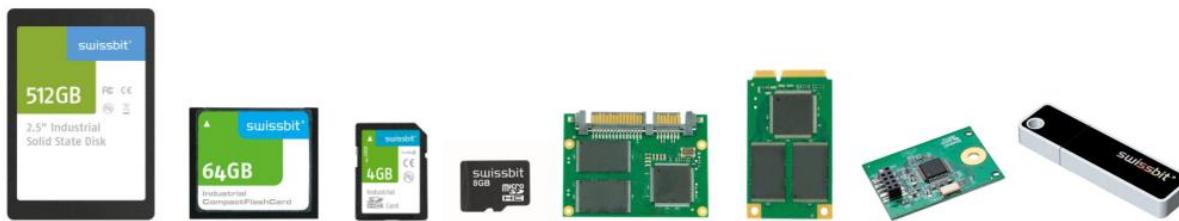


I. SSD ARCHITECTURE

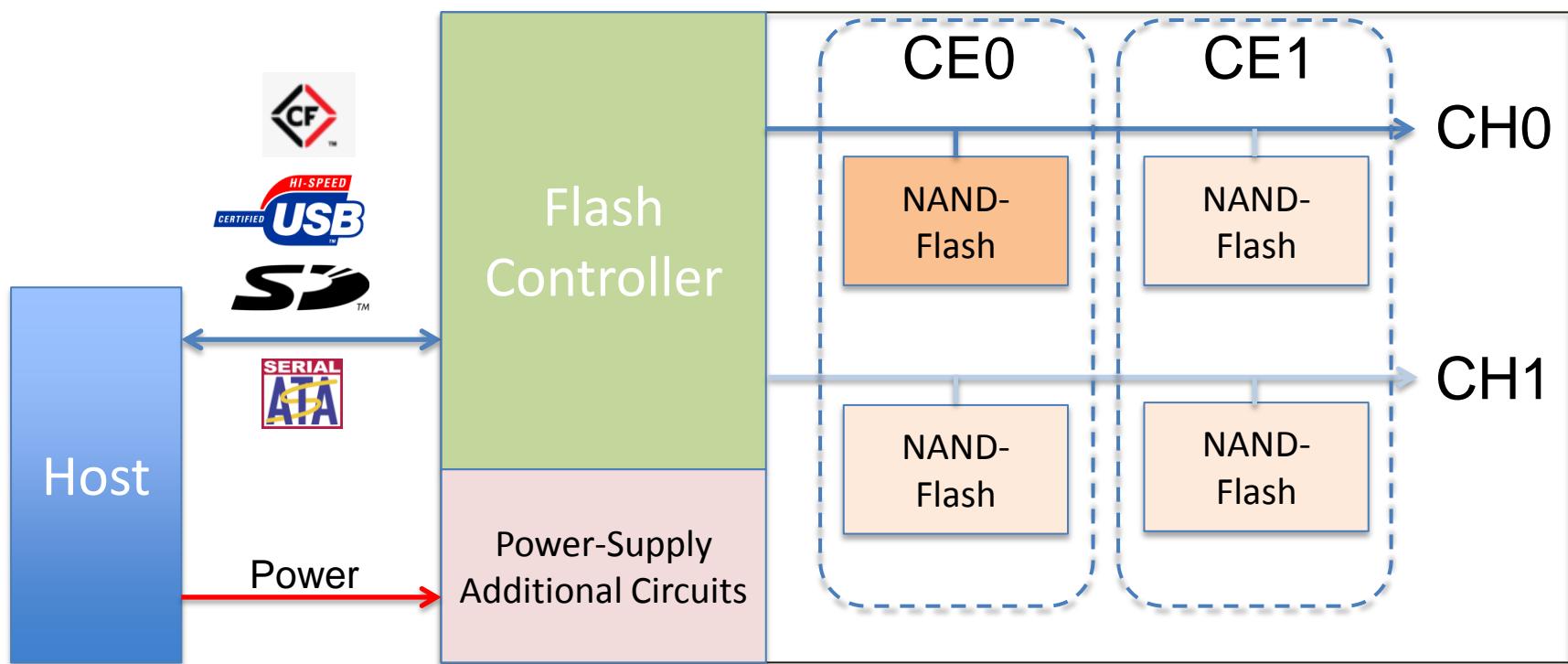
PART 1, INTRODUCTION



SSD ARCHITECTURE

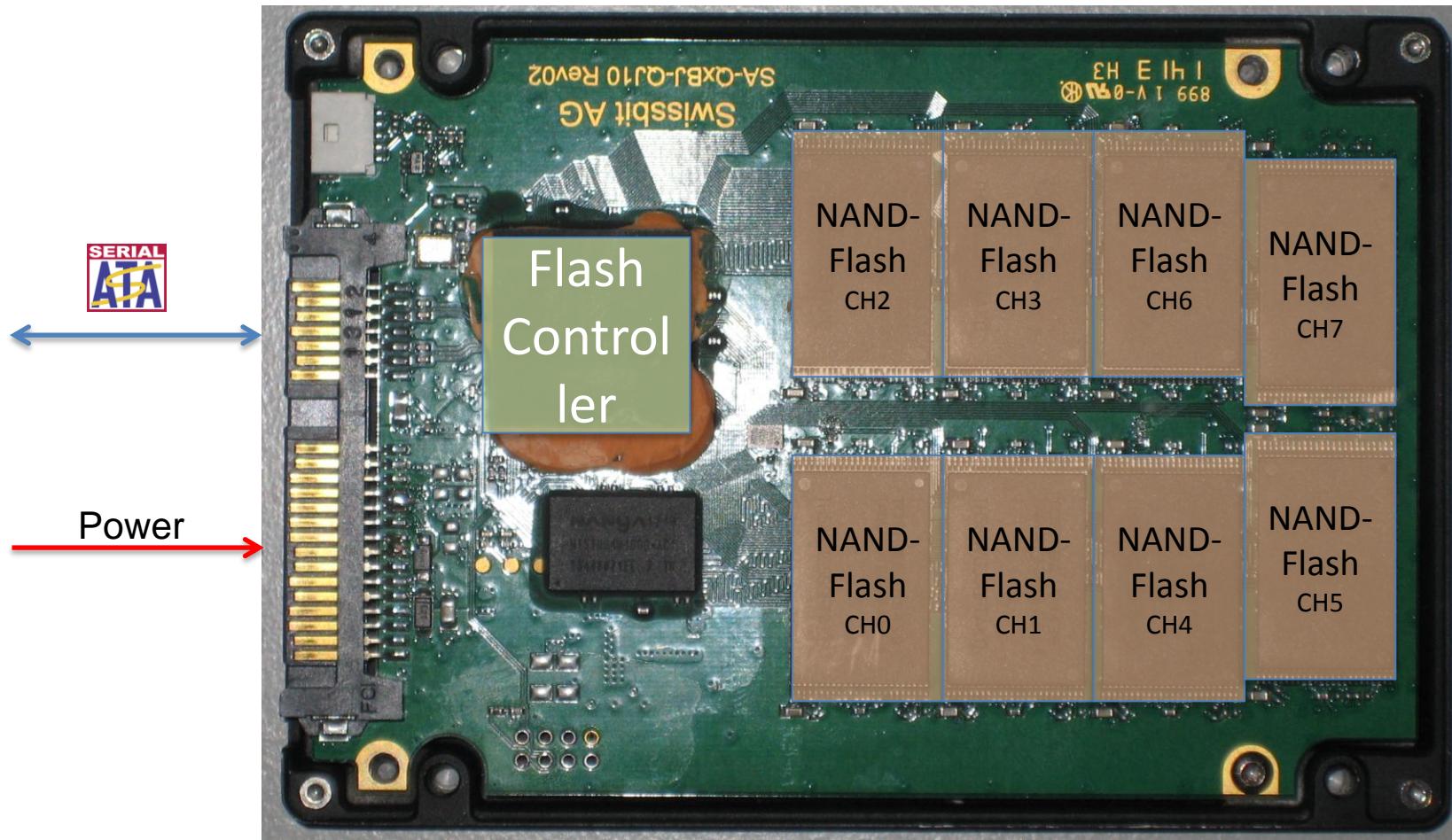


CH: Channel
CE: Chip Enable



SSD ARCHITECTURE

X-500/X-55 SSD

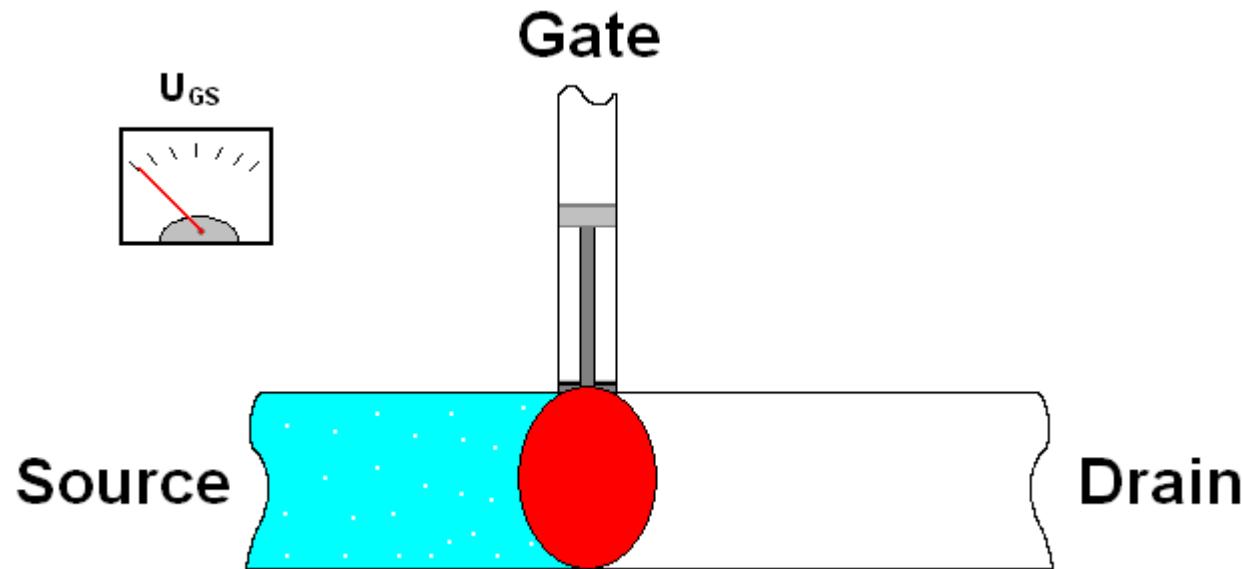


II. MEDIA NAND FLASH, WHERE IS IT AND WHERE IS IT GOING?

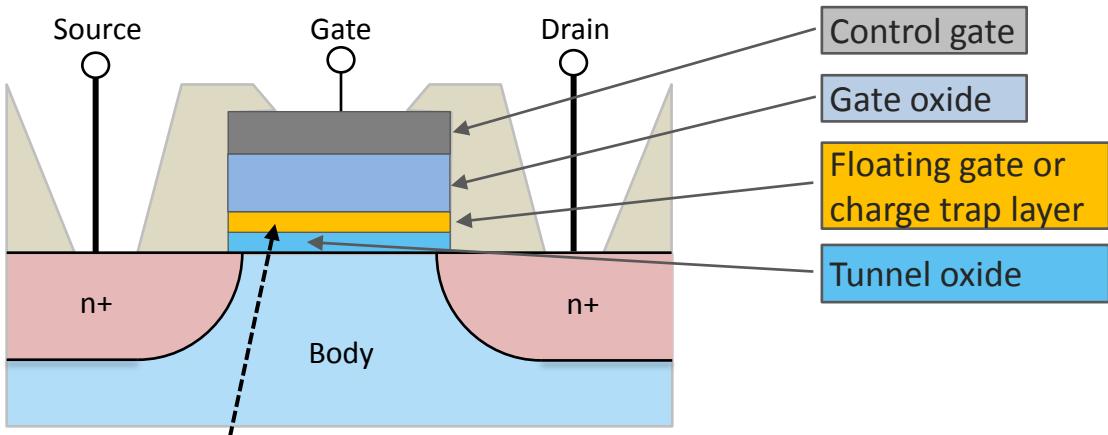


NAND CELL

FET TRANSISTOR



NAND CELL



Polysilicon, conductive,
total discharge through pin holes



Layered Silicon Nitride, non
conductive, partial discharge
through pin holes

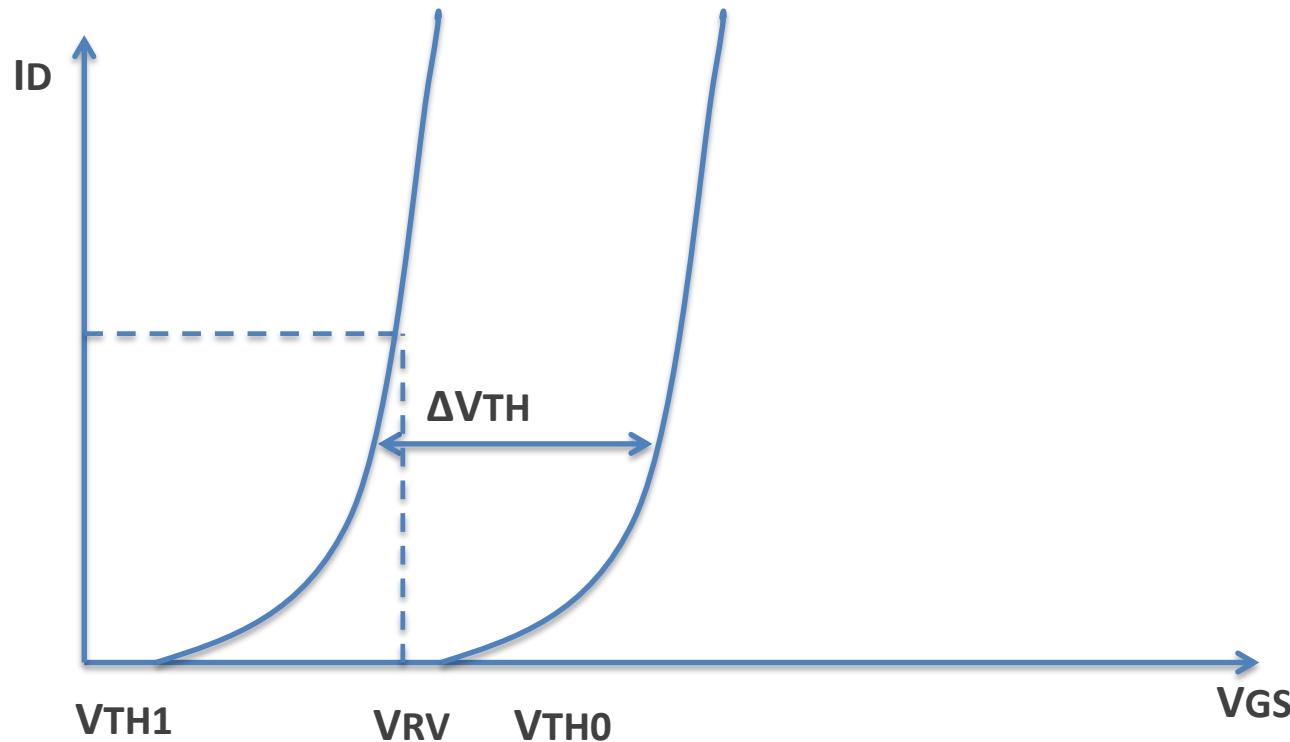
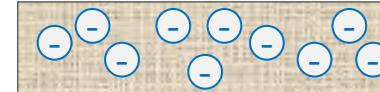
NAND flash uses a control gate to switch on or off a pass transistor and a floating gate to store information by different levels of charge

NAND CELL

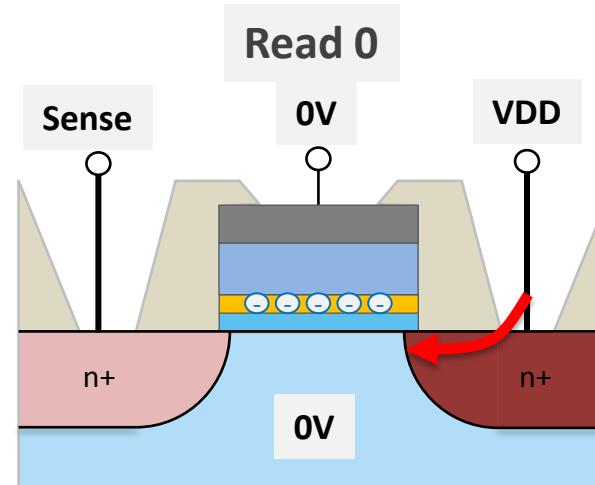
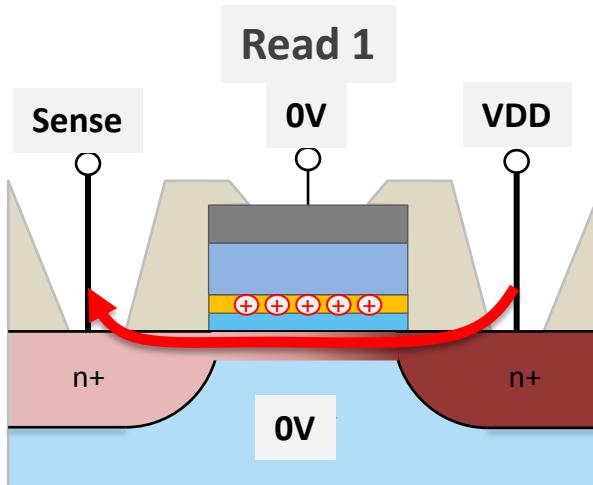
Erased = „1“



Programmed = „0“

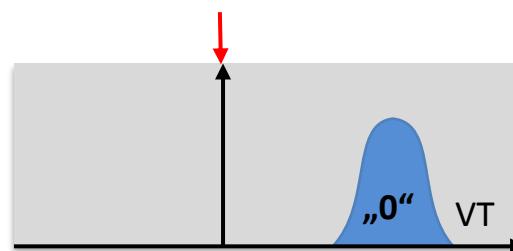
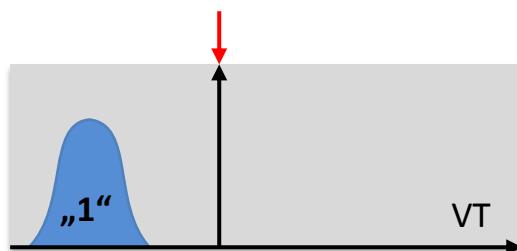


NAND CELL OPERATION



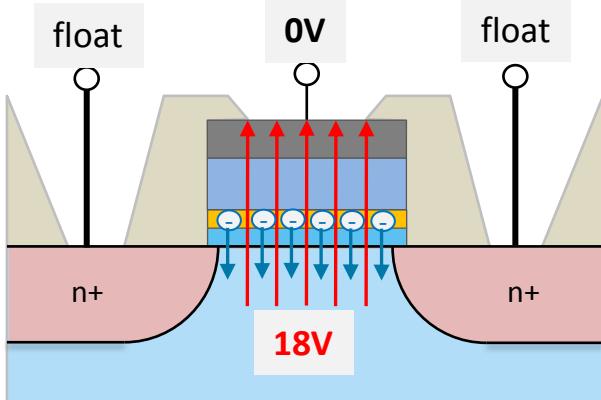
negative VT @ 0V ->
Channel conducting ->
Read VDD = „1“

positive VT @ 0V ->
