# 2D MLC vs. 3D TLC

## Why not all NAND flash TLC is the same when it comes to 3D storage

One way to measure the drop-in price of flash memory is to regularly visit exhibitions. Back in the early 2000s, when USB flash drives (Figure 1) were new, novel, and capable of storing 8 MB of data, it was unlikely that you would have received such a device as a gift unless you were a special customer. As the years have rolled on, the capacities and access to these corporate gifts has grown and grown. Initial storage capacity growth could be implemented thanks to improvements in lithography processes. Moving from 43 nm, through 24 nm, down to 15 nm, ever smaller transistors ensured that more memory cells could be packed into each NAND flash die, thereby feeding the industry's hunger for storage capacity.

The challenge at this process point was not, however, the transistor. Flash memory makes use of a floating-gate MOSFET (FGMOS) where the gate is electrically isolated thanks to a capacitor. Because the surrounding material is highly resistive, the charge stored in this component can remain unchanged for long periods of time. The challenge, therefore, is the basic physics of the capacitor, whose capacitance is dependent upon its area. Regardless of the lithography used, if the capacitance required for the design remains constant (in the range of a few femtofarads), so will the size of the capacitor from process shrink to process shrink.

*Figure 1a and 1b: USB storage comes in a variety of capacities and housings, even utilizing the latest fast USB interfacing technology.*

Improvements in lithography were not the only capacity improvement being undertaken. First generation flash technology stored a single bit of data in each memory cell, termed Single Level Cell (SLC) architectures. These used the presence of charge in the cell, or the lack of charge, to indicate the storage of a '0' or '1' bit of data. In order to push storage capacities for the same die area, the density of the bits was increased by storing more than one bit of data in each cell.

## The move to more bits per cell

Two bits per cell, known as Multiple Level Cell (MLC) devices were achieved by storing one of four different voltage in each cell (Figure 2). Each level coded one of the values '00', '01', '10', or '11'. With the continuing demand for ever higher capacities showing no sign of abatement, the next natural step was to progress to eight voltage levels. This allowed the encoding of value between '000' and '111' in Triple Level Cells (TLC).

Of course, like in almost every situation, improvements come at a cost. In the case of NAND flash this was the durability of the resultant device. NAND flash memory has a limited life span that is measured in terms of the number of erases and writes to the cells. Similar to a mechanical system, the cells wear, resulting in their eventual failure. The additional stress placed upon the technology by supporting different voltage levels further reduced the lifetime of the cells. Storage devices, such as USB flash drives, can also implement functionality in their driver hardware, such as wear levelling, that can help compensate for this effect and lengthen the products lifetime. However, as is the case with popular trends, this led to a flood of high capacity, but low quality, USB flash drive products on the market that failed after relatively little use. Their use of MLC and TLC memory was often thought to be the cause of their mediocre lifetime and performance.

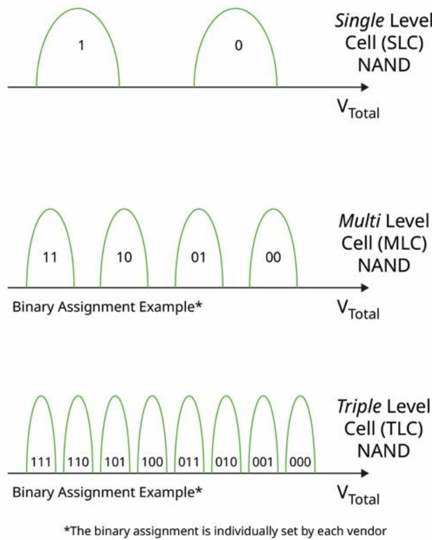


*Figure 2: By storing more than just two voltage levels per cell, each flash memory cell can two or more bits of data.*

For embedded developers, working on automotive applications and industrial systems, this raised a number of red flags. Unlike consumer products such as USB flash drives, their products needed to guarantee operation over a wide temperature range and under challenging environmental conditions. While flash memory utilizing MLC or TLC were financially attractive and provided the higher capacities desired, the risks associated with the reduced number of write/erase cycles could not always be accommodated.

In high-reliability applications, SLC would often be specified to deliver on the reliability guarantees the application demanded. As an alternative, pseudo SLC (pSLC) solutions were also offered. These offered MLC or TLC technology memories with an SLC partition for storing critical sections of the application code, such as the bootloader, at the expense of reduced capacity for this memory region. The overall outcome has been the perception that MLC and TLC flash memory is not of the same quality as SLC options.

## From planar to 3D

As it became clear that a blind rush towards fulfilling Moore's Law would not resolve the capacity challenges of memory technology, alternative innovation approaches were sought. KIOXIA, as the inventor of flash memory, applied its significant research and development resources to 3D flash manufacturing approaches. The idea was simple: take the existing

planar 2D design and fold it in the middle to create a vertical memory structure. However, this was not the only innovation undertaken.

