



# Post-NAND Flash: How Does Schiltron Compare?

**Andrew J. Walker**

**Schiltron Corporation**

1638 Cornell Drive  
Mountain View, CA 94040 USA  
(email: [andy@schiltron.com](mailto:andy@schiltron.com) ; tel: +1 408 425 4150)

## **ABSTRACT**

There are a number of competitors claiming to be the successor of classic Floating Gate NAND Flash. These include: various forms of vertical 3-D NAND such as BiCS, pBiCS and TCAT; VG-NAND; various forms of impedance switching materials collectively known as RRAM; and the Schiltron approach. This white paper compares these main approaches using eight attributes: 3-D stackability; scalability; manufacturability; retention after cycling; disturbs; sense currents; program/erase voltages and design maturity. It is shown that the Schiltron approach is the only one where all attributes are optimized simultaneously.

## **I. INTRODUCTION**

The two main branches of technologies claiming to be the successor of classic Floating Gate NAND Flash are:

- Charge Trap Flash (CTF) such as BiCS, pBiCS, TCAT, VG-NAND and Schiltron;
- Various forms of impedance switching materials also known as RRAM.

Within the CTF approaches, two main subdivisions are discernible, namely:

- Vertical current flow (BiCS, pBiCS, TCAT);
- Horizontal current flow (VG-NAND, Schiltron).

Within the RRAM space, there are many different approaches with many different materials. It is not the intention of this white paper to deal in detail with every variant but to highlight the main attributes that all approaches should be measured against. References to academic papers are not made since these can all be found in Schiltron's other white papers and peer-reviewed articles<sup>1</sup>

## II. COMPARISONS

### 3-D STACKABILITY

This measures the ability to monolithically stack several memory cell layers on top of each other and all on top of a control circuit that would reside in the bulk of the silicon. The important yardsticks are process thermal budget and the ability to integrate a select device in the 3-D stack (such as a diode or transistor).

Any monolithic 3-D approach that requires excessive temperatures during manufacturing to create the cell layers will cause parameter shifts in both the underlying bulk control circuitry and in the first cell layers. Such shifts may be so great as to be beyond correction. An example is dopant diffusion in polycrystalline silicon which is usually orders of magnitude greater than in monocrystalline silicon. Therefore, any critical dopants in the cell layers may move to such an extent as to make any device unusable.

---

<sup>1</sup> <http://www.schiltron.com/WhitePapers.html>

Any of the CTF approaches depend on similar deposition steps of dielectrics and amorphous or polysilicon layers all of which can be done at low enough temperatures. RRAM approaches for minimum cell footprints require either a select device such as a diode integrated into the memory stack or must have an inherent diode on programming. In any case, the impedance switching material needs to be able to withstand the process temperatures of subsequent layers. This remains to be shown.

### SCALABILITY

This measures the ability to shrink the device in the lateral dimension. All approaches seem to be viable in this regard except for an interesting challenge associated with any of the vertical 3-D NAND structures such as BiCS, pBiCS and TCAT. In these, the lateral dimension includes twice the thickness of the charge trapping dielectric stack. From published work on the TANOS structure, it is known that this dielectric stack must be of a certain thickness to try to minimize pass disturbs in the NAND series string. Therefore, to laterally shrink any vertical 3-D NAND structure will undoubtedly require a thinning of this dielectric stack and a resultant increase of the pass disturbs. This is exacerbated even further if the string is also lengthened since each cell would be disturbed more often than in a shorter string. The TANOS gate dielectric stack is on the order of ~20 nm thick. Therefore, the cylinder containing the channel silicon and the gate dielectric stack would be at least 40nm plus the diameter of the silicon channel. An absolute minimum would appear to be about 50nm.

