way to increase density is to stack the cells vertically. Figure 6 displays (c) a 3-D schematic view of vertical NAND flash memory (Katamura et al., 2009), where SG is selecting gate, CG is control gate and PC is pipe connection. The stacking of memory cells via 3-D technology looms on the horizon, in particular, for NAND flash memory.

![Fig. 6. (a) A cross-sectional TEM image of 3-D stacking SRAM (Left) and its schematic diagram (Right) (Jung et al., 2004). (b) 3-D stacking NAND flash memory (Jung et al., 2006). (c) A 3-D schematic view of vertical NAND flash memory (Katsumata et al., 2009), where SG is selecting gate; CG is control gate; and PC is pipe connection. (d) A cross-sectional SEM image of memory array after the removal of the sacrificial film (See Katsumata et al., 2009)](image-url)

It is also believed that logic technology will shift to 3-D integration after a successful jumpstart in silicon business. The nature of a logic device, where transistors and interconnections are integrated as key elements, is not much different from those of stacked memory cells. It may be very advantageous to introduce 3-D integration technology to a logic area. Note that implementation of interconnection processes seems to be more efficient in vertical scale. For example, a dual or quad-core CPU can be realized with only a half or quarter of the chip size, which will result in significantly greater cost-effectiveness. Another promising use would be to improve logic performance by cutting down on the length of metallization. Decrease in interconnection length means a huge amount of reduction in parasitic $RC$ components, i.e., a high speed and power saving. In addition, 3-D technology will make it easy to combine a memory device and a logic device onto one single chip.
Future Memory Technology and Ferroelectric Memory as an Ultimate Memory Solution

through hierarchical stacking. Since most parts of SoCs (system-on-chips) in the future will be allocated to memory, this combining trend will be accelerated. The next step will be to stack multi-functional electronics such as RF (radio frequency) modules, CISs (CMOS image sensors) and bio-sensors over the logic and memory layers.

2.2.3 Chip level of 3-D integration

The early version of 3-D integration in chip level has been commercialized already in a multi-chip package (MCP), where each functional chip (not device) is stacked over one another and each chip is connected by wire bonding or through the ‘through-via hole’ bonding method within a single package. Figure 7 exhibits (a) a bird’s eyes view of multi-chip-package (MCP) by wire bonding; (b) wafer-level stack package with through-via-hole; (c) a photograph of 3-D integrated circuit; and (d) a schematic drawing of a 3-D device for use in medical applications. The advantages of the MCP are a small footprint and better performance compared to a discrete chip solution. It is expected that the MCP approach will continue to evolve. However, the fundamental limitation of MCP will be lack of cost-effectiveness due to a number of redundancy/repair requirements. In this respect, ‘through-silicon-via’ (TSV) technology is able to overcome MCP limitations through an easy implementation of redundancies and repairs. Many groups have reported TSV-based integrated circuit (TSV IC), where a single integrated circuit is built by stacking silicon wafers or dies and interconnecting them vertically so that they can function as one single device (Topol et al., 2006; Arkalgud, 2009; Chen et al., 2009). In doing so, key technologies include TSV formation, wafer-thinning capability, thin wafer handling, wafers’ backside processes, and 3D-stacking processes (e.g., die-to-die, die-to-wafer and wafer-to-wafer). In detail, there are many challenging processes such as etching profiles of TSV sidewall, poor isolation liners and barrier-deposition profiles. All of these are likely to provoke TSV’s reliability concerns due to lack of protection from metal (e.g., Cu) contamination. A report of silicon-based TSV interposers (Rao et al., 2009) may have advantages over traditional PCB or ceramic substrate in that it has a shorter signal routing. This results from vertical interconnect and improved reliability due to similarity to silicon-based devices in thermal expansion and extreme miniaturization in volume. TSV-IC technologies together with the 3-D interposers will accelerate an adoption of 3-D system-in-package (SiP) with heterogeneous integration (See Fig. 7d). And this might be a next momentum for genuine 3D IC devices in the future because of tremendous benefits in footprint, performance, functionality, data bandwidth, and power. Besides, as the use of 3-D silicon technology has great potential to migrate today’s IT devices into a wide diversification of multi-functional gadgetry, it can also stimulate a trend that merges one technology with another, ranging from new materials through new devices to new concepts. In this regard, new materials may cover the followings: carbon nano-tube (CNT) (Iijima, 1991), nano-wire (NW) (Yanson et al., 1998), conducting polymer (Sirringhaus et al. 1998), and molecules (Collier et al., 1999). New devices could also be comprised of many active elements, such as tunneling transistors (Auer et al., 2001), spin transistors (Supriyo Datta & Biswajit Das, 1990), molecular transistors (Collier et al., 1999), single electron transistors (SETs) (Fulton & Dolan, 1987) and others. We may be able to extend this to new concepts, varying from nano-scale computing (DeHon, 2003) and FET decoding (Zhong et al., 2003) to lithography-free addressing (DeHon et al., 2003). To a certain extent, some of these will be readily integrated with 3-D silicon technologies. This integration will further enrich 3-D silicon technologies to create a variety of new multi-functional electronics, which will provide further substantive boosts to silicon industry, allowing us to make a projection of a nano-silicon era into practical realities tomorrow.
These realities will be manifested in highly desirable applications of combining of information technology (IT), bio-technology (BT), and nano-technology (NT), to become so called fusion technology (FT). Given that key obstacles to realize this are tackled by bridging the gap between previously incompatible platforms in silicon-based CMOS technology and new technological concepts, a vast number of new applications will unfold. One example may be many applications related to health sensor technology, in particular, the early recognition of cancer diseases and the screening of harmful and poisonous elements pervasive in the environment. Further, when a nano-scale bio-transistor is available, lab-on-a-chip (LoC) will become a single solution integrating all of its essential components, such as micro-array, fluidics, sensors, scanners and displayers. Then, by its very nature, one will have tons of benefits from a mass of disposable LoCs, which will stimulate the future silicon industry.