Channel isolated ->
Read „0“

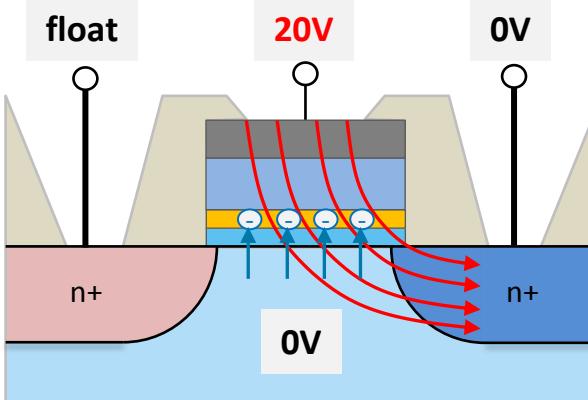


NAND CELL OPERATION

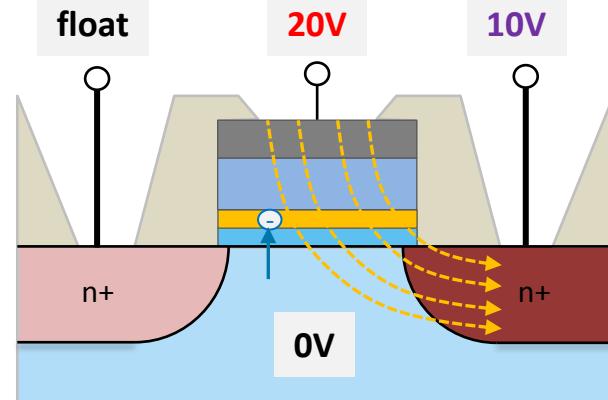
erase „1“



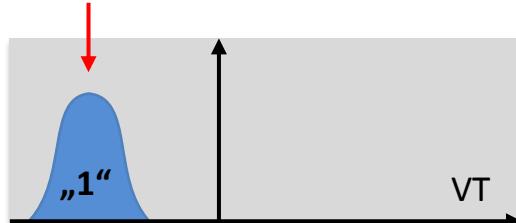
write „0“



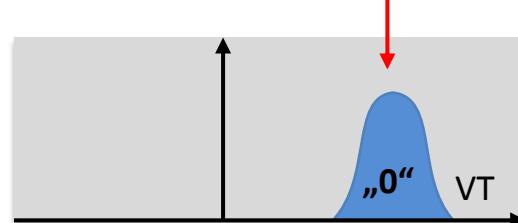
write „1“



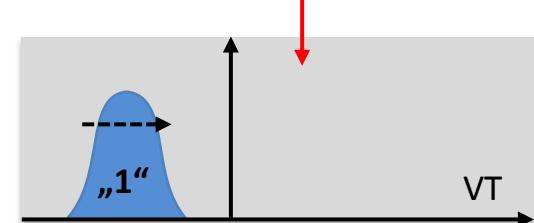
Positive charge (holes) on storage gate, negative VT
Channel conducting



Negative charge (electrons) on storage gate, positive VT
Channel isolated



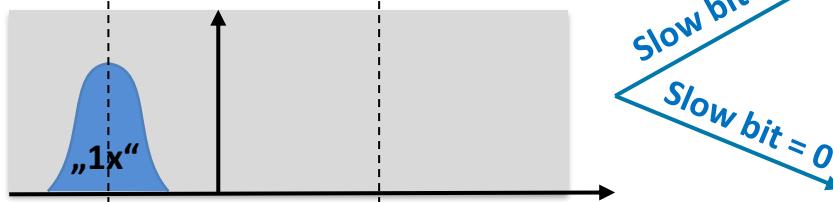
Electrical field not sufficient to inject electrons, negative VT
Channel conducting



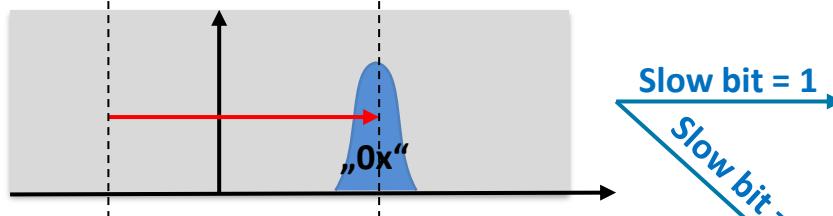
“FAST PAGE” ACCESS

Fast (strong) page programming only uses 2 of the 4 levels
 $1x$ = erased, $0x$ = programmed

Fast bit = 1

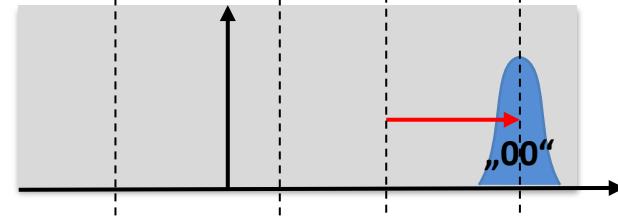
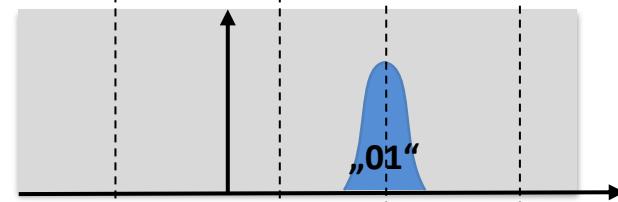
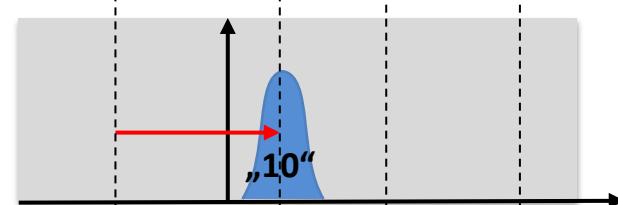
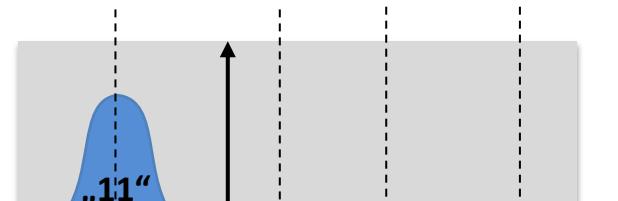


Fast bit = 0



Usage of fast page has higher reliability and faster access
 Fast page must be written first

Slow (weak) page programming modifies these levels by adding additional charge
 $x1$ = no programming, $x0$ = added charge