In the case of the Schiltron architecture, the dual-gate approach results in excellent electrostatic control within the channel. As regards alignment of the top gate to the bottom



gate, several comments can be made. The 193 nm ASML NXT 1950i scanner has an overlay capability of  $2.5 \text{ nm}^2$ . This is for a single machine overlay including chuck dedication. Actual overlay in process with product wafers may be a nm or so more due to wafer stress effects. In the DG-TFT-SONOS approach from Schiltron, the top gate to bottom gate alignment requirement is in one direction only. This then results in a cell dimension in this direction of  $2F + 2\delta$  where F is the minimum feature size or space and  $\delta$  is an increase on each side of the bottom gate to allow for misalignment of the top gate. The worst case is to assume that this is the overlay tolerance plus process effects. The other cell direction can be a true 2F size. This “asymmetric” scaling allows for the following cell size per node.

<b>F (nm)</b>	<b>Cell Size (nm<sup>2</sup>)</b>	<b>Multiplier of F<sup>2</sup> for Cell Size</b>
20	1880	4.7
17	1394	4.8
15	1110	4.9
13	858	5.1
10	540	5.4

**TABLE 1: Schiltron cell size per 3-D layer at each technology node (F). The overlay between top and bottom gate was 3.5 nm.**

<sup>2</sup> <http://www.asml.com/asml/show.do?ctx=6720&rid=36951>



In actual manufacturing, part of  $\delta$  can be taken from the space between the bottom gates which would allow for an even smaller cell size at each technology node.

### MANUFACTURABILITY

This attribute is the result of materials, fab tool maturity and thermal budget interactions. All approaches except RRAM use existing and historically well known materials and tooling with known thermal budget interactions. RRAM approaches tend to use materials that are not common in silicon fabs. This in itself may not be such a problem but questions remain about material stability, interactions, thermal budget effects and tool maturity.

All the SONOS type approaches (BiCS, pBiCS, VG-NAND and Schiltron) use existing well-proven fab infrastructure.

### RETENTION AFTER CYCLING

Any nonvolatile memory claiming to be the replacement for NAND Flash must have the ability to retain its programmed and erased states after endurance cycling between these states and at elevated temperatures. The Schiltron approach has already shown 10 years retention at 85 °C after 100,000 program/erase cycles. This is the best performance of any NAND replacement technology published so far (as far as the author knows). Both the vertical 3-D NAND and the VG-NAND need the gate dielectric stack to be engineered in such a way as to minimize pass disturbs in the strings. This results in a tunnel oxide (the thin oxide in the gate dielectric stack that is in contact with the channel) that is thicker than classic SONOS values where the latter are usually between 1.5 nm and 3.0 nm. A thicker tunnel oxide results in damage during endurance cycling which also leads to retention



problems after cycling. Retention results published for vertical 3-D NAND and VG-NAND are usually given on cells with either no cycling or very limited cycling and sometimes just at room temperature.

The outstanding results for Schiltron can be understood as follows. The dual-gate architecture allows the decoupling of pass disturbs from the memory gate dielectric stack engineering. This allows classic SONOS “ONO” thicknesses to be used. These in turn allow program/erase endurance without discernible damage out to a million cycles or more. The lack of damage results in excellent retention after cycling.

The physics of SONOS retention is well understood. Its predictable nature is a significant advantage over other non-charge-trapping approaches. The understanding of RRAM retention is perhaps not as well developed and understood. Besides its apparent stochastic nature (where a bit can flip state without any warning), retention results across various materials have been decidedly mixed and need to be established before doubt is removed.