As a successful booster for the silicon industry, whatever will be, it should be a high volume product at a reasonable price. PCs are high volume products, and hand-held phones are too. In that sense, LoC is very promising because its potential market is the entire population.
2.3 Remarks
Not only do many challenges await silicon industries as technology enters the deep nano-dimension era but promising opportunities are also there. Equipped with new technologies such as 3-D scaling and a wealth of new materials, alongside fusing of related technologies, we will overcome many hurdles ahead and respond technological challenges we will stumble along the way. All plausible solutions described earlier tell us that planar-based technology will reach an impasse limit. 3-D technology begins to provide clear signs of serving as a foundation for a refuel of the silicon industry. The advantages of 3-D integration are numerous. They include: elimination of uncertainty in the electrical characteristics of deep nano-scale transistors; extendable use of silicon infrastructures, especially optical lithography tools; and formation of a baseline for multi-functional electronics and thus facilitation of implementing a hierarchical architecture, where each layer is dedicated to a specific functional purpose. Over the next decade, we will see great endeavors in numerous areas that will greatly stimulate the semiconductor business. Successful evolutions of device structures will continue and even accelerate at a greater pace in the not-too-distant future. In addition, device designs will converge onto a single mobile platform, covering many different capacities and services from telecommunication through broadcasting and a much higher degree of data processing. In line with this, silicon technology will still play a critical role in realizing functionally merged solutions. All of these will permit us to have invaluable clues not just on how to prepare future silicon technology but also on how to positively influence the entire silicon industry. This will allow us to attain an even more sophisticated fusing of technologies. As seen in the past, silicon technology will continue to provide our society with versatile solutions and as-yet unforeseen benefits in much more cost-effective ways.

3. Ferroelectric memory as an ultimate memory solution

3.1 Introduction
There has been great interest to understand ferroelectric properties from the point of view of both fundamental physics and the need of nano-scale engineering for memory devices. On the one hand, since electric hysteresis in Rochelle salt was in 1920 discovered by Valasek (Valasek, 1921), there have been tremendous efforts to look through ferroelectricity in a comprehensive way over the past many decades. As a consequence, the phenomenological theory of ferroelectricity has been presented by many researchers: Devonshire (Devonshire, 1949; Devonshire, 1951); Jona and Shirane (Jona and Shirane, 1962); Fatuzzo and Merz (Fatuzzo and Merz, 1959); Line and Glass (Line and Glass, 1979); and Haun (Haun, 1988). The series of their works have been successful to express the internal energy of a ferroelectric crystal system. This theory has also been examined experimentally in detail, and extended by Merz (Merz, 1953); by Drougard et. al. (Drougard et al., 1955); and by Triebwasser (Triebwasser, 1956). Especially, Devonshire’s phenomenological theory (Devonshire, 1949; Devonshire, 1951) gives the free energy of BaTiO$_3$, as a function of polarization and temperature. From this free energy we know what the possible state and meta-stable states of polarization are in the absence of an applied field. We also know how polarization changes as a function of field applied to the crystal. In short, according to the theory, a ferroelectric possesses two minima (e.g., a second-order phase transition) in the internal energy. These two minima are separated by an energy barrier $\Delta E$. Essential feature of a ferroelectric is that these two minima corresponds to two different spontaneous
polarizations that can be changeable reversibly by an applied field. Under an assumption that applied electric field is able to surmount the energy barrier, the advent of smart thin-film technology in evolution of CMOS technology, has enabled to consider a ferroelectric crystal a useful application. Thinning a ferroelectric film with high purity means that there could be an opportunity to use ferroelectrics as a memory element.

On the other, integrated ferroelectrics are a subject of considerable research efforts because of their potential applications as an ultimate memory device due to 3 reasons: First, the capability of ferroelectric materials to sustain an electrical polarization in the absence of an applied field, means that integrated ferroelectric capacitors are non-volatile. They can retain information over a long period of time without a power supply. Second, the similar architectural configuration of memory cell-array to conventional ones, means that they are highly capable of processing massive amounts of data. Finally, nano-second speed of domain switching implies that they are applicable to a high-speed memory device. Since ferroelectric capacitors was explored for use in memory applications by Kinney et al. (Kinney et al., 1987); Evans and Womack (Evans & Womack, 1988); and Eaton et al. (Eaton et al., 1988), it has been attempted to epitomize ferroelectrics to applicable memory solutions in many aspects. In the beginning of 1990’s, silicon institutes have begun to exploit ferroelectrics as an application for high-density DRAMs (Moazzami et al., 1992; Ohno et al., 1994). This is because permittivity of ferroelectrics is so high as to achieve DRAM’s capacitance extremely high and thus appropriate for high density DRAMs. An early version of non-volatile ferroelectric RAM (random-access-memory) used to be several kilo bits in packing density. This lower density (NB. at that moment, DRAM had several ten mega bits in density) is because of two: One is that its memory unit was relatively large in size, being comprised of two transistors and two capacitors (2T2C) to maximize sensing signal. The other is that a ferroelectric capacitor stack has required not only novel metal electrodes such as platinum, iridium and rhodium, all of which are hard to be fine-patterned due to processing hardness, but also reluctant metal-oxide materials to conventional CMOS integration due to possible cross contaminants such as lead zircornate titanate (PbZrTiO₃) and strontium bismuth titanate (Sr₁₋ₓBiₓTiO₃). Next steps for high density non-volatile memory have been forwarded (Tanabe et al., 1995; Sumi et al., 1995; Song et al., 1999). In similar to DRAM, an attempt to build smaller unit cell in size was in the late 1990’s that one transistor and one capacitor (1T1C) per unit memory was developed (Jung et al., 1998). Since then, many efforts to build high density FRAM have been pursued, leading to several ten mega bits in density during the 2000s (Lee et al., 1999; Kim et al., 2002; Kang et al., 2006; Hong et al., 2007; Jung et. al. 2008).