NAND CELL

Number of electrons

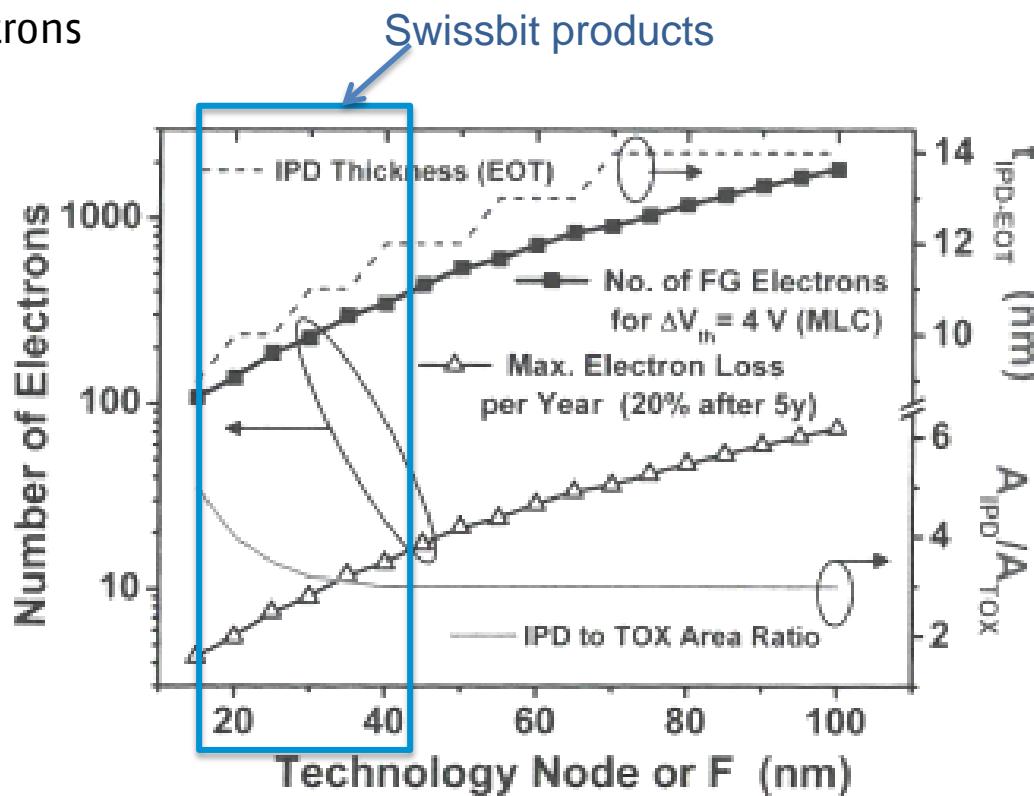


Fig. 5.25 Number of electrons as a function of the technology node F. To have the programming voltages remain similar over different technology generations, the gate coupling ratio is optimized by means of an IPD EOT reduction and an increase of the CG-FG to TOX area ratio

FLASH ARRAY

We use NAND-Flash

NICHT	
A	Y
1	0
0	1

UND		
A	B	Y
1	1	1
1	0	0
0	1	0
0	0	0

NAND		
A	B	Y
1	1	0
1	0	1
0	1	1
0	0	1

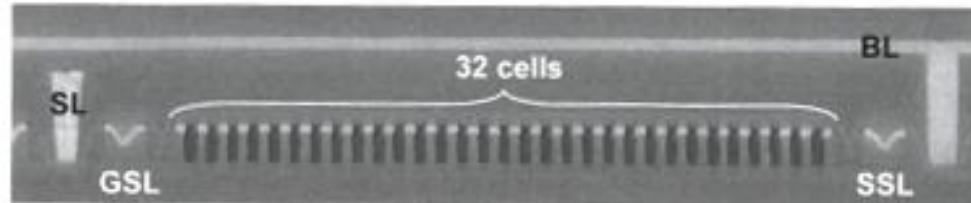
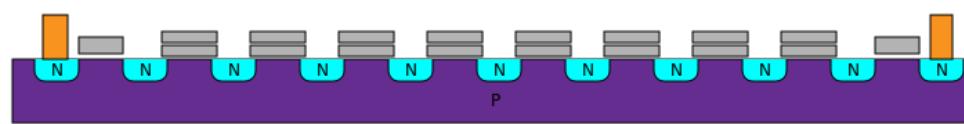
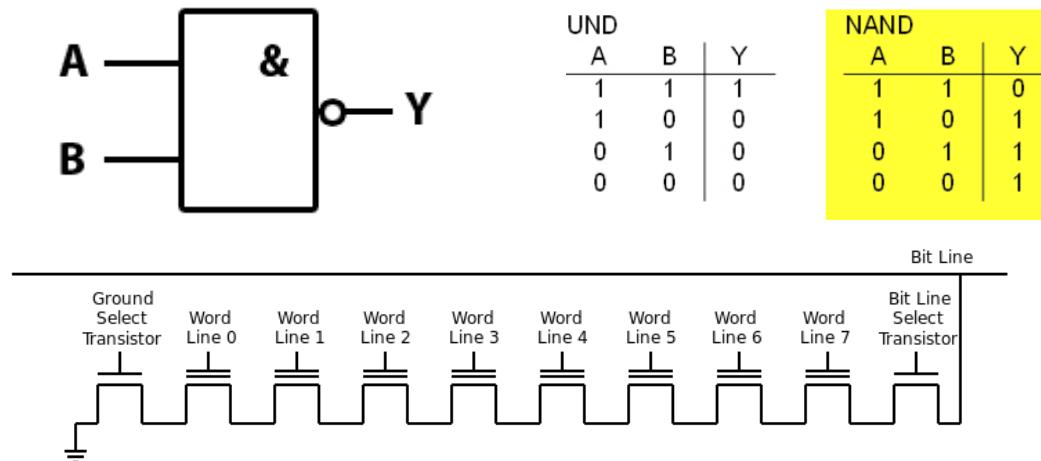
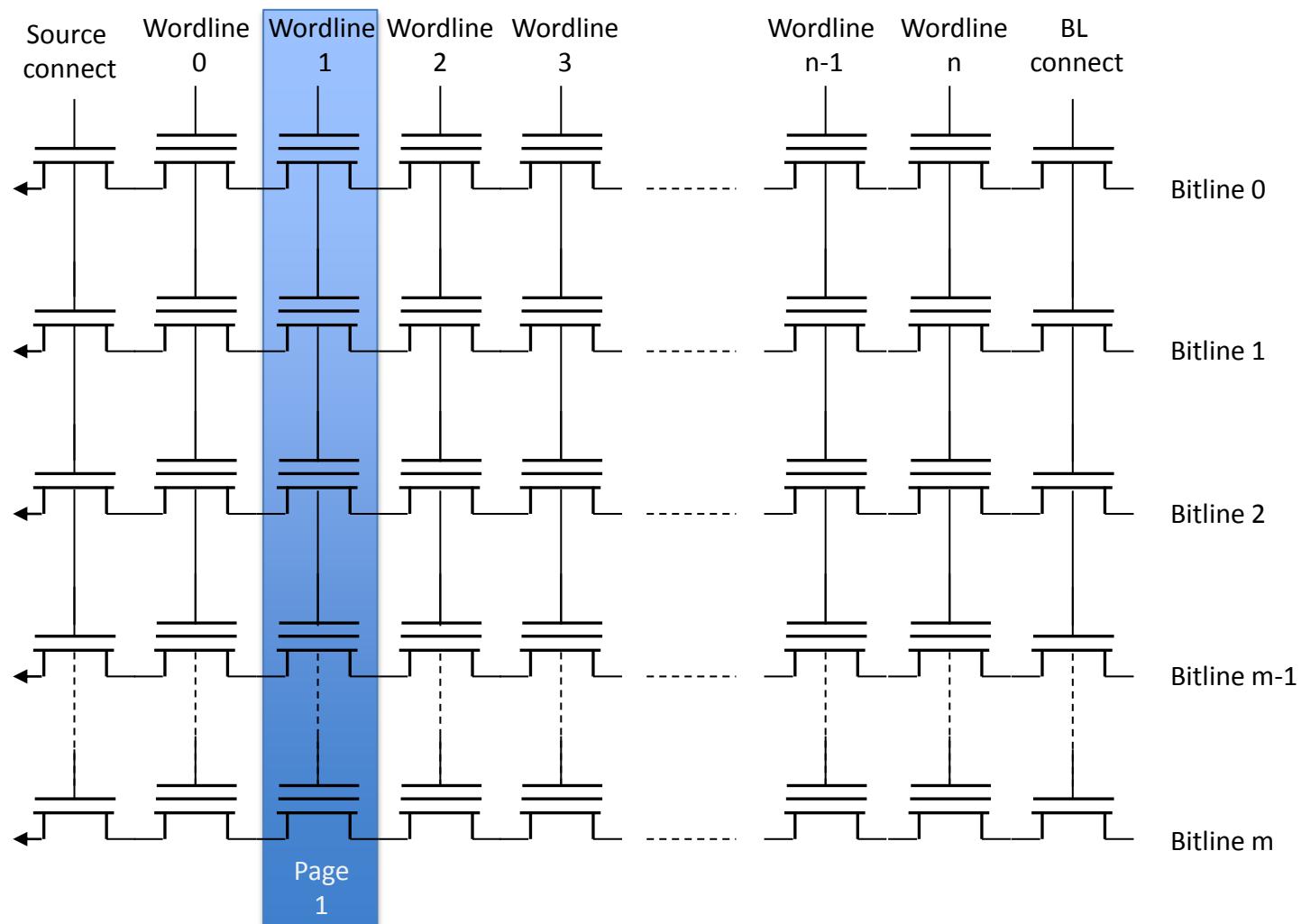
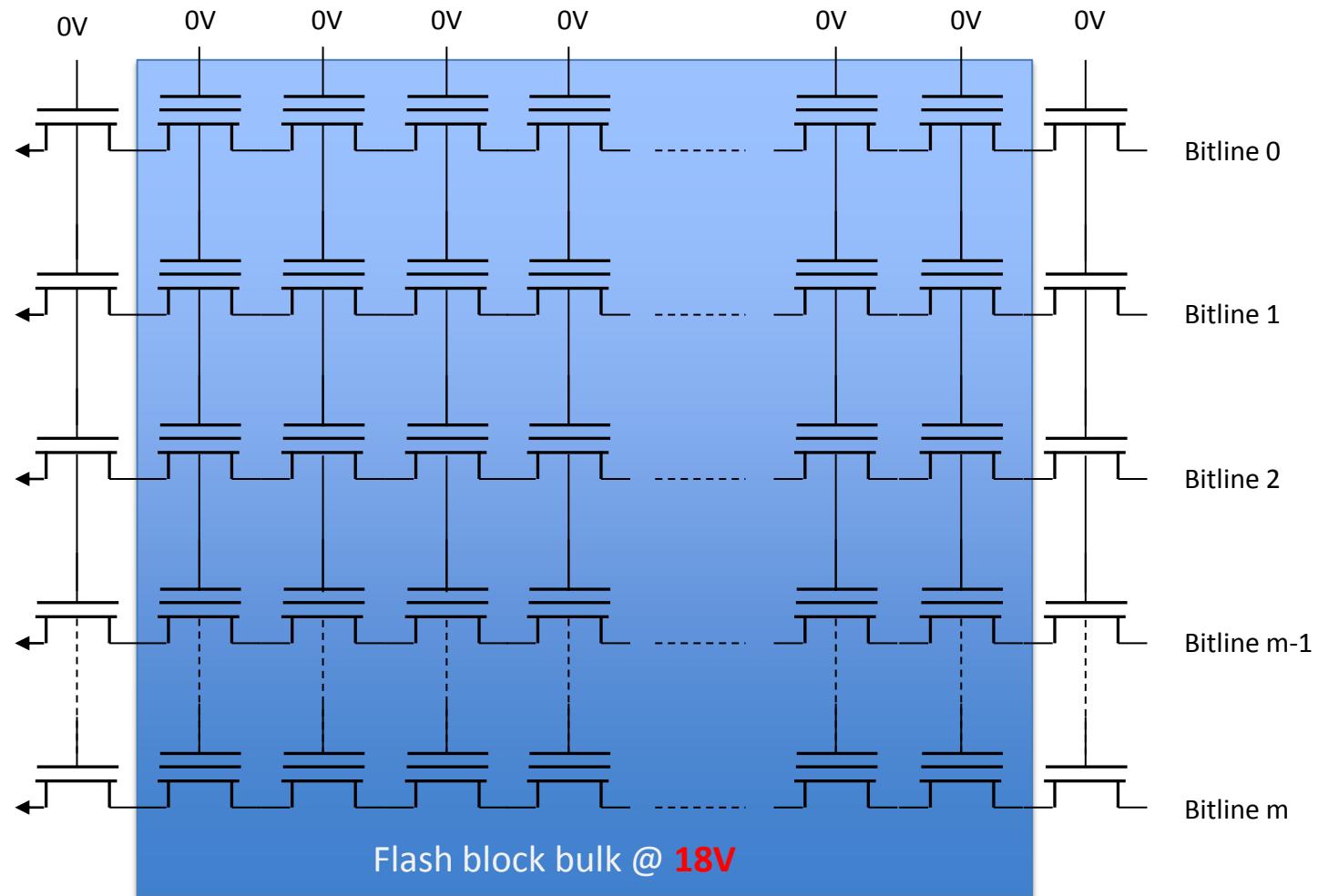


Fig. 5.2 SEM picture of a NAND string with 32 cells per string in a 48 nm floating gate NAND technology [2]