The FGMOS technology used up to this point made use of polycrystalline silicon gates. Silicon is a non-isolating semiconductor, so the charge accumulated still has mobility after being stored. This requires strong isolation to maintain the charge and, in turn, the data stored. As a result, the oxide used to restrict the flow of electrons from the floating gate (FG) needs to be relatively thick. Here there is another challenge. The process of charging and discharging the cell (to store a 0 or 1) requires that the charge pass, or tunnel, through this oxide. Tunneling through the oxide leads to the material's degradation as a result of continued erase and write cycles, which leads to a decrease in the reliability of the cell.

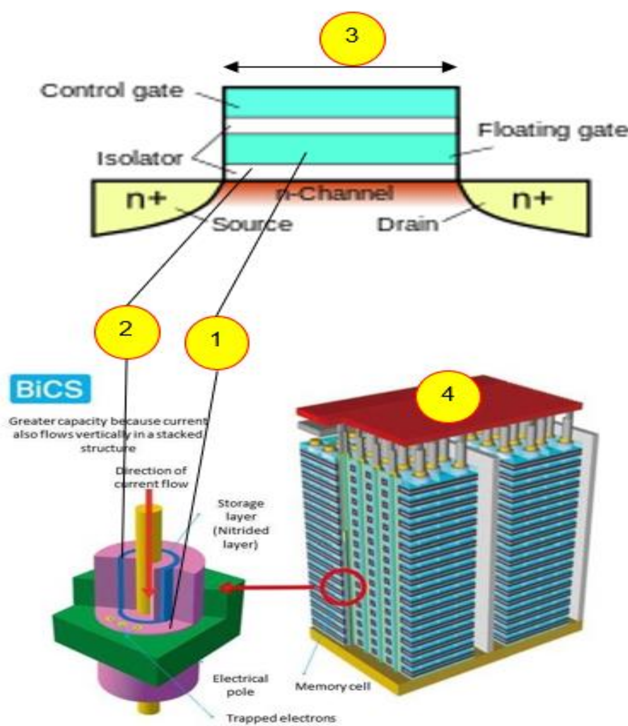


*Figure 3: The trapped charge capability of silicon-nitride NAND memory cells leads to more charge per cell, reduced cell-to-cell crosstalk, and support for the 3D BiCS implementation.*

| Topic | Floating Gate (FG) | Charge Trap (CT) |
|---|---|---|
| 1 – Charge storage | Polycristalline Silicon gate (floating gate) | Silicon-Nitride isolator (charge trap) |
| 2 – Isolation oxide | Thicker; leads to degradation and decreasing reliability | Thinner; less sensitive to degradation effects |
| 3 – Cell dimensions | 1x nm process node is at the technology limit | Larger cells, each storing more charge |
| 4 – Topology | Planar – capacity increase requires smaller cells | 3D – capacity increases achieved in third dimension (stacking) |

*Table 1: Explanation for technical improvements highlighted in 3D BiCS flash in Figure 3.*

The move to the smallest lithographic processes had also reduced the amount of charge that could be stored per cell to very low levels. While this proved challenging enough to deal with for SLC devices, MLC and TLC were further impacted by the effects of crosstalk due to the limited cell-to-cell distances involved. This led to an increase 'read disturb' errors. With this collection of challenges, it was clear that a simple vertical implementation of an FGMOS process would not deliver the desired level of reliability and quality demanded.

## Charge trap memory

The challenges of the FGMOS approach were overcome through the use of charge trap (CT) technology. By moving from using silicon to silicon-nitride, an isolator, ensured that the charge stayed caught in the charge trap. As a result, the isolation oxide can be much thinner, in turn making the material less sensitive to the degradation suffered by floating gate technology (Figure 3 and Table 1). This also allowed a lower programming voltage to be used, saving energy.

Because capacity could then also additionally be defined by the number of layers of cells implemented, it is no longer necessary to use the finest lithography processes. This allows the CT memory cells to be significantly larger than FG NAND cells, enabling them to store more charge. Not only does this simplify the implementation of TLC cell states, as the cell-to-cell crosstalk is reduced, it opens up the possibility of quad-level cells (QLC). As a result of this move to silicon-nitride CT technology, cell reliability, even for multiple bits per cell, is

significantly improved. This, coupled with the move into the vertical domain, has provided the necessary step improvement in device capacity demanded.

CT memory forms the basis of KIOXIA's 3D BiCS (Bit Column Stacked) technology. The initial demonstration devices, BiCS2, were implemented using a 48-layer design. Studied under an electron microscope, the structure looks very similar to that of a high-rise building (Figure 4). Current flows through the cylindrical columns that stretch from the base up to the bits lines that extend along the surface of the device. Since its inception, further iterations have added additional layers so that the current BiCS5 offers 112 layers, making a single package 1 TB device possible.