Fig. 8. (a) Evolution of electronic components in data throughput performance. (b) NVM (non-volatile memory) filling price/performance gap.
Among integrated ferroelectrics, one of the most important parameters in FRAM is sensing signal margin. The sensing signal of FRAM is proportional to remanent polarization ($P_r$) of a ferroelectric capacitor as follows:

$$\Delta V_{BL} = \frac{2P_r A}{C_{BL}} = \frac{2\varepsilon_0 \varepsilon A}{C_{BL}d}$$

(10)

where $A$ is capacitor’s area; $d$ is capacitor’s thickness. As seen in equation (10), in principle, we have to compensate the area reduction when technology scales down. However, in practice, when the thickness of PZT ferroelectric thin film decreases, degradation of polarization tends to appear in the ferroelectric capacitor due to a dead layer between the ferroelectric and electrodes (See section 3.3.3). Unlike the requirement of DRAM’s CAT, the array transistor of FRAM is not necessarily constrained from the off-leakage current due to no need of the refresh cycles, but from on-current, which is at least greater than several $\mu$A in order for a reasonable read and write speed. Thus, this will greatly relieve technology scaling quandaries and enable fast technology migration to the high end. This is because designing of a less leaky cell transistor becomes very difficult in incumbent memories such as DRAM and NAND/NOR flash due to need of lower doping concentration.

As witnessed in the Moore’s law, there has been enormous improvement in VLSI (very large-scale integration) technology to implement system performance of computing platforms in many ways over the past decades. For instance, data throughput of central processing unit (CPU) has been increased by thousand times faster than that of Intel 286™ emerged in the beginning of 1980’s. Alongside, a latest version of DRAM reaches a clock speed of more than 1 GHz. By contrast, state-of-the-art HDD (hard disk drive) transfers data at 600 MB/sec around (See Fig. 8a). Note that data rate of the latest HDD is still orders of magnitude slower than the processor/system-memory clock speed (see Fig. 8b). To achieve the throughput performance in more effective way, it is therefore needed to bridge performance gap in between each component. To compensate the gap between CPU and system memory, a CPU cache has been required and adopted. In line with this, ferroelectric memory is non-volatile, high-speed. But it has a destructive read-out scheme in core circuitry, whose memory cells need to return the original state after being read. This is because the original information is destroyed after read. As a result, it is essential to return the information back to its original state, which is so-called restore, necessarily following the read. This operation is so inevitable in the destructive read-out memory such as DRAM and FRAM. In particular, when the ferroelectric memory are used as one of the storage devices in computing system, such as a byte-addressable non-volatile (NV)-cache device, the memory has to ensure lifetime endurance, which is regarded as the number of read/write (or erase if such operation is required) cycles that memory can withstand before loss of any of entire bit information. Thus, authors are here trying to attempt not only how FRAM provides NV-cache solutions in a multimedia storage system such as solid state disk (SSD) with performance benefits but also what should be satisfied in terms of lifetime data-retention and endurance in such applications. Here, we also put forward size effect of ferroelectric film in terms of temperature-dependent dielectric anomaly because a dead layer plays an adverse role in thickness scaling. In addition, it is very important to ensure that integration technology of FRAM in nano-dimension is extendable to one of the

---

9File system cache is an area of physical memory that stores recently used data as long as possible to permit access to the data without having read from the disk.
conventional memories. Accordingly, we will present key integration technologies for ferroelectric memory to become highly mass-productive, highly reliable and highly scalable. This covers etching technology to provide a fine-patterned cell with less damage from plasma treatments; stack technology to build a robust ferroelectric cell capacitor; encapsulation technology to protect the ferroelectric cell capacitors from process integration afterwards; and vertical conjunction technology onto ferroelectric cell capacitors for multi-level metallization processes.

3.2 Non-volatile RAM as an ultimate memory solution

SSD, one of the multimedia storage systems, in general, consists of 4 important devices. First is a micro-controller having a few hundreds of clock speed in MHz, with real-time operating system (firmware). Second is solid-state storage device such as HDD or NAND flash memory, which has several hundreds of memory size in gigabyte. Third is host interface that has the primary function of transferring data between the motherboard and the mass storage device. In particular, SATA (serial advanced technology attachment) 6G (6th generation) offers sustainable 100 MB/s of data disk rate in HDD. In addition, bandwidth required in DRAM is dominated by the serial I/O (input/output) ports whose maximum speed can reach 600 MB/s. SATA adapters can communicate over a high-speed serial cable. Last is a buffer memory playing a considerable role in system performance. As such, DRAM utilization in SSD brings us many advantages as a buffer memory. For example, in DRAM-employed SSD, not only does I/O shaping in DRAM allow us to align write-data unit fitted into NAND flash page/block size but collective write could also be possible. As a result of sequential write, the former brings a performance benefit improved by 60% at maximum, and also the latter gives us another performance benefit improved by 17% due to increase in cache function, as shown in Fig. 9a and b, respectively.

![Fig. 9. Impact of DRAM utilization in SSD on system performance. (a) Increase in sequential read/write by I/O shaping. (b) Performance improvement by collective WRITE. (c) Additional performance benefit for DRAM plus FRAM in SSD.](image-url)

As an attempt to implement system performance further, not only does DRAM have been considered but FRAM has also been taken into account because of its non-volatility and random accessibility. Before that, it is noteworthy that, in SSD with no NV-cache, system-log manager is needed to record and maintain log of each transaction\(^\text{10}\) in order to ensure that

\(^\text{10}\)Each set of operations for performing a specific task.
file system maintains consistency even during a power-failure. A log file that contains all the changes in metadata, generally serves as a history list of transactions performed by the file system over a certain period of time. Once the changes are recorded to this log, the actual operation is now executed. This is so-called power-off recovery (POR). By contrast, POR is redundant in FRAM-employed-SSD as a NV-cache because metadata can be protected by FRAM. Elimination of POR overhead is the single most critical implementation by utilization of FRAM. This is because FRAM provides such system with byte-addressable and non-volatile RAM function. Thus, in spite of sudden power failure, system can safely be protected by adopting FRAM even without POR overhead, ensuring integrity of metadata stored in the ferroelectric memory. Through many benchmark tools, we have confirmed that by eliminating this overhead, system performance has been increased by 250% in random write (See Fig. 9c). This also brings the system to no need of flush operation in file system. As a consequence, this also brings the system to no need of flush operation in file system. As a consequence, additional 9.4% increase in performance, maximizing cache-hitting ratio. Since metadata frequently updated do not necessarily go to NAND flash medium, endurance of the flash memories can be increased by 8% at maximum as well. Besides, failure rate of operations can be reduced by 20% due to firmware robustness increased mostly by elimination of the POR overhead.