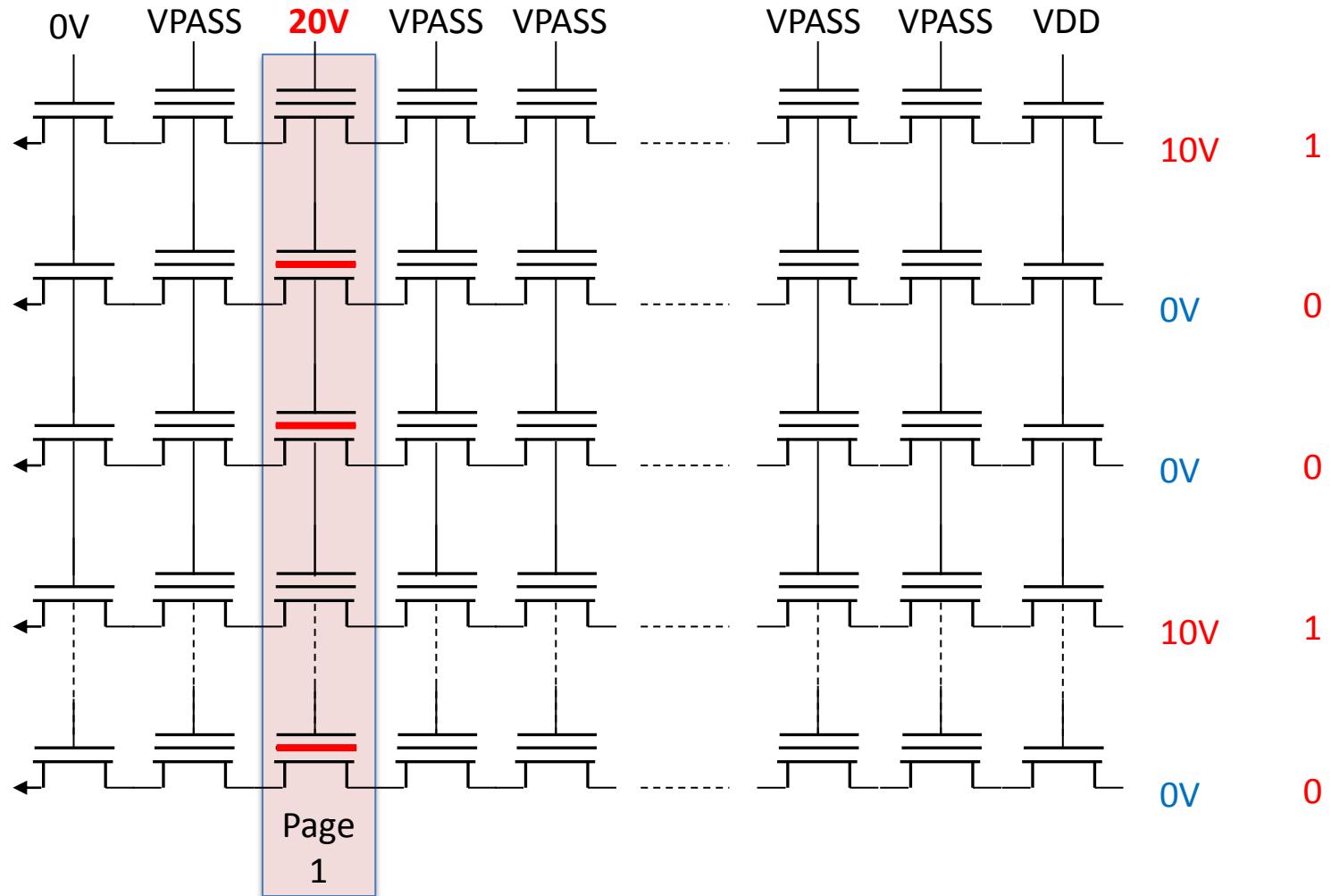
FLASH ARRAY



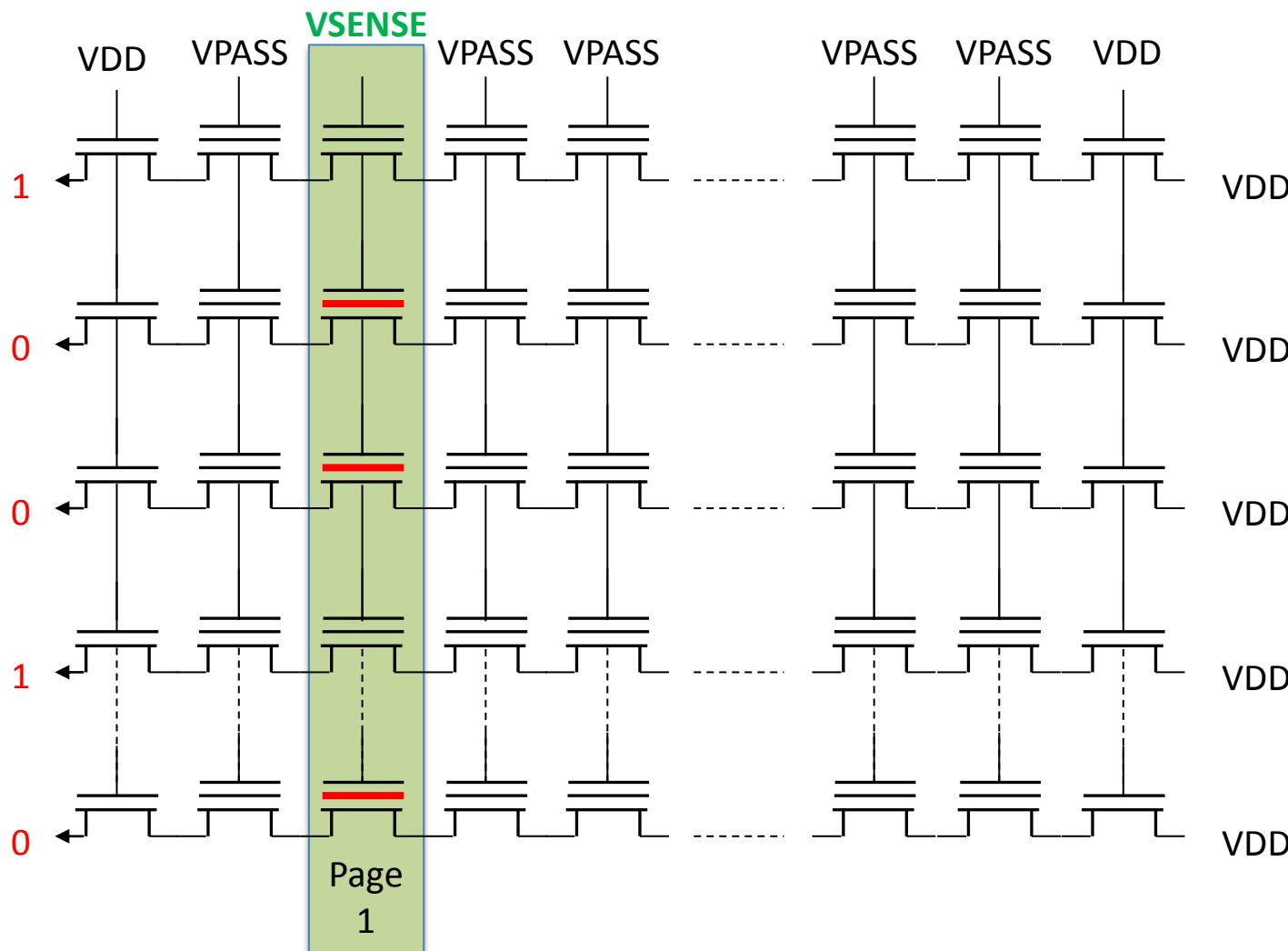
FLASH ARRAY ERASE



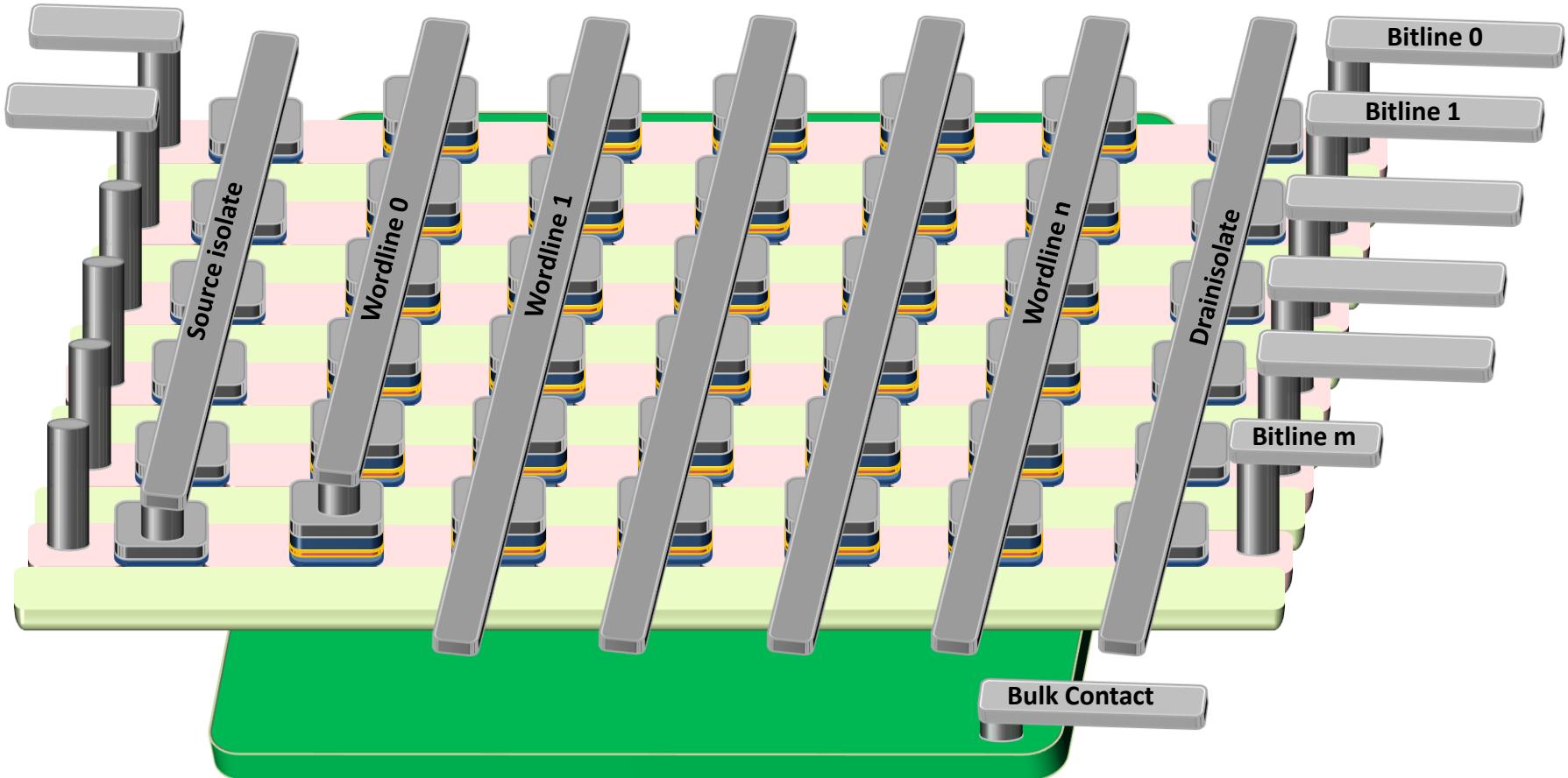
FLASH ARRAY PROGRAM



FLASH ARRAY READ



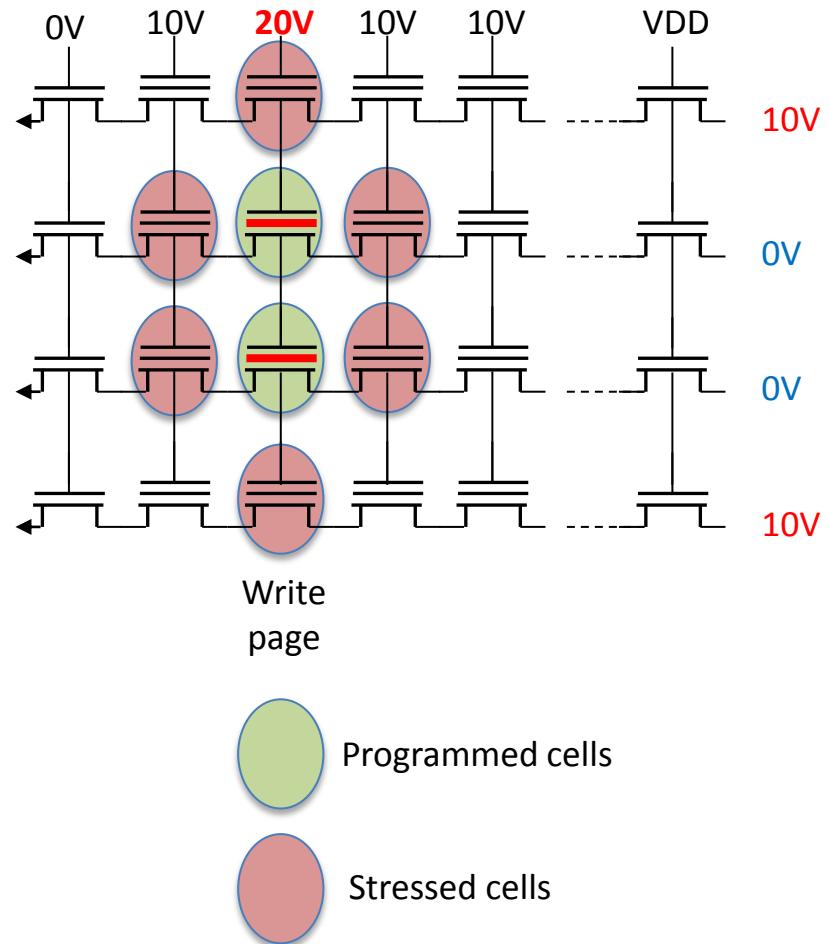
FLASH ARRAY



- Flash erase happens through the bulk
- For high density routing the bulk is shared for a full block
- The minimum erasable unit is one block