### *DISTURBS*

The acts of programming and reading cells must not affect the states of other cells not being programmed or read. If the states of these cells are affected beyond a budgeted amount, we speak of disturbs. The main problem of connecting SONOS Flash cells in a NAND series string is that they get disturbed very easily during read and program. During read, a small voltage is applied to the gate of the device being read but all other cells have a large voltage applied to their gates to make sure they turn on no matter what their threshold voltage is. This read-pass condition “soft programs” these cells. This is due to the thin

tunnel oxide associated with classic SONOS. The first technique to try to minimize this is to thicken up the gate dielectric stack and especially the tunnel oxide. In the case of the TANOS publications between 2003 and 2006, the tunnel oxide increased in thickness from 2.5 nm, through 3.5 nm to 5 nm. What this does of course is to increase program and erase voltages from the classic SONOS values towards 20 V. Also, the thicker stack results in more limited program/erase cycling endurance and retention after cycling challenges. A second technique is to replace the tunnel oxide by a stack of oxide and silicon nitride layers. This also reduces the effect of pass disturbs but does not eliminate it. The thicker gate stack leads to higher program and erase voltages than classic SONOS and retention after cycling challenges.

The vertical 3-D NAND and the VG-NAND approaches have this linkage between the structure of the gate dielectric stack and the pass disturbs and tend to have program/erase voltages approaching 20V and limited retention after extended endurance cycling. Both approaches have two times the gate dielectric stack thickness within the lateral scaling dimension. Therefore, lateral shrinking to increase capacity will involve re-engineering the gate dielectric stack and affecting pass disturbs and retention after cycling.

The Schiltron dual-gate architecture eliminates pass disturbs and allows the use of classic SONOS gate dielectric thicknesses in a series string of Flash cells allowing SONOS advantages (low program and erase voltages, extended endurance cycling capabilities) to be used in such strings. This decoupling effect of ONO engineering from pass disturbs allows additional benefits such as:

- Ability to overdrive pass gates to minimize worst case string currents;
- No program/erase pattern sensitivity on string currents and programmability;

- No difference between best case and worst case string currents with both being larger than other TFT SONOS NAND approaches;
- Complete absence of any over-programming issue in single level cell version.

Disturbs in any RRAM array have not been extensively published in peer reviewed literature as far as the author knows. The magnitude of these will no doubt depend on the actual mechanism that allows switching. Their relative immaturity leaves a question mark that still needs to be addressed.

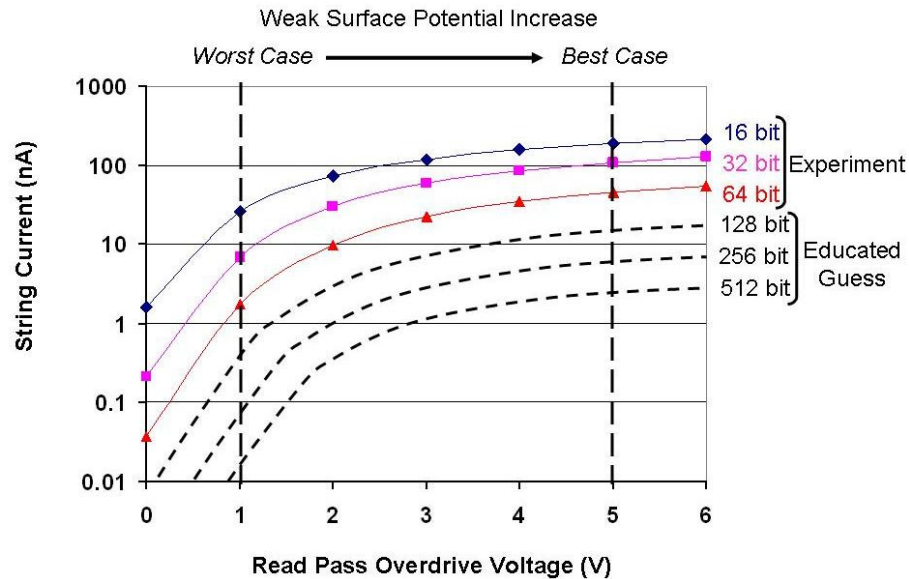
### SENSE CURRENTS

The minimum amount of current coming out of a cell in the low impedance state must be distinguishable from a cell in the high impedance state. Also, this current needs to be large enough to charge circuit nodes in a certain amount of time to be able to switch sense amplifiers. In MLC NAND, published worst case string currents are in the order of 100's of nA's. In RRAM, several variants show uA's of current.