Fig. 10. Data locality of FRAM as a code memory.

Meanwhile, how many endurance cycles are necessary for use in applications of NV-cache solutions such as data memory and code memory? To answer this question, we need to understand access patterns of NV-cache devices in multimedia system. Now, we take into account the followings: First is the ratio of read/write per cycle in data memory (likewise, number of data fetching per cycle in code memory). Generally, the ratio for data memory and code memory is 1.00 and 0.75, respectively. Second is data locality\textsuperscript{11}. Figure 10 is a simulation result showing strong locality of 1.5% when FRAM has been considered a code memory.

\textsuperscript{11}The locality of reference is the phenomenon that the collection of data locations often consists of relatively well predictable clusters of code space in bytes.
memory. As shown in Fig. 10, less than 200 bytes of code space is more frequently accessed. Provided wear-leveling in read/write against the strong locality and taking an example of 20 MHz clock frequency of main memory (CPU clock ~ 200 MHz), what has been found is that the endurance cycles for 10-year lifetime becomes less than $9.5 \times 10^{13}$. This number of cycles is far less than the cycles we presumably assumed, which is more than $10^{15}$ cycles. Thus, authors believe that more than $1.0 \times 10^{14}$ of the endurance cycles is big enough to ensure that the ferroelectric memory as a NV-cache is so endurance-free as to be adopted to a multimedia storage system.

3.3 Reliabilities
3.3.1 Retention
Since Merz’s exploration of domain switching kinetics in the mid 1950’s (Merz, 1954), it is now believed that polarization reversal occurs in a way of domain nucleation and growth (Landauer, 1957; Pulavari & Kluebler, 1958; Key & Dunn, 1962; Du & Chen, 1997; Jung et al., 2002; Kim et al., 2005; Jo et al., 2006). The retention time of FRAM is closely related to a decay rate of the polarization reversal of a ferroelectric capacitor as expressed in formula (11).

$$P(t)/P_0 = \left(\frac{t}{t_0}\right)^{-n}$$  

(11)

$$\Delta F^* = -2EP_0V + \sigma_wA + Ud$$  

(12)

where $P_0$ is initial remanent polarization; $P(t)$ is remanent polarization at time $t$; $t_0$ is a time constant; $n$ is an exponent; $\Delta F^*$ is domain free energy; $E$ is homogeneous electric field applied externally; $V$ is the volume of domain nucleus; $\sigma_w$ is domain wall energy; $A$ is domain wall area. While the first term of Eq. (12) represents the electrostatic energy gained by formation of a domain nucleus, the second is the surface energy, and the last term is the field energy of the depolarizing field (Merz, 1954). Provided that length of domain nuclei is much smaller than thickness of a ferroelectric, half-prolate spheroidal nuclei tends to be formed and finally reaches a cylindrical shape (Key & Dunn, 1962; Jung et al., 2002). Under such an assumption, if one can measure depolarization energy of Eq. (12), we can now
estimate $\Delta F^*/k_B T$, where $k_B$ is Boltzmann constant. Based on experimental values of depolarization field $E_d$ that ranges from 300 to 800 kV/cm (Kim et al., 2005), the corresponding $\Delta F^*/k_B T$ is estimated to 4 to 20 at ambient temperature (Jo et al., 2006). Figure 11 represents (a) a decay exponent $n$ plot against estimated thermal energy $\Delta F^*/k_B T$ in various thickness of BaTiO$_3$ films and (b) thermal energy barrier $\Delta F^*/k_B T$ as a function of thickness in different ferroelectric stacks. As seen in Fig. 11a, in most of interesting nano-ferroelectrics with thickness ranging from 5 to 30 nm, the energy barrier is evaluated to $\Delta F^*/k_B T \approx 150 k_B T$ for $n \approx 0.017$, which is the exponent corresponding to 50% of polarization decay during 10 years in Eq. (11). Thus, as shown in Fig. 11b, if one takes into account a stack of SrRuO$_3$-PbTiO$_3$-SrRuO$_3$ (SRO-PTO-SRO), the energy barrier of polarization reversal via the formation of domain nuclei during 10 years is more than $150 k_B T$, which means that there is virtually no retention conundrum in FRAM as long as a ferroelectric stack is properly chosen.

### 3.3.2 Endurance

In FRAM, it is not readily achieved to assure whether or not a memory device can endure virtually infinite read/write cycles. This is because of memory size that is several tens or hundreds megabits typically. For instance, a HTOL (high temperature operational life) test during 2 weeks at 125 °C, is merely a few millions of endurance cycles for each memory cell in 64-Mb memory size, for example. Even taking into account minimum number of cells (in this case 128 bits because of 16 I/Os), time to take evaluation of $10^{13}$ cycles is at least more than 20 days. Therefore, it is essential to find acceleration factors to estimate device endurance through measurable quantities such as voltage and temperature. However, direct extraction of acceleration factors from memory chips is not as easy in practice as it seems to be in theory. This is because VLSI circuit consists of many discrete CMOS components that have a temperature and voltage range to work. Generally, more than 125 °C is supposed to be a limit to operate properly. A voltage range of a memory device is also specified in given technology node (±10% of $V_{DD}$=1.8 V in this case). Despite those difficulties, it has been attempted to figure out acceleration factors in terms of temperature and voltage, together with information obtained from capacitor-level tests.