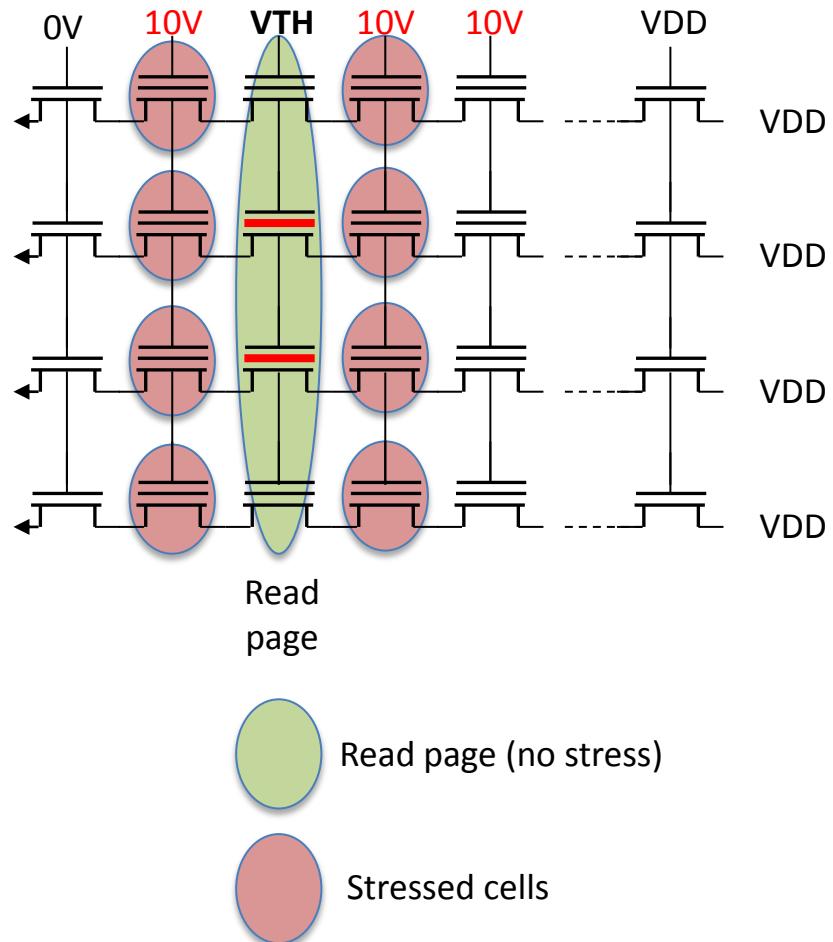
FLASH ISSUES: PROGRAM DISTURB

- Cells not being programmed still see elevated voltage (~10V V_{gate-drain})
- Stress occurs only in same block as programmed cells
- Stress occurs
 - in pass through pages
 - in page under programming for cells that should remain erased
- Unintentional charge accumulation in floating gate shifts VT to higher level which can cause future read errors.
- Worse case stress on low number pages
- Read fails will be handled by ECC
- Block will recover after block erase



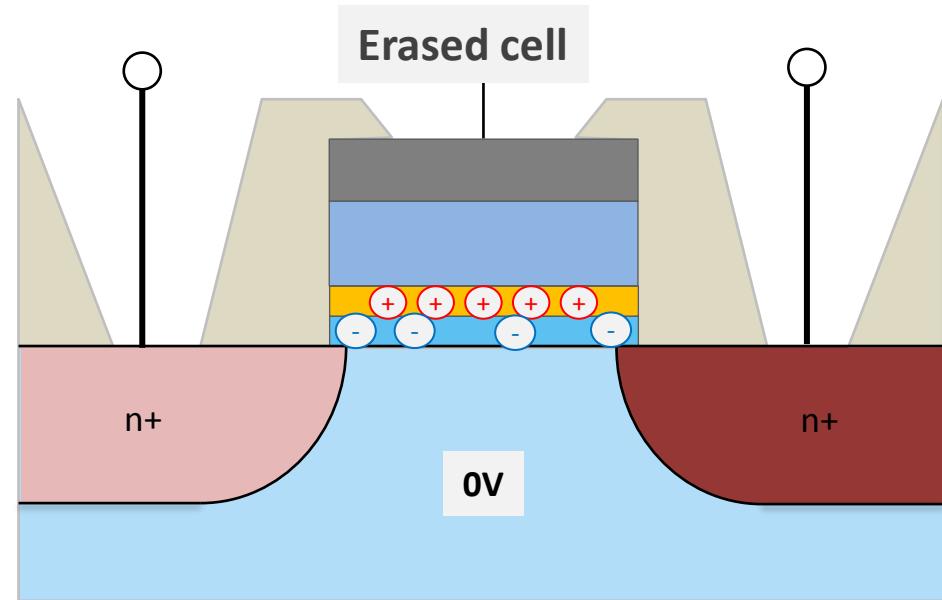
FLASH ISSUES: READ DISTURB

- Pages that are not read are in transparent mode and see high select gate voltage
- Stress occurs only in same block as read page
- Stress occurs in pass through pages
- Unintentional charge accumulation in floating gate shifts VT to higher level which can cause future read errors .
- Effect less critical than program disturb due to reduced voltage
- Cell charge shift happens mostly in read only applications with intense read to few blocks
- Read fails will be handled by ECC
- Block will recover after block erase

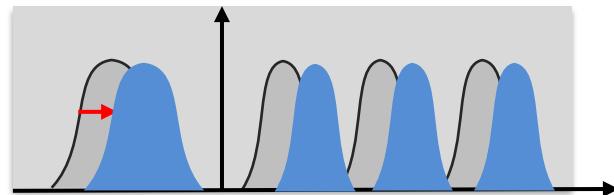


FLASH ISSUES: ENDURANCE

- Multiple programming of the cell injects hot electrons in gate oxide instead of floating gate
- These electrons are trapped permanently and cause a VT shift of this cell
- Read fails will increase over time
- Block erase cannot recover the fails, damage is permanent
- Block needs to be marked as „Bad block“

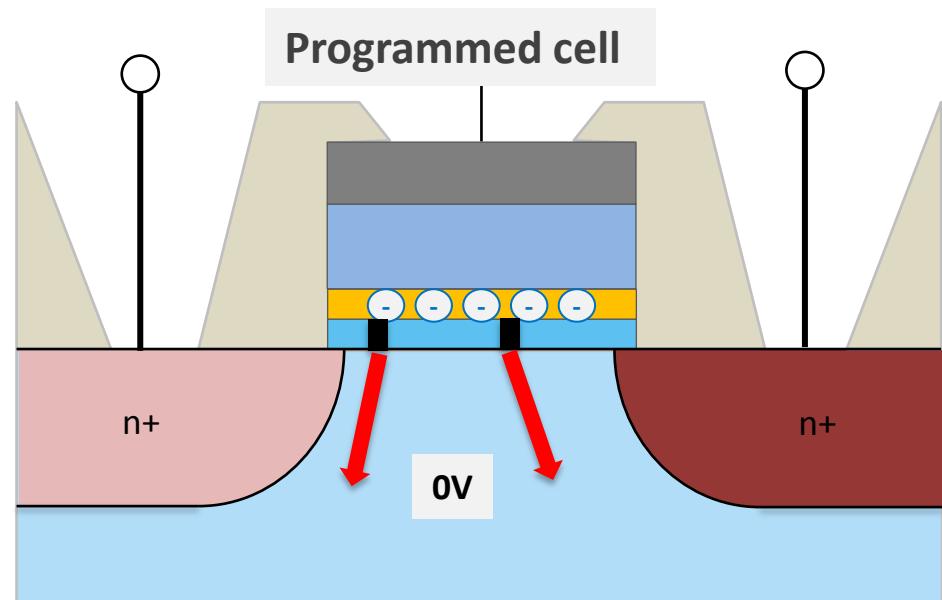


VT of each bit coding shifts to higher levels

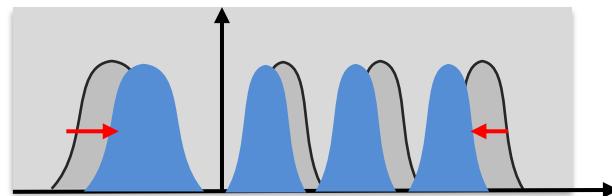


FLASH ISSUES: RETENTION

- Multiple programming of the cell cause degradation of the tunnel gate oxide and create conductive paths from the storage gate to substrate
- Stored electrons will dissipate into substrate, stored holes will be filled
- Effect is accelerated at high temp
- Fail mechanism mostly at end of allowed P/E cycles
- Block needs to be marked as „Bad block“ if retention fails intensify
- Charge trap technology improves retention due to non-conductive storage cell



VT of each bit coding shifts towards GND



FLASH ISSUES: RETENTION

PE cycles	0	300	1,000	3,000	10,000	30,000	100,000
EM-MLC 24nm	5	5	5	5	1	0.25	-
standard MLC 24nm	10	10	3	1	-	-	-
SLC 24nm	10	10	10	10	10	4	1
MLC 19nm	5	3	1	-	-	-	-

Table 3: Typical data retention (years) at 40°C

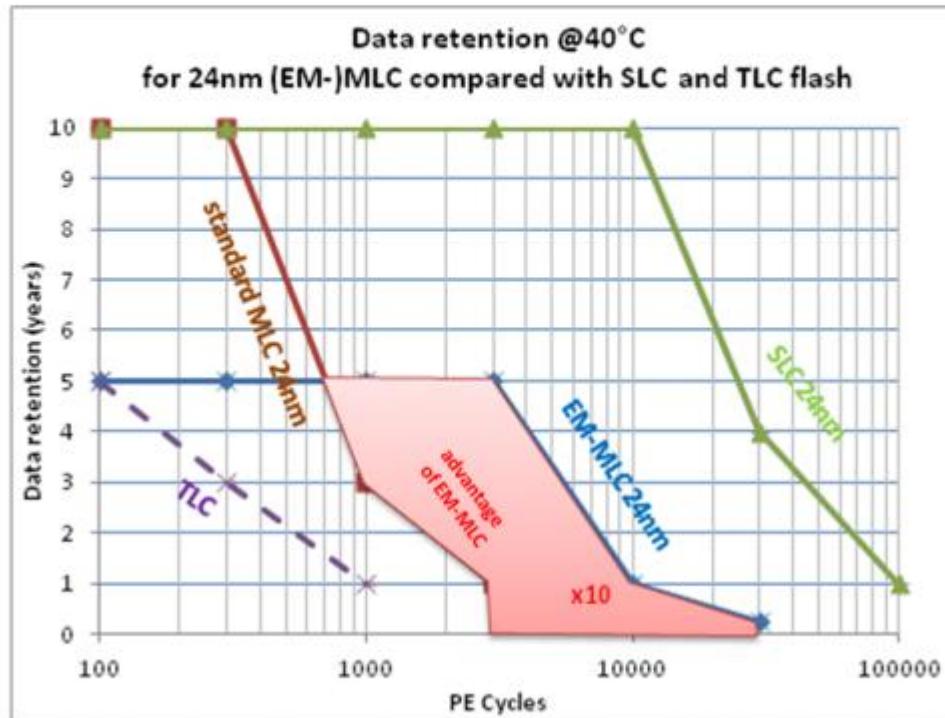


Figure 4: Typical data retention at 40°C

FLASH ISSUES: RETENTION

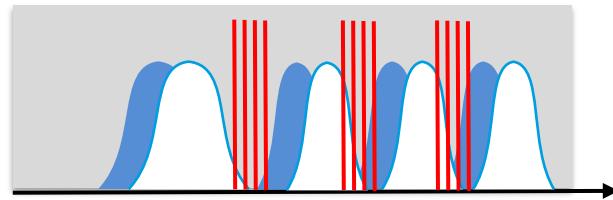
Temperature	40°C	55°C	70°C	85°C	100°C
Acceleration Factor	1	6.4	35	168	706
Specified retention	10 years	565 days	103 days	11 days	5 days

Table 2: Temperature dependence of the data retention, specified for 10 years at 40°C

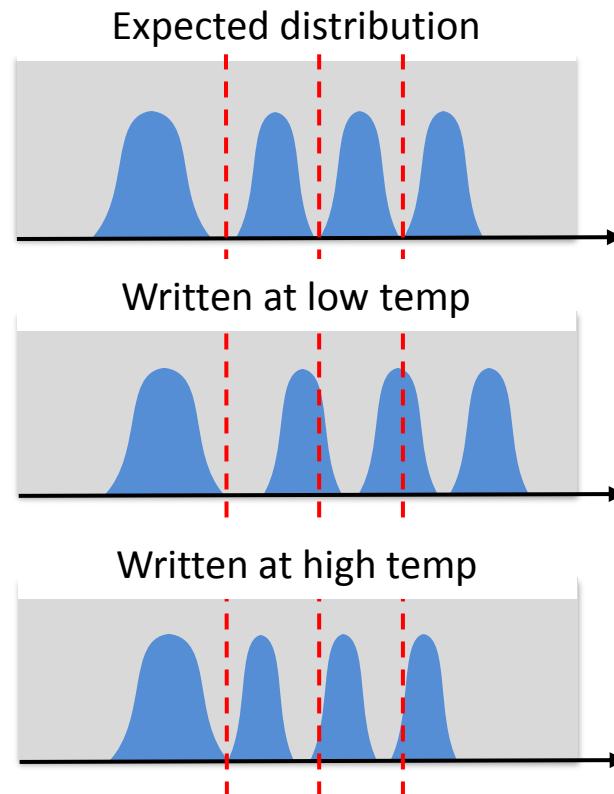
-> Retention is highly temperature dependent!