In any approach that uses a Thin-Film Transistor (TFT), the current is in danger of being submerged by background noise since the carrier mobilities in TFTs is usually much less than in bulk silicon. This effect is exacerbated by placing the TFTs in a series string whether vertically or horizontally. Further worsening is achieved when all cells in the TFT string are in the high threshold voltage programmed state while the cell being read is in the erased state. This worst case string current is heavily dependent on the read-pass voltage applied over and above the highest programmed threshold voltage.



The worst case string current that is expected from such series strings of TFT cells has been published in a previous white paper<sup>3</sup> and is reproduced here in Fig.1 and Table 2. Note the importance of read pass overdrive voltage in trying to maximize the worst case string current. As has been said before, this voltage is strongly linked to the read pass disturb.



**Figure 1: String currents with a polycrystalline channel measured as a function of read pass overdrive voltage for different string lengths. The worst case and best case conditions are marked using a 1 volt and a 5 volt read pass overdrive condition respectively. “Educated guesses” for the longer strings are given.**

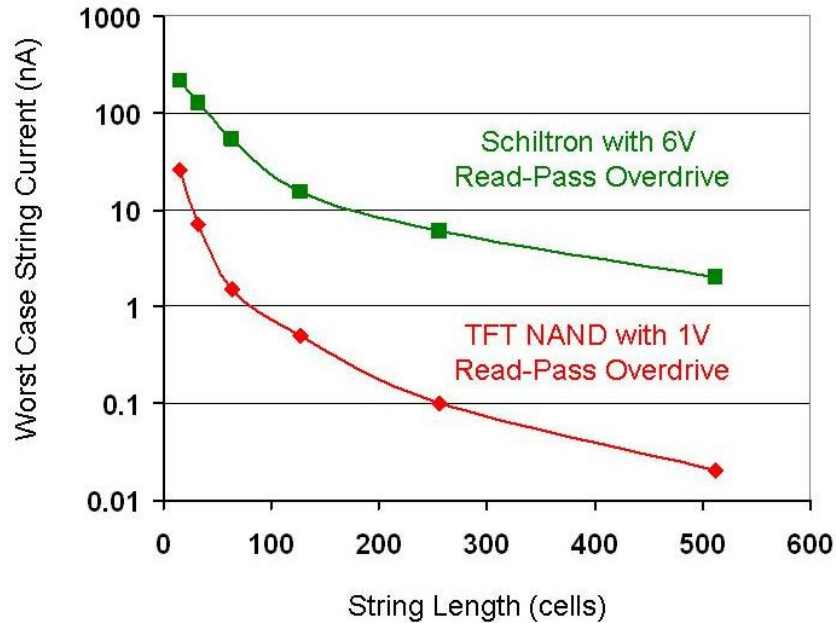
In the Schiltron case, the worst case string current can be maximized by overdriving the pass gates. This allows the best case string currents to be achieved even in the worst case condition.

<sup>3</sup> [http://www.schiltron.com/PDF\\_files/Vanishing\\_String\\_Currents\\_White\\_Paper\\_2011.pdf](http://www.schiltron.com/PDF_files/Vanishing_String_Currents_White_Paper_2011.pdf)

String Length (cells)	Worst Case String Current at 1 volt read pass overdrive (nA)	Worst Case String Current at 2 volt read pass overdrive (nA)	Worst Case String Current at 3 volt read pass overdrive (nA)
16	26	75	120
32	7	30	60
64	1.5	10	20
128	0.5	3	7
256	0.1	1	3
512	0.02	0.5	1

**TABLE 2: Data and educated guesses for the worst case string current as a function of string length and read pass overdrive voltage. Data in blue, educated guess in red.**

Figure 2 shows the worst case string currents for any TFT NAND approach with 1V as read-pass overdrive voltage (this then would be 7V read-pass voltage if the programmed threshold voltage is 6V) compared to the Schiltron approach where we can apply a 6V read-pass overdrive voltage (this would be 7V if the non-memory access device threshold voltage were 1V). The Schiltron multiplier is an order of magnitude at short string lengths and two orders at long string lengths.