In regard to package-level endurance, figure 12 represents changes in (a) peak-to-peak sensing margin ($SM_{pp}$) and (b) tail-to-tail sensing margin ($SM_{tt}$) as read/write cycles continues to stress devices cumulatively at 125 °C. Both $SM_{pp}$ and $SM_{tt}$ have been obtained by averaging out 30 package samples for each stress voltage. Function-failed packages have been observed when $SM_{pp}$ and $SM_{tt}$ reach 10% and 25% loss of each initial value, respectively. As seen in Fig. 12a and b, voltage acceleration factors ($AF_v$) between 2.0 V and 2.5 V has been calculated by these criteria ($AF_v = 81$ at $SM_{pp}$ and $AF_v = 665$ at $SM_{tt}$). In other words, the test FRAMs can endure $1 \times 10^{12}$ of read/write cycles at the condition of 125 °C and 2.0 V. Second, in capacitor-level endurance, figure 13 is (a) a normalized polarization plot against cumulative fatigue cycles at 145 °C in a variable voltage range and (b) a logarithm plot of cycle-to-failure (CTF) as a function of stress voltage in a various range of temperature. Here, we introduce a term of CTF which is referred to as an endurance cycle at which remanant polarization (or sensing margin) has a reasonable value for cell capacitors (or memory) to operate. Polarization drops gradually as fatigue cycles increase and the collapsing rate is accelerated as stress voltage goes higher. Likewise, provided 10% loss of polarization is criteria of CTF, the CTF at 145 °C and 2.0 V approximates $2.2 \times 10^{13}$. (NB. This
is reasonable because samples of 10% loss in $SM_{pp}$ turned out to be defective functionally.) Considering temperature- and voltage-acceleration factors from Fig. 13a, acceleration condition of 145 °C, 3.5V is more stressful in 5 orders of magnitude than that of 85 °C, 2.0 V. In other words, $1.0 \times 10^9$ of CTF at 145 °C, 3.5 V is equivalent to $6.0 \times 10^{14}$ at 85 °C, 2.0 V.

Results of the acceleration factors obtained from device-level tests differ from those in capacitor-level. For example, while $AF_V(2.5 V/2.0 V)$ of 81 in device-level tests, that of 16 in capacitor-level. We have yet to find a reasonable clue of what makes this difference. But it could be thought that the difference might arise from the fact that a memory device contains many different functional circuitries such as voltage-latch sense amplifiers, word-line/plate-line drivers, all of which make tiny amount of voltage difference magnify each effect on cell capacitors. This tendency can also be observed in the big gap of $AF_V$ obtained from two different definitions between $SM_{tt}$ ($AF_V = 665$) and $SM_{pp}$ ($AF_V = 81$). Tail-bit behaviors of memory cells could include a certain amount of extrinsic imperfection, in general. Thus, we believe that results tested in capacitor-level seem to be close to a fundamental nature of CTF.

Fig. 12. Changes in (a) peak-to-peak sensing margin ($SM_{pp}$) and (b) tail-to-tail sensing margin ($SM_{tt}$) as a function of endurance cycles at 125 °C. (c) $SM_{pp}$ vs. endurance cycles at 125 °C, 2.5 V. (d) $SM_{tt}$ vs. endurance cycles at 125 °C, 2.5 V. $SM_t$ and $SM_i$ of the ordinate in Fig. 12a and b is sensing margin at time $t$ and initial time, respectively.

---

It is thought that $AF_V$ in capacitor-level tests follows $AF_V$ of $SM_{pp}$ in device-level rather than that of $SM_{tt}$ because of nature of capacitor-level tests that average out all the cell capacitor connected in parallel.
Future Memory Technology and Ferroelectric Memory as an Ultimate Memory Solution

than those in device-level tests due to lack of extrinsic components. Figure 14 is (a) a logarithm plot of CTF as stress voltage increases in a various range of temperature and (b) Weibull distribution of endurance life in package samples tested at 125 °C in a various voltage range. The distributions in a 2.2-3.0 V range of voltage have a similar shape parameter, \( m \approx 2.4 \). This suggests that evaluation of endurance tests in device-level makes sense in physical term. As seen in Fig. 14a, voltage-endurance stress at less than 2.0 V does not allow us to obtain any sign of degradation in sensing margins within a measurable time span. Nor does temperature-endurance stress above 125 °C due to off-limits of operational specifications of the device.

![Fig. 13. (a) A normalized polarization plot against cumulative fatigue cycles at 145 °C in a variable voltage range. (b) Logarithm of CTF vs. stress voltage, \( V_{DD} \) at 145 °C.](image1)

![Fig. 14. (a) A logarithm plot of CTF as stress voltage increases in a various range of temperature and (b) distributions of endurance life in device-level tests at 125 °C.](image2)

### 3.3.3 Temperature-dependent dielectric anomaly
Since ferroelectricity involves the cooperative alignment of electric dipoles responding external field applied, there should be a critical volume below which the total energy associated with domain nucleation and growth, is outweighed by the entropic desire to
There has been a trend in recent literature to use the term “size effect” relating to the stability of spontaneous polarization to specifically describe the manner in which reduced size leads to progressive collapse of ferroelectricity (Saad et al., 2006). Finding the point at which this size-driven phase transition occurs is obviously interesting and fundamentally important, and thus various groups have done excellent works to elucidate, via both theory (Li et al., 1996; Junquera & Ghosez, 2003) and experiment (Streiffer et al., 2002; Tybell et al., 1999; Nagarajan et al., 2004), the dimensions at which ferroelectricity is lost. In that sense, one of the most critical quantities in ferroelectrics is remanent polarization $P_r$, which can be expressed as below:

$$P_r^2 \approx P_S^2 = -\frac{\alpha}{\beta},$$

(13)

and

$$\frac{1}{\varepsilon_0
\varepsilon_r} \approx \frac{1}{\chi} = -2\alpha = -\frac{2(T - T_c)}{\varepsilon_0 C},$$

(14)

where $P_S$ is spontaneous polarization; $\alpha$ and $\beta$ are standard bulk LGD (Landau-Ginzburg-Devonshire) coefficients, provided that ferroelectric materials have a second-order phase transition while neglecting the $P^6$ terms due to lack of contribution in the free energy expansion of the LGD theory (then, a hysteresis loop would be a cubic equation); $\chi$ is the dielectric susceptibility; $T_c$ is the transition temperature; and $C$ is the Curie constant. As denoted in Eq. (10) and (13), the sensing signal depends strongly on spontaneous polarization $P_S$ which is also varying material constants such as $\alpha$ and $\beta$. Eq. (14) is temperature-dependent dielectric anomaly, so-called, the Curie-Weiss law. Thus, in this section, we will examine whether or not size effect of ferroelectrics is intrinsic.