FLASH ISSUES: TEMPERATURE

- The charge written into the floating gate depends strongly on the temperature during the programming
- The sensing reference levels (select gate voltage) must be adjusted to the shifted charge levels of the page
- In case of uncorrectable ECC failure rate the voltage levels can be incremented in small steps until the re-read attempt delivers correctable information.
(Read Retry)



Read Retry



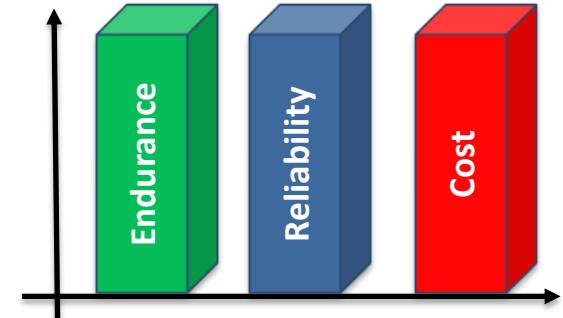
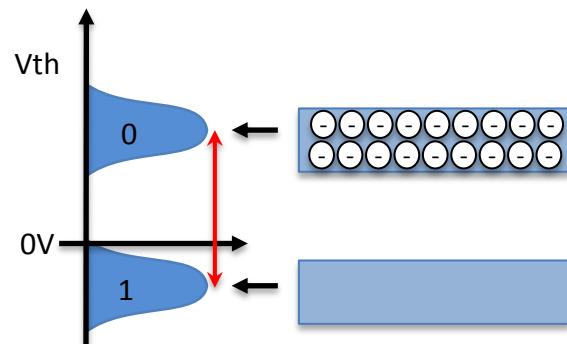
NAND FAILURE MODES

Mode	Description	Trigger Mechanism	Detection Method	Impact	Mitigation Strategy
Endurance	P/E failures can occur at the cell, page or block level	P/E cycle count approaches NAND EOL Spec.++ P (Bad Block) ↑	Poll NAND status register following a program or erase command	Once all Spare Blocks are consumed, drive enters EOL Read-Only mode	Usage Model & Service Life steer NAND technology selection
Data Retention	Static Data may change state over a period of time, errors caused by Charge loss or gain	High Read-only count, P/E cycles and Temperature	CRC & Error Correcting Code (ECC) via SSD Controller	UECC (risk of static data loss – i.e., code)	Background Media Scan, Overprovisioned ECC, Read Retry
Program/Read Disturb	Program/Read Operations disturb adjacent page data in the block, due to charge gain	After a large number of read cycles charge can build up causing a cell to be soft programmed to a different state.	CRC & Error Correcting Code (ECC) via SSD Controller	UECC (risk of static data loss – i.e., code)	Background Media Scan, Overprovisioned ECC, Read Retry
Early Life Block Failure	One or more blocks fail during Program, Erase or Read operations <10% of P/E	Random Particle contamination in the peripheral control circuits	Program/Erase detected via Status Register. Read Failures (UECC) are detectable but cannot be corrected	Program/Erase fails are mapped out (BB) and replaced with (SB) UECC (risk of static data loss – i.e., code)	Burn-In Test

NAND TECHNOLOGY COMPARISON

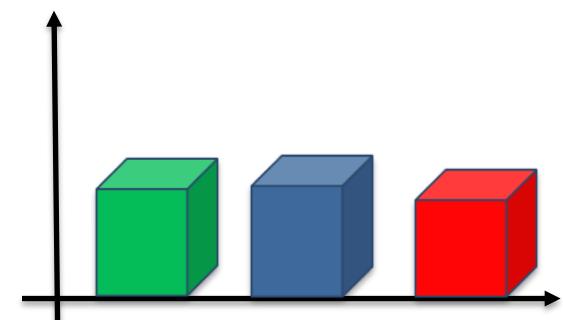
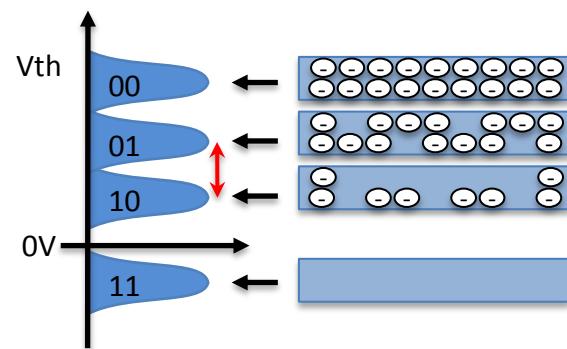
Single Level Cell (SLC)

P/E Cycles: 100K
 Data Retention: 1 year @ EOL
 Initial: 10 years
 RBER = 1×10^{-7}



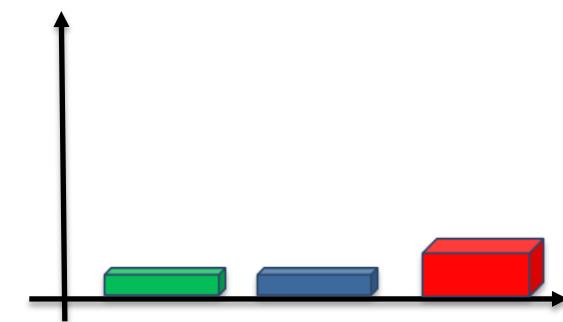
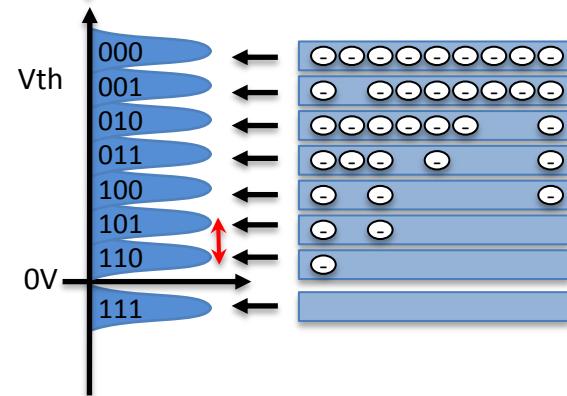
Multi Level Cell (MLC)

P/E Cycles: 3K
 Data Retention: 1 year @ EOL
 Initial: 10 years
 RBER = 1×10^{-6}



Triple Level Cell (TLC)

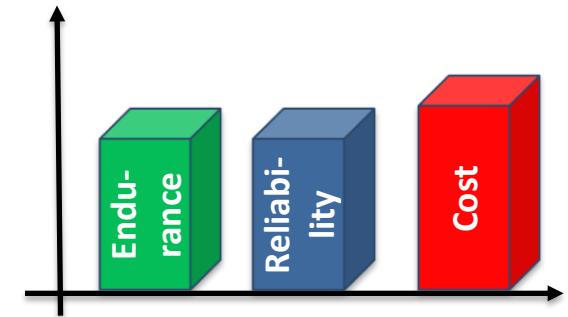
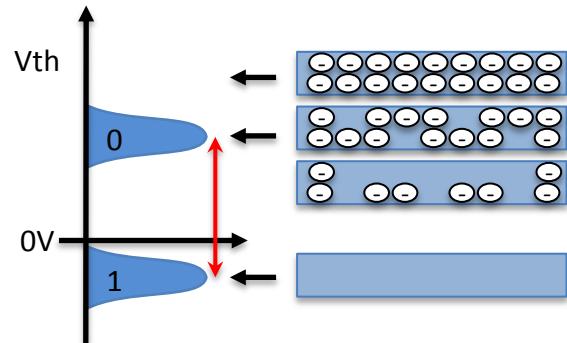
P/E Cycles: 500
 Data Retention: 3 mnt @ EOL
 Initial: 1 years
 RBER = 1×10^{-3}



NAND TECHNOLOGY COMPARISON

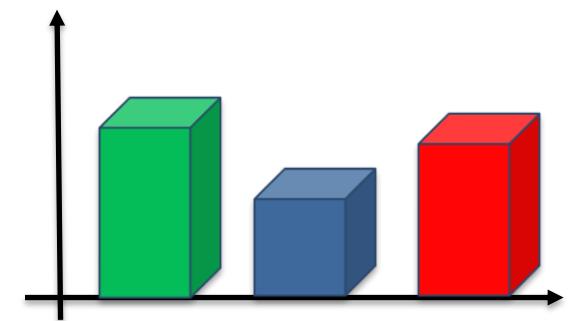
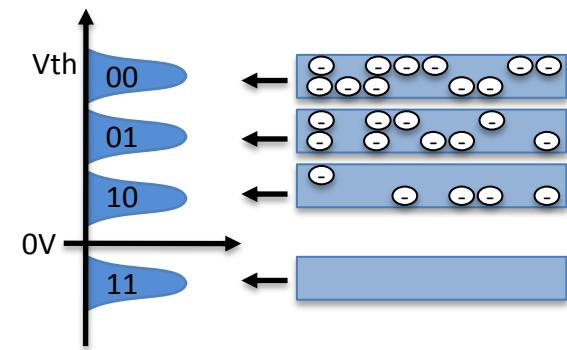
Pseudo SLC

P/E Cycles: 50K
 Data Retention: 1 year @ EOL
 Initial: 10 years
 RBER = 1×10^{-6}



Enhanced/Enterprise Multi Level Cell (eMLC)

P/E Cycles: 30K
 Data Retention: 3 mnt @ EOL
 Initial: 5 years
 RBER = 1×10^{-6}



3D NAND

What is 3D NAND?

3D (aka BiCS, aka V-NAND) Standard NAND flash bit string is arranged vertically vs. in the plane

What is the promise of 3D NAND?

Enhance overall drive capacity, improved reliability and performance

Who are key players in 3D NAND technology?

Samsung, Toshiba, and Intel/Micron (IMFT)

When will we see it in the market?

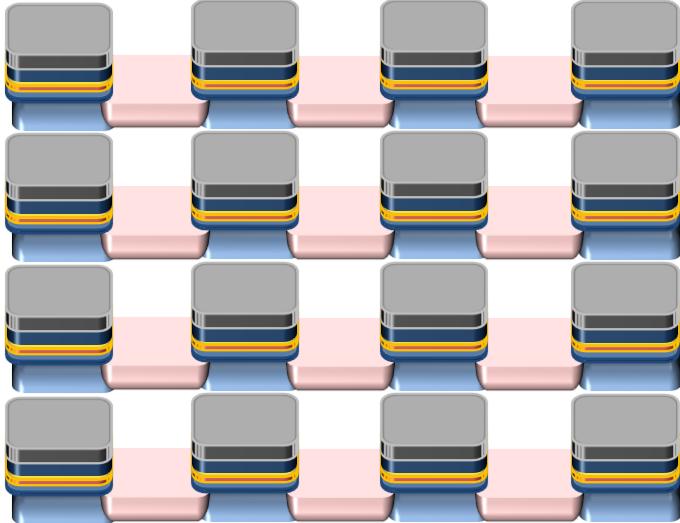
Samsung shipping V-NAND SSD's Now

Toshiba 128Gb (BiCS) eMLC 2015, Q4

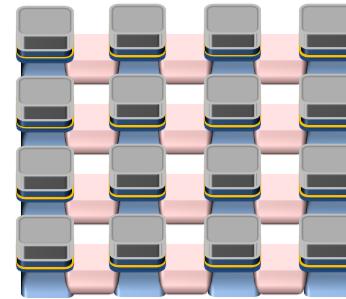
Intel/Micron (IMFT) 3D NAND CY2015, 2H

Type	15nm MLC	BiCS
Generation	15nm	Gen2
Interface	Toggle 2.0	Toggle 2.0
Die Density	128Gb	256Gb (32 stack) 384Gb (48 stack)
P/E Cycle	3K	3K+
Data Retention	1 Yr (@ 40C, EOL)	1 Yr (@ 40C, EOL)
ECC	40b/1KB	40b/1KB +
Page Size	16KB + 1376B	16KB + 1280B
Block Size	4MB	6MB
Schedule	2015/Q1	2016/Q2

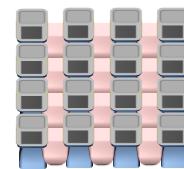
3D NAND



- 40 nm technology
- 16 cells = 16 squares



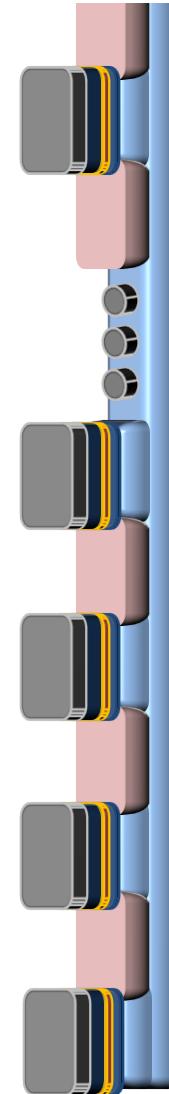
- 20 nm technology
- 16 cells = 4 squares