**Figure 2: Worst case string currents compared between Schiltron and any other TFT NAND approach as a function of string length with 6V read-pass overdrive in Schiltron and 1V in TFT NAND. The 6V in Schiltron can be applied without pass disturbs.**

PROGRAM/ERASE VOLTAGES

RRAM approaches tend to have the lowest program and erase voltages at below 5V (although reset currents from low impedance to high impedance tend to be in the uA range).

Schiltron can take advantage of SONOS type tunneling voltages that can reach below 10V (at very low tunneling currents).

All other TFT NAND approaches (vertical and horizontal) tend towards 20V for program and erase because of the need to thicken up the gate dielectric stack for pass disturb reduction. Such voltages may cause scaling problems due to breakdowns.



In addition, erasing cells within a TFT NAND string may be problematic in some cases since there is no contact to the channel. This challenge is avoided in Schiltron's case as has been explained in a previous article<sup>4</sup>

### DESIGN MATURITY

This metric measures the ability of the technology to take advantage of existing design know-how and infrastructure. For example, much work has been done over the last 20 years to perfect techniques for program and erase and program inhibit in NAND. Also, various program-read-verify schemes have evolved to shape threshold voltage distributions in NAND especially for multi-bit per cell storage.

All 3-D TFT NAND approaches and the Schiltron technology can take advantage of this large amount of knowledge.

RRAM on the other hand would cause most of these techniques to be jettisoned in any transition from the classic NAND approach to an RRAM approach. This is not to say that it cannot be done but such a large step away from established knowledge always entails risk.

## **III. SUMMARY**

Table 3 summarizes the comparisons between the various NAND replacement contenders using the attributes introduced and discussed above.

It is clear that the Schiltron dual-gate approach is the only one that simultaneously optimizes all attributes.

---

<sup>4</sup> [http://www.schiltron.com/PDF\\_files/Erase\\_2010\\_Schiltron.pdf](http://www.schiltron.com/PDF_files/Erase_2010_Schiltron.pdf)



Needless to say, a transition away from a tried and trusted technology like classic Floating Gate NAND will always be fraught with risk. Therefore, choosing a technology that is evolutionary and satisfies key attributes is the path most likely to lead to success.



<i>Attribute</i>	<i>Vertical 3-D NAND</i>	<i>VG - NAND</i>	<i>RRAM</i>	<i>Schiltron</i>	<i>Comments</i>
<b>3-D Stackability</b>			?		<i>RRAM needs integrated select device or needs robust inherent select device</i>
<b>Scalability</b>					<i>Schiltron's dual gate architecture ideal for electrostatic control</i>
<b>Manufacturability</b>			?		<i>Schiltron uses existing materials and fab infrastructure</i>
<b>Retention after Cycling</b>	?	?	?		<i>10 years at 85 C after 100,000 cycles already shown in Schiltron's case</i>
<b>Disturbs</b>	?	?	?		<i>Schiltron with ~zero read- and program-pass disturbs</i>
<b>Sense Current</b>	?	?			<i>Schiltron worst case string current &gt;~ 10 - 100 times greater than any other vertical and VG-NAND</i>
<b>Program/Erase Voltages</b>	?	?			<i>Schiltron approach allows low SONOS voltages for program/erase since ONO optimization independent of pass disturbs</i>
<b>Design Maturity</b>			?		<i>Schiltron uses existing design approaches for program/erase/read</i>

**TABLE 3: How the different technologies compare. The Schiltron dual-gate approach is the one that optimizes all attributes simultaneously.**