Fig. 15. Changes in dielectric constants as a function of temperature in BST materials: (a) Comparison of temperature-dependent dielectric constants between a ceramic bulk and a film 100-nm thick (Shaw et al., 1999). (b) Variation of relative permittivity as a function of temperature with a variety of thickness ranging from 15 to 580 nm (Parker et al., 2002)
In many ferroelectrics, ferroelectric phenomena could be ascribed to a dielectric origin, so-called, temperature dependent dielectric anomaly (Wieder, 1958; Pulavari & Kluebler, 1958). Since most integrated ferroelectrics are embedded as a thin film, it is desirable to pay much attention to the temperature-dependent dielectric properties in thin-film ferroelectrics. In this regard, there have recently been good approaches to evaluate size effects of ferroelectrics on their dielectric behaviors, in particular, in terms of temperature dependence. Figure 15 shows changes in the dielectric constant as a function of temperature in Ba$_{0.7}$Sr$_{0.3}$TiO$_3$ (BST) materials. As seen in Fig. 15a, Shaw et al. (Shaw et al., 1999) observed that temperature-dependent dielectric constant in a Ba$_{0.7}$Sr$_{0.3}$TiO$_3$ bulk ceramic undergoes a sudden change in value i.e., a first-order transition near ambient temperature at which a peak of dielectric constant in thin-film Ba$_{0.7}$Sr$_{0.3}$TiO$_3$ 100 nm thick, suffers a collapse of dielectric constant by orders of magnitude with severe broadening of Curie anomaly. This suggests a second-order transition. Along with the observation of Shaw et al., Parker et al. (Parker et al., 2002) measured variations of dielectric constant as a function of temperature over a variety of thickness ranging from 15 to 580 nm for Ba$_{0.7}$Sr$_{0.3}$TiO$_3$. They found that the temperature dependence of the dielectric constant exhibits diffusive shapes, also suggesting second-order transitions shown in Fig. 15b. They also found that the temperature maxima in the relative permittivity plots tend to decrease as the film thickness decreases, implying reduction of the transition temperature, $T_C$.

There are many possible origins to explain these temperature-dependent dielectric properties: First, these effects could arise from an intrinsic size effect that results in a drop in permittivity with decreasing sample dimension. Second is a model suggesting that a dead layer of grain boundary in BST films could have a low permittivity value compared to that of their grain interior; although the microstructure in the films has a columnar shape, resulting in a parallel rather than series capacitance contribution. Third, this is because of structural imperfection at film-electrode interfaces, consisting of interfacial dead layers and the biaxial strain caused by the thermal expansion mismatch with the substrate (Shaw et al., 1999; Parker et al., 2002). It is necessary to know whether the first case weights less severely
than the others, because the first is intrinsic. In this respect, Saad et al. (Saad et al., 2004a, 2004b) devised a method to thin bulk single-crystal BaTiO$_3$ using a focused ion beam (FIB) in order to evaluate the size effects of single crystal ferroelectrics thus excluding grain boundaries. The dielectric behaviors as a function of temperature in BaTiO$_3$ single crystals has been evaluated with a range of thickness from 447 nm to 77 nm (Morrison et al., 2005), fabricated from a bulk single crystal BaTiO$_3$. Figure 16 shows (a) a relative permittivity plot as a function of temperature in these single crystals of BaTiO$_3$ and (b) the reciprocal relative permittivity plot of the 77 nm BaTiO$_3$ as a function of temperature. Startlingly, dielectric constants have similar behavior to that of bulk BaTiO$_3$ single crystal even down to 77 nm thick. The dielectric constant in BaTiO$_3$ 77 nm thick gradually decrease over a range from 2,738 to 2,478 at temperature corresponding to 300 to 365 K, considerably increases and abruptly soars up to 26,663 at 410 K. The dielectric constant reaches a peak of 26,910 at 413 K.

In general, the dielectric constant in bulk BaTiO$_3$ single crystal are regarded as 160 for $\varepsilon_r$ (parallel to the polar axis) and 4100 for $\varepsilon_a$ (normal to the polar axis) at ambient temperature (Landauer et al., 1956; Benedict & Duran, 1958). In addition, the sudden change in dielectric constants due to the phase transition from FT (ferroelectric, tetragonal) to PC (paraelectric, cubic), occurs either 122 °C upon heating or at 120 °C on cooling (Merz, 1953; Drougard & Young, 1954). In Fig. 16a, the transition temperature $T_C$ is a little bit different from one of bulk BaTiO$_3$. Morrison et al. (Morrison et al., 2005), however, think that this difference may be caused by the fact that the temperature of thermocouple placed on a heater block is a little bit higher than that on the sample. Thus, considering the temperature artefact, the abrupt change in dielectric constant occurs at a temperature close to that observed in bulk BaTiO$_3$. Alongside the dielectric constant as a function of temperature, the inverse of the Curie constant as a function of temperature is shown in Fig. 16b for the 77-nm BaTiO$_3$ single crystal. According to the Curie-Weiss law, the Curie-Weiss temperature $T_0$ can also be estimated at 382 K from the extrapolation as shown in Fig. 16b. As a result, for the 77-nm BaTiO$_3$ single crystal, they can obtain that the difference $\Delta Temp$ between $T_C$ and $T_0$ is approximately 13 °C, which is quite good agreement with experimental results obtained from bulk BaTiO$_3$ single crystal, in which $\Delta Temp = 14$ °C (Merz, 1953; Drougard & Young, 1954). These results provide a very interesting and promising clue, because ferroelectric properties even in 77-nm thickness are expected to show a similar dielectric behavior with that of bulk BaTiO$_3$. In addition, the first-order transition from FT to PC in ferroelectrics can appropriately be described by the dielectric behaviors near the transition tempemperatures. They conclude therefore that, down to 77 nm dimension, the intrinsic size effect has negligible influence on the temperature-dependent dielectric properties. Moreover, it is not difficult to estimate the Curie constant $C$ from the Curie-Weiss plot because the 77-nm sample of BaTiO$_3$ exactly follows the typical Curie-Weiss law as shown in Fig. 16b. From the slope of 1/ $\varepsilon_r$ vs. $T$, the Curie constant is approximately $4.53 \times 10^3$ °C, which is compared to experimental values of $1.56 \times 10^3$ and $1.73 \times 10^3$ °C, obtained by Merz (Merz, 1953) and Drougard and Young (Drougard & Young, 1954), respectively. The Curie constant is in the same order of magnitude but is roughly 3 times larger than those compared. This may be because of two