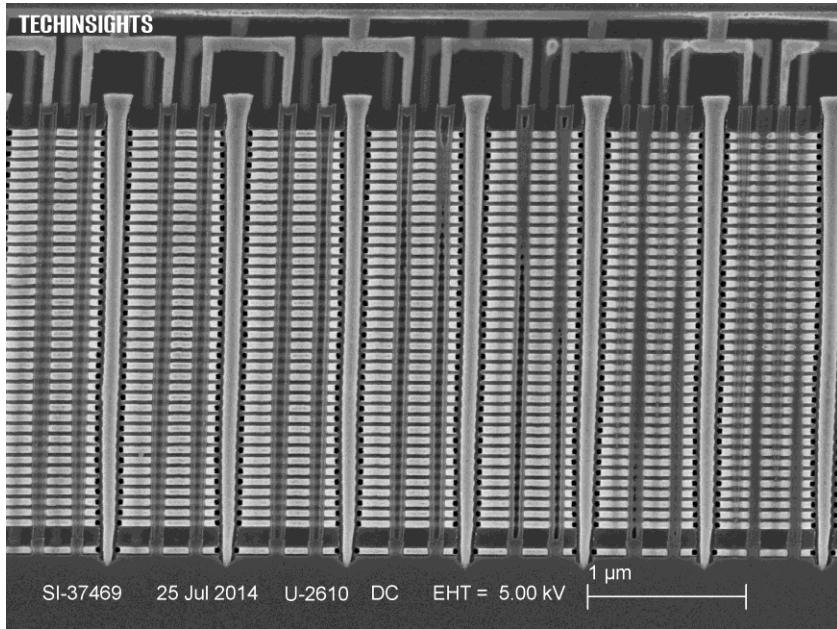
- 10 nm technology
- 16 cells = 1 square

- Vertical NAND allows high density growth with moderate technology
- Usage of charge trap material increases retention

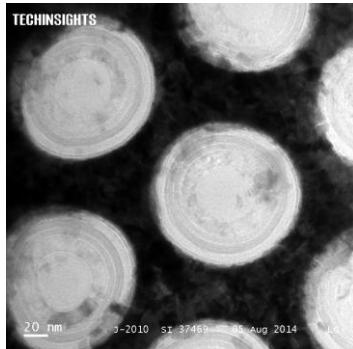
- Vertical NAND
- 40 nm technology
- 16 Cells = 1 square



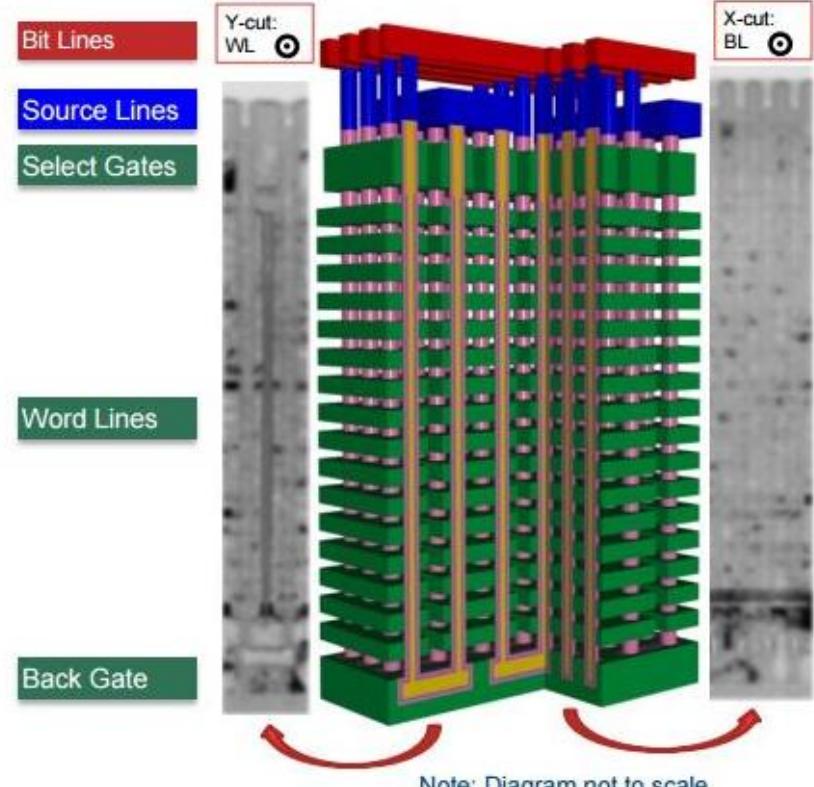
3D NAND



Source: Techinsights / Samsung



Source: Techinsights / Samsung



Source: Samsung

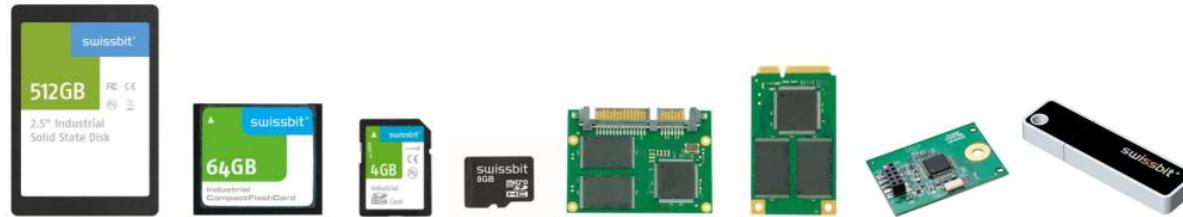
DISRUPTIVE NVM TECHNOLOGIES

Technology	Basic Premise	Figures of Merit					
		Density	Endurance	Reliability	Scalability	Dev. Stage	Cost
STT-MRAM Spin Transfer Torque	Uses spin-polarized current to change the state of a magnetic tunnel junction.	■	★★ (10 ¹⁶)	★	★	★	■
MRAM Magneto resistive	Uses magnetic fields within a grid of nanoscopic power rails to store data which is detected via current flow.	■	★★ (10 ¹⁶)	★★	● (90nm)	★	■
FRAM Ferro-electric	Uses ferro-electric layer versus a dielectric layer to store charge	● (4Mb)	★★ (10 ¹⁴)	★★ (10 ¹⁵)	● (65nm)	★	■
PCM (PRAM) Phase Change Memory	Uses nanoionics to change the state of a Chalcogenide glass from amorphous ($\uparrow R$) to crystalline ($\downarrow R$) & back	■	★ (10 ⁸)	●	★★ (<5nm)	■	■
RRAM Resistive	Stores data by creating a resistance in a circuit rather than trapping electrons in a cell	★ (3D)	★★	■	★★ (<5nm)	■	■
PMC (CBRAM) Programmable Metallization Cell	Based on the physical re-location of ions within a solid electrolyte	● (1Mb)	★ (10 ⁶)	★	★ (10nm)	★	■
Racetrack (DWM) Domain Wall Memory	Uses spin-coherent current to move magnetic domains along a nanoscopic perm-alloy wire	★★	★	★	■	■	■

Nearly all NVM Technologies present performance advantages over today's FG technology
 Several struggle with a tradeoff between Scaling and Power Consumption

I. SSD ARCHITECTURE

PART 2



SSD ARCHITECTURE

TWO PLANE

- Programming two pages at the same time
- Not supported by every flash

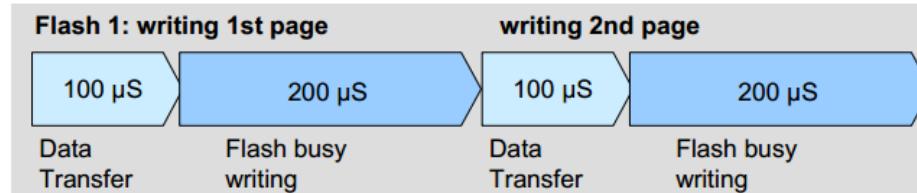


Figure 15: Example of writing to an SLC Flash with single channel configuration and without 2 plane operations where writing takes about 600 µs

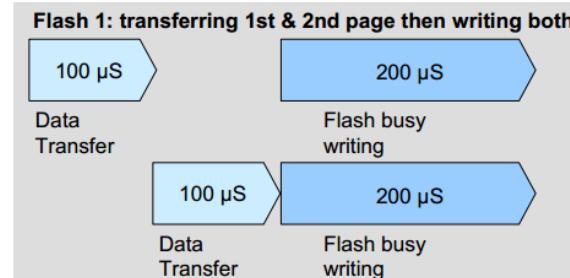


Figure 16: Example of writing to an SLC Flash with single channel configuration and with 2 plane operations where writing takes about 400 µs

SSD ARCHITECTURE

MULTI CHANNEL/INTERLEAVE

Multi Channel:

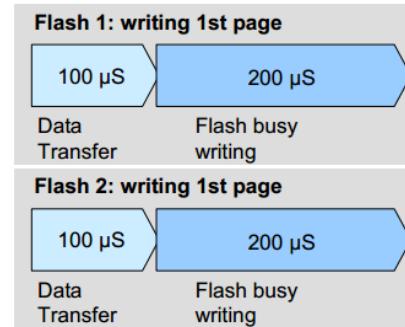


Figure 17: Example of writing to an SLC Flash with 2-channel configuration and without 2 plane operations where writing takes about 300 µs

FLASH HANDLING

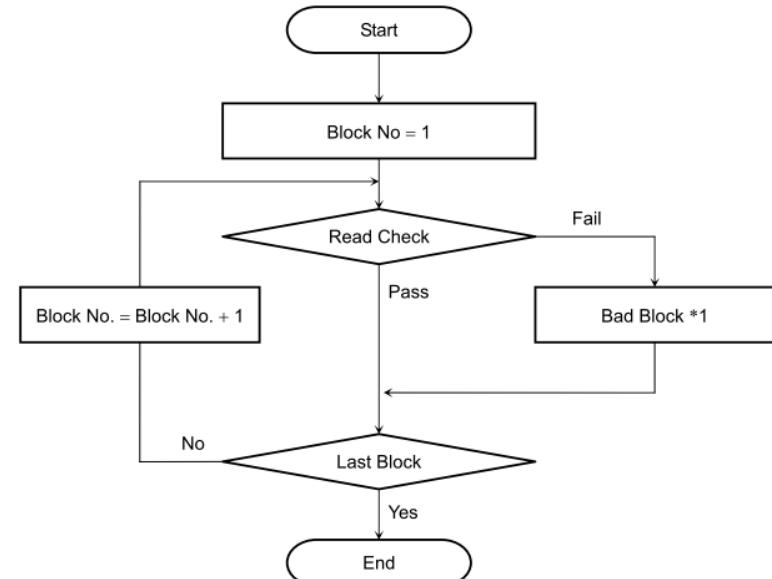
ORIGINAL BAD BLOCKS

Bad or “invalid” blocks are integral part of NAND Flash design

Bad blocks are found in NAND Flash production → “original bad blocks”

The Flash manufacturer specifies the maximum number of bad blocks for full lifetime – this includes both original bad blocks and runtime bad blocks

→ Example:



The number of valid blocks over the device lifetime is as follows:

	MIN	TYP.	MAX	UNIT
Valid (Good) Block Number	2008	—	2048	Block

FLASH HANDLING

RUN TIME BAD BLOCKS

Bad blocks are also detected by the Flash while programming or erasing
→ "runtime bad blocks"

- Need to be handled by the firmware (by checking program status / erase status)

Hyperstone specific:

- "Remap for Erase"
- "Remap for Program"

FLASH HANDLING

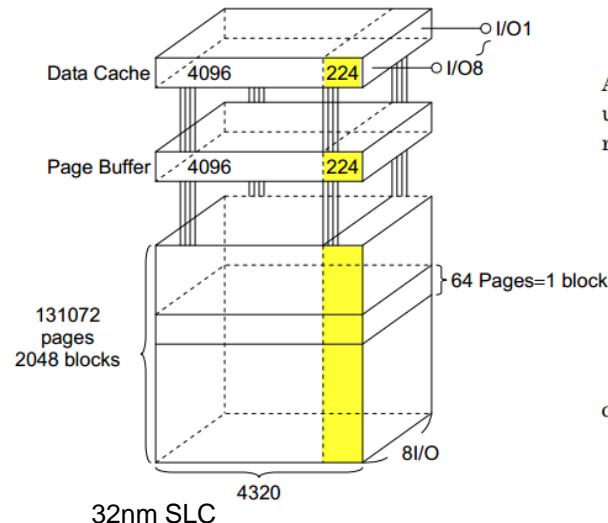
BIT ERROR CORRECTION

It's a requirement

The ECC (Error Correction Code) is responsible for ensuring the quality of data being read, by adding information during the writing phase to restore partially corrupted data.

Schematic Cell Layout and Address Assignment

The Program operation works on page units while the Erase operation works on block units.



A page consists of 4320 bytes in which 4096 bytes are used for main memory storage and 224 bytes are for redundancy or for other uses.

1 page = 4320bytes

1 block = 4320 bytes × 64 pages = (256K + 14K) bytes

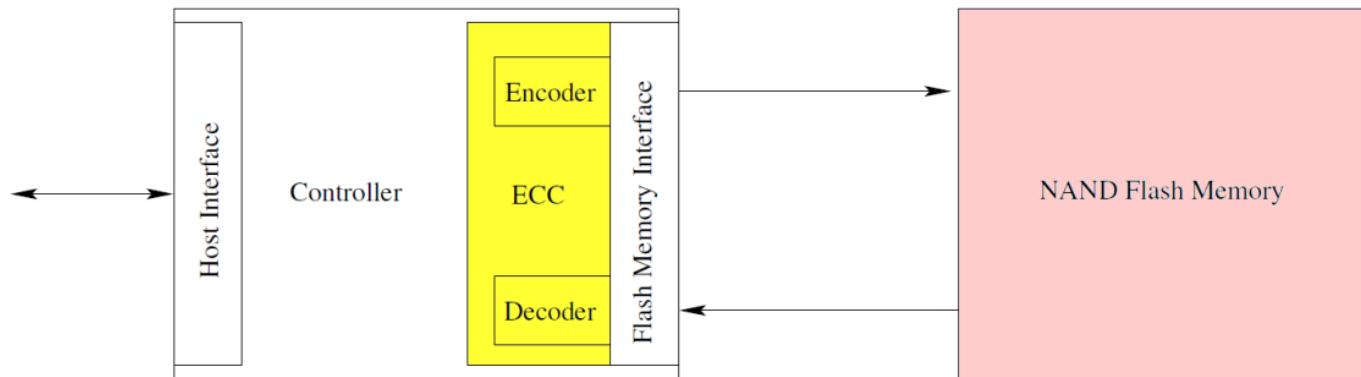
Capacity = 4320 bytes × 64pages × 2048 blocks

An address is read in via the I/O port over five consecutive clock cycles, as shown in Table 1.