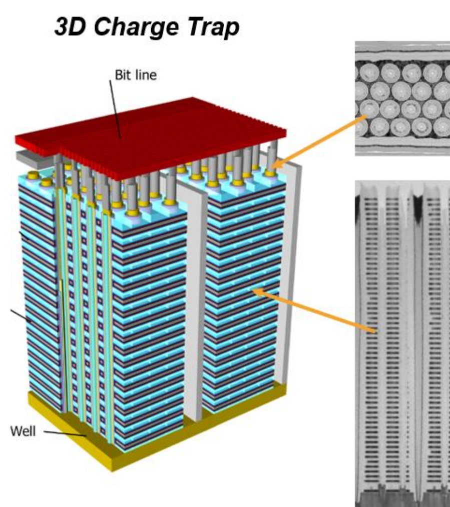


*Figure 4: Under an electron microscope, 3D BiCS flash memories resemble a high rise building due to their tall, cylindrical charge-trap towers.*

Packaged memories have relied on a stacked-die approach for some time in order to fulfil market demand for high capacity in a single device. At the same time, the data interfaces (eMMC, UFS) have continued to push the bandwidth boundaries. The use of bond wires to interconnect die and connect die to the lead frame became increasingly challenging to bond due to their length, as well as proving a bottle neck for data transfer speeds. To resolve this, the use of Through Silicon Via (TSV) technology was also adopted. This allows die of equal size to be stacked on top of one another without the use of bonding wire. The shortened electrical pathways result in reduced parasitic effects and delivered the desired data transfer improvements, while also reducing power consumption.

## Not all TLC is the same

While it may seem prudent to take small steps when undertaking major technological advancements, such as starting with an SLC implementation, BiCS technology was implemented using TLC from its inception (Figure 5). And, while future generations of the technology will see it continue to grow vertically, it will also be offered in a QLC architecture too as an additional way of improving capacity. This naturally results in alarm bells ringing amongst those who have had poor experiences with 2D TLC devices in the past.

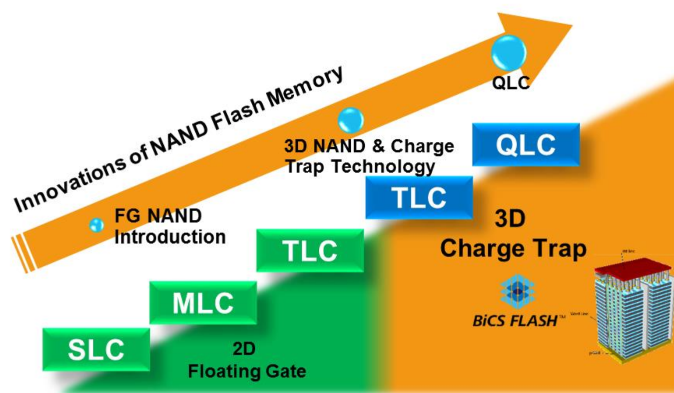


*Figure 5: 3D NAND BiCS has made use of TLC storage since its inception.*

However, as highlighted, the move to larger cells and their implementation using silicon-nitride, provides a significantly more reliable NAND flash product. In fact, 3D BiCS TLC technology can be considered on par with 2D FG MLC implementations, while the upcoming 3D BiCS QLC implementation will match 2D FG TLC, both with respect to reliability and lifetime. Furthermore, developers also profit from a lower cost per bit, high performance, reduced power requirements, and high storage density per die. This latter point means that a single-package die, featuring 16-die stacked architecture QLC, will deliver an incredible 2.66 TB of storage.

BiCS technology has also established itself as a leading NAND flash solution for a range of consumer and enterprise applications, through SD cards and SSDs, to the world of embedded systems, such as in e-MMC and UFS storage.

## Summary

It is fully understandable that some market segments, such as those that have the highest demands on reliability and product lifetime, view TLC NAND flash technology with a certain level of skepticism. Some past 2D implementations of TLC technology simply did not meet their expectations. However, like in all areas of technology, things do not stand still. 3D NAND BiCS, with its charge trap silicon-nitride approach, is a different technology to 2D floating gate solutions. The advancements made not only provide a more reliable and robust cell, it is more than suited for use in TLC architectures. Thanks to the resultant larger cell size, it can be considered on par with 2D MLC devices with respect to reliability and lifetime. As a result, design teams can draw upon this new technology secure in the knowledge that their data is safe and their product quality is maintained, while also taking advantage of the significant capacity and throughput improvements 3D BiCS solutions have to offer.

**Disclaimer:**
Definition of capacity: KIOXIA defines a megabyte (MB) as 1,000,000 bytes, a gigabyte (GB) as 1,000,000,000 bytes and a terabyte (TB) as 1,000,000,000,000 bytes. A computer operating system, however, reports storage capacity using powers of 2 for the definition of 1GB = 2^30 = 1,073,741,824 bytes and therefore shows less storage capacity. Available storage capacity (including examples of various media files) will vary based on file size, formatting, settings, software and operating system, such as Microsoft Operating System and/or pre-installed software applications, or media content. Actual formatted capacity may vary.

All company names, product names and service names may be trademarks of their respective companies.

Product illustrations may present design models or previous versions. Actual product may vary.

Images are for illustration purposes only.

© 2020 KIOXIA Europe GmbH. All rights reserved. Information, including product pricing and specifications, content of services, and contact information is current and believed to be accurate on the date of the announcement, but is subject to change without prior notice. Technical and application information contained here is subject to the most recent applicable KIOXIA product specifications.