---

3It was widely accepted that the Curie point of undoped crystal and ceramic BaTiO$_3$ was near 120 °C. Measurements on highly purified ceramics and on crystals grown by Remica’s process (Remica & Morrison Jackson, 1954) but without the addition of Fe$^{3+}$ have shown that their Curies temperature is near 130 °C (Jaffe et al., 1971).
factors: Errors in electrode area and thickness can affect the Curie constant dramatically; and the temperature difference between sample and thermocouple may not be constant.

3.4 Key technologies

Etching damage: It is widely accepted that as a device shrinks, node separation of cell capacitors is not readily achievable due to necessity of novel metals that served as electrodes of the MIM (Metal-Insulator-Metal) cell capacitor, such as iridium, iridium oxide, strontium ruthenium oxide (SrRuO$_3$). In typical, remanent polarization depends heavily on processing temperature at which ferroelectric PZT (PbZr$_{0.4}$Ti$_{0.6}$O$_3$) is etched. The remanent polarization ($P_r$) value drops drastically as temperature of the processing chuck in an etching chamber increases. According to a report of etching impact on ferroelectrics (Jung et al., 2007), there is no direct evidence how higher-temperature etching makes a $P_r$ value smaller. But it is believed that a certain amount of halides or halide ions might accelerate chemical reduction during the etching process at higher temperature, in particular, at the interfaces of the cell capacitors. Thus, Jung et al. (Jung et al., 2007) reported that ferroelectric cell-capacitors suffering a severe etching damage, are likely to follow bulk-limited conduction such as space-charge-limited current (SCLC), rather than those of electrode-limited.

Fig. 17. Cross-sectional micrographs both (a) in a peripheral circuitry region and (b) in a cell region, (c) in which one of the cell capacitors is pictured (Jung et al., 2008).

Stack technology: Building a stack for a robust ferroelectric cell capacitor is a more important part of the entire integration than any other process due to the fact that the preparation of a ferroelectric thin-film plays a crucial role in whether the cell capacitors have the ferroelectric properties in a certain level of integration. For example, Qos-retention charge of a sol-gel derived PZT film is severely degraded if one evaluates non-volatile polarization by using the two-capacitor measurement technique$^{14}$. This tells us how a ferroelectric film is

$^{14}$Qos-retention means opposite-state charge retention that is change in non-volatile polarization values elapsed after a certain amount of time and temperature stress, before which the two capacitors are written to data 1 (D1) and data 0 (D0). In general, the Qos-retention has a faster decay rate than Qss-retention (same-state charge retention) does under the same acceleration condition because imprint change has a much more severe impact on degradation of non-volatile polarization than depolarization increases.
vulnerable to loss of ferroelectricity when film preparation is poor. The memory device integrated with CVD (chemical vapor deposition)-derived PZT film has twice bigger sensing margin than that the sol-gel-based device has even after severe suffering of a thermal acceleration test during 1000 hours at 150 °C. In addition, it is also important to regulate deposition temperature in CVD preparation of PZT films. SMpp of FRAM with the PZT film prepared at adequate temperature is more than 650 mV, otherwise FRAM with a less optimized PZT film has SMpp less than 550 mV (See Fig. 18).

<table>
<thead>
<tr>
<th>Integration technologies</th>
<th>Case A</th>
<th>Case B</th>
<th>Case C</th>
<th>Case D</th>
<th>Case E</th>
<th>Case F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Etching temperature</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>PZT deposition</td>
<td>Regulated</td>
<td>Regulated</td>
<td>Not</td>
<td>Regulated</td>
<td>Regulated</td>
<td>Regulated</td>
</tr>
<tr>
<td>Capping thickness</td>
<td>Thick</td>
<td>Thick</td>
<td>Thick</td>
<td>Thick</td>
<td>Thin</td>
<td>Thick</td>
</tr>
<tr>
<td>Recovery Anneal</td>
<td>No anneal</td>
<td>No anneal</td>
<td>No anneal</td>
<td>No anneal</td>
<td>No anneal</td>
<td>Anneal</td>
</tr>
</tbody>
</table>

Table 2. A list of combination of different integration conditions.

**Encapsulation Technology:** In general, ferroelectric capacitors comprise a perovskite-oxide-based ferroelectric film and novel metals that have a catalytic effect on oxide layers. The metallic electrodes of the ferroelectric capacitors consist of top-electrode (TE) SRO underneath iridium and bottom-electrode (BE) iridium. Due to these novel metals, oxide of the perovskite ferroelectric is very prone to chemical reduction during many hydrogen-based processes such as interlayer dielectrics (ILD) and inter-metallic dielectrics (IMD). Thus, it is essential for protecting the capacitors from these integration processes in order to build a robust capacitor. Thus, a ferroelectric cell capacitor seems to be capped with Al₂O₃ that needs to be deposited conformally on its sidewall. The Al₂O₃ layer is, typically, prepared by an atomic-layer-deposition (ALD) method. By opting a thicker Al₂O₃ layer, one can have not only a sharper distribution of bit-line potential but 33% increase in SMpp as well, compared with the case of an Al₂O₃ layer thinner.

**Vertical conjunction:** FRAM has similar architecture with one of the DRAMs, featured by folded bit-line and voltage-latch sense amplifiers. But a prominent difference between FRAM and DRAM is, in architecture, how to form the plate node of a cell capacitor—the other end is connected to the storage node of a cell transistor in both DRAM and FRAM. While a bunch of plate nodes in DRAM is connected together, a few plate nodes in FRAM should be separated. The reason of the separation is to give a plate pulse independently to each plate line. Due to this essential contact between cell capacitors and the plate lines, metallization in FRAM needs a special care in integration. This is because contact forming onto the top electrode of a cell capacitor may provoke another root-cause of capacitor degradation during the process integration. Since it is suitable for protecting ferroelectric capacitors from any involvement of aluminum when forming the plate line and strapping line, an addition-top-electrode (ATE) scheme has been adopted for this contact formation (Kim et al., 2002). The ATE landing pad consists of iridium oxide and iridium. Through a proper anneal process, what has been achieved is to decrease data 0 population of bit-line potential as low as possible, so that 8% improvement in SMpp is attained.

Figure 18 summarizes (a) populations of bit-line potential as integration differently applied and (b) tail-bit populations of V_{BLD1} and V_{BLD0} for the integration scheme of the case F in table 2. The number of dies is 150 in total. Table 2 also summaries how each integration technology to combine. The overall population of bit-line potential has a strong impact on changes in data 1 distribution when each technology varies as shown in table 2. First,
imperfect encapsulation of the cell capacitor causes bit distribution to become wider and bigger loss of the peak value in data 1 that corresponds to switching charge quantity in ferroelectric cell capacitors. This charge lessening effect may be accelerated under the severe etching condition, for example, etching at high temperature. That is why the case E shows the smallest bit-line distribution in Fig. 18a in spite of the fact that the PZT thin film is properly deposited at a regulated condition. Second, when one applies a poorly regulated deposition condition to a ferroelectric thin-film preparation, broadness of cell-charge distribution appears dominantly as seen in the case C of Fig. 18a. Third, etching of ferroelectric capacitors in highly reduced ambient could result in tailing of data 1 distribution, giving rise to a certain loss of sensing margin as seen in the case B of Fig. 18a. Last, the contact formation onto the top electrode of cell capacitors should be emphasized because it might have an advantageous effect in the distribution of data 0 not only on lessening of the peak value but on being sharp without any loss of the data 1 distribution, as shown in the case F of Fig. 18a. Through the combination of key integration technologies, 525 mV of $SMtt$ in sensing margin has been achieved (Jung et al., 2007). To recapitulate it, preparation of ferroelectric capacitors is very important to realize highly reliable and scalable FRAM. But all the integration technologies followed by the capacitor stacking is equally important, in particular, in a smaller dimension. This is because nano-scaled ferroelectric capacitors are so vulnerable as to lose the ferroelectric properties during ever-growing integration processes as reported here.

Fig. 18. (a) Data 1/Data 0 distributions of bit-line potential as integration technology varies from case A to F (See Table. 2). (b) Tail-bit populations of $V_{BLD1}$ and $V_{BLD0}$ for an integration scheme in table 2. The number of dies is 150 in total.

3.5 Conclusions
Utilization of FRAM as a NV-cache solution in a multimedia storage system such as SSD, gives users critical advantages. By elimination of POR overhead due to its non-volatility, random-write throughput can be enhanced by more than twice. In spite of strong data locality of FRAM, 10-year lifetime endurance has been estimated to be less than $1.0 \times 10^{14}$ cycles in such system. This endurance is much less than that we presume (e.g., $\sim 10^{15}$ due to every-time access for 10 years). From the investigation of acceleration factors both in device-level and in capacitor-level, CTF of the FRAM evaluated has been estimated to
approximately $6.0 \times 10^{14}$ at a system operating condition. To be in a nutshell, ferroelectric memory as a NV-cache seems to be a very plausible scenario for increase in data throughput performance of SSD. In assertion of endurance, lifetime endurance is no longer problematic even in the FRAM based on a destructive read-out scheme. On top of that, the introduction of ferroelectric materials to conventional CMOS technologies has brought us to realize non-volatile, byte-addressable and high-speed memory. This is thanks not only to bi-stable states of a ferroelectric but also to tremendous efforts done by many institutes around the world, trying to epitomize it in two folds. One is, mostly done by silicon institutes, development of thin-film technology with high precision and high purity for a ferroelectric cell capacitor. The other is, mainly pursued by academia, to scrutinize thin-film ferroelectrics for whether or not their intrinsic properties (e.g., order parameters) are restricted by scaling of capacitor’s thickness, so-called size effect. What both found is that ferroelectric properties is not restricted by scaling of thin ferroelectrics, at least within a concerned integration range of thickness, e.g., less than 10 unit perovskite-cells in polar axis are enough to have stable minima in dipole energy. Note that lattice constant of ferroelectrics is several Angstroms. Also, what they found is that a dead layer is not fundamental one in extremely thin ferroelectric capacitors. This suggests that gigabit density NV-RAMs by using ferroelectrics will be in the market place in the future, under an assumption that FRAM follows DRAM’s approach to build ferroelectric cell capacitors in a 3-D way. Such assumption is not an illusion because physical thickness of storage dielectrics in state-of-the-art DRAM, is several ten Angstroms.

4. References


Future Memory Technology and Ferroelectric Memory as an Ultimate Memory Solution


Morrison, F.; Jung, D. & Scott, J. (2005). In collaboration with Saad et al., 2004a, we independently evaluate the dielectric behaviors as a function of temperature in BaTiO$_3$ single crystal with a range from 447 to 77 nm.


www.intechopen.com


Ferroelectric materials have been and still are widely used in many applications, that have moved from sonar towards breakthrough technologies such as memories or optical devices. This book is a part of a four volume collection (covering material aspects, physical effects, characterization and modeling, and applications) and focuses on the application of ferroelectric devices to innovative systems. In particular, the use of these materials as varying capacitors, gyroscope, acoustics sensors and actuators, microgenerators and memory devices will be exposed, providing an up-to-date review of recent scientific findings and recent advances in the field of ferroelectric devices.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