FLASH HANDLING

BIT ERROR CORRECTION

ECC is done by the Controller



Flash manufacturers do specify the required ECC engine

Does have an influence to endurance and retention, example:

1 Bit/512 Bytes	-> 60'000 P/E Cycles
6 Bits/512 Bytes	-> 100'000 P/E Cycles

FLASH SPEED COMPARISON

Mode	Read Page	Program Page	Erase Block	P/E Cycles	ECC
SLC (5x nm)	25us	200us	1.5ms	100000	1 Bit / 512
SLC (4x nm)	25us	250us	0.7ms	100000	1 Bit / 512
SLC (3x nm)	30us	300us	3ms	100000	4 Bits / 512
SLC (2x nm)	35us	400us	4ms	100000	24 Bits / 1K
MLC (3x nm)	50us	700us	3ms	5000	24 Bits/1K
MLC (2x nm)	75us	1300us	4ms	3000	24 Bits / 1k
MLC (1x nm)	115us	1600us	5ms	2000	40 Bits / 1k
Pseudo SLC (1x nm)	75us	800us	5ms	5000	40 Bits / 1k
eMLC (3x nm)	50us	1400us	5ms	30000	24 Bits / 1K
3D (~3x nm)	50us	900us	3ms	30000	40 Bits / 1k

Values are average of different vendors or assumptions

FLASH HANDLING

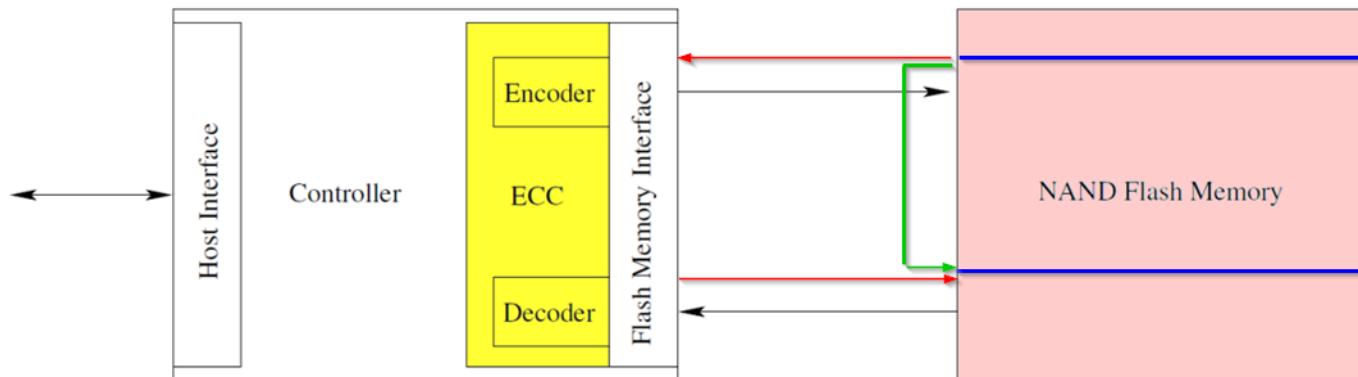
COPY-BACK

Command to copy a page flash-internally.

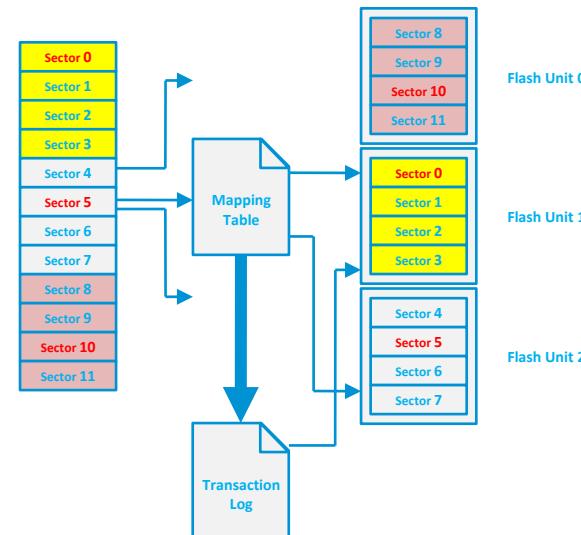
Benefit: No data transfer to controller required -> Better performance!

Disadvantage: Bit error commulation

- Unusable for MLC Flash (Too many bit-errors)
- Periodic check required

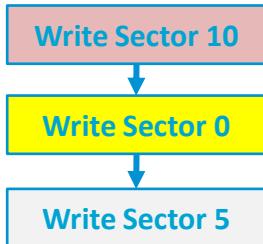


II. FIRMWARE MANAGEMENT ARCHITECTURE, OPTIONS



FTL MAPPING

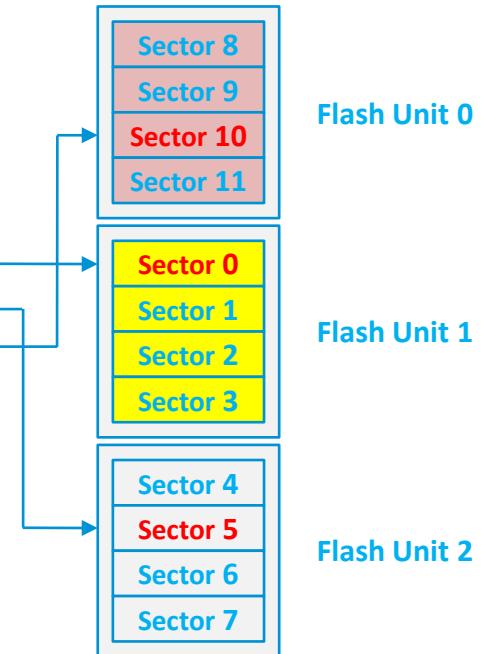
Write sequence



Logical Sector Address

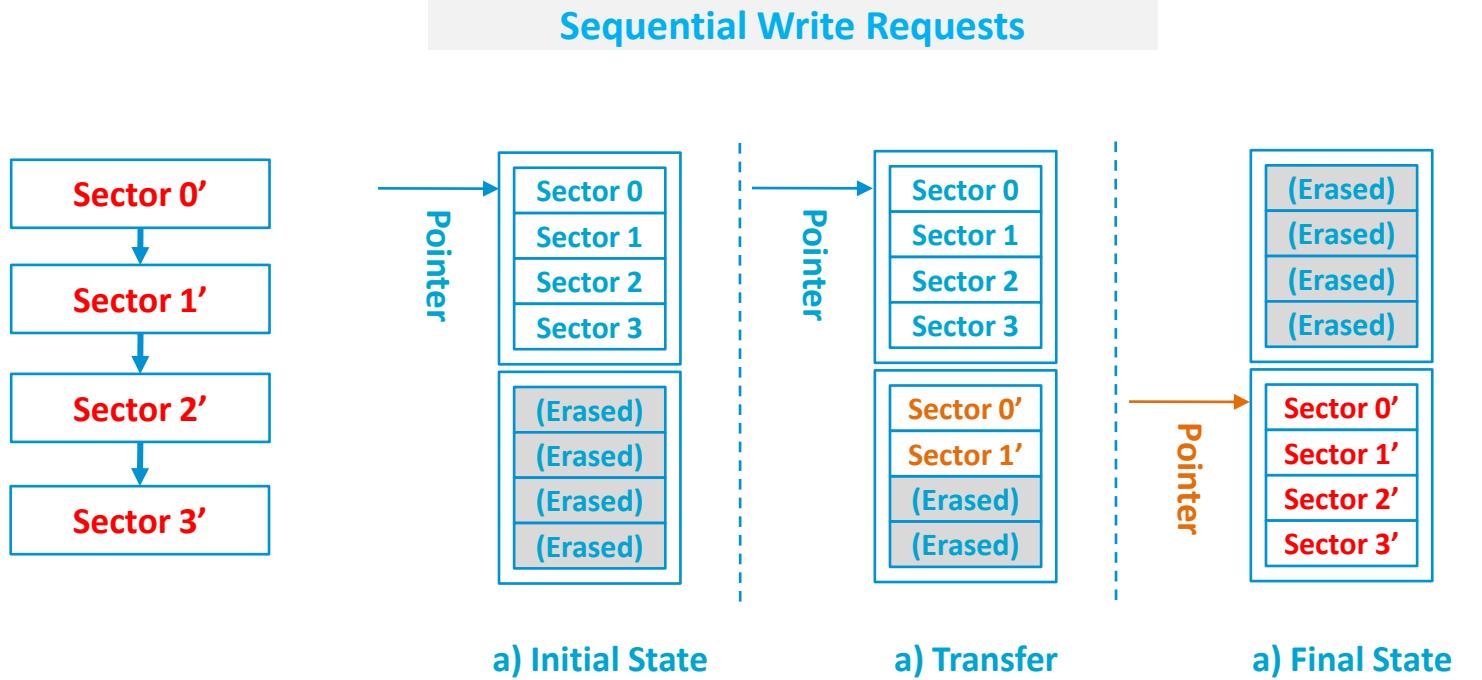


Physical Sector Address



Flash Translation Layer

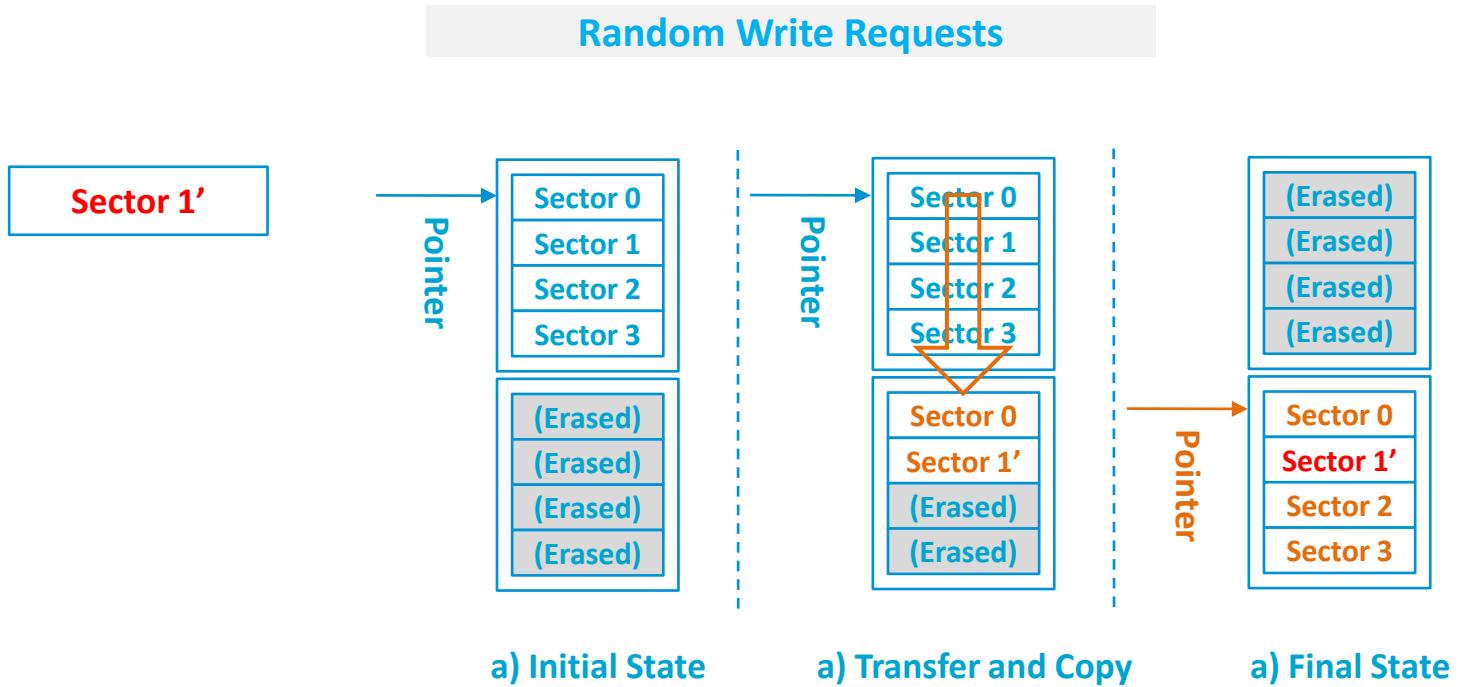
BLOCK MANAGED WRITE



WAF consideration

- Sequential writes can be combined into one block
- Writing one logical block leads to writing one physical block and one block erase -> WAF = 1

BLOCK MANAGED WRITE

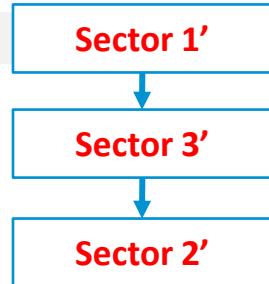


WAF consideration

- Random writes require one block write and one block erase for each sector
- WAF depends on physical block size and is $>> 1$ (can be up to 100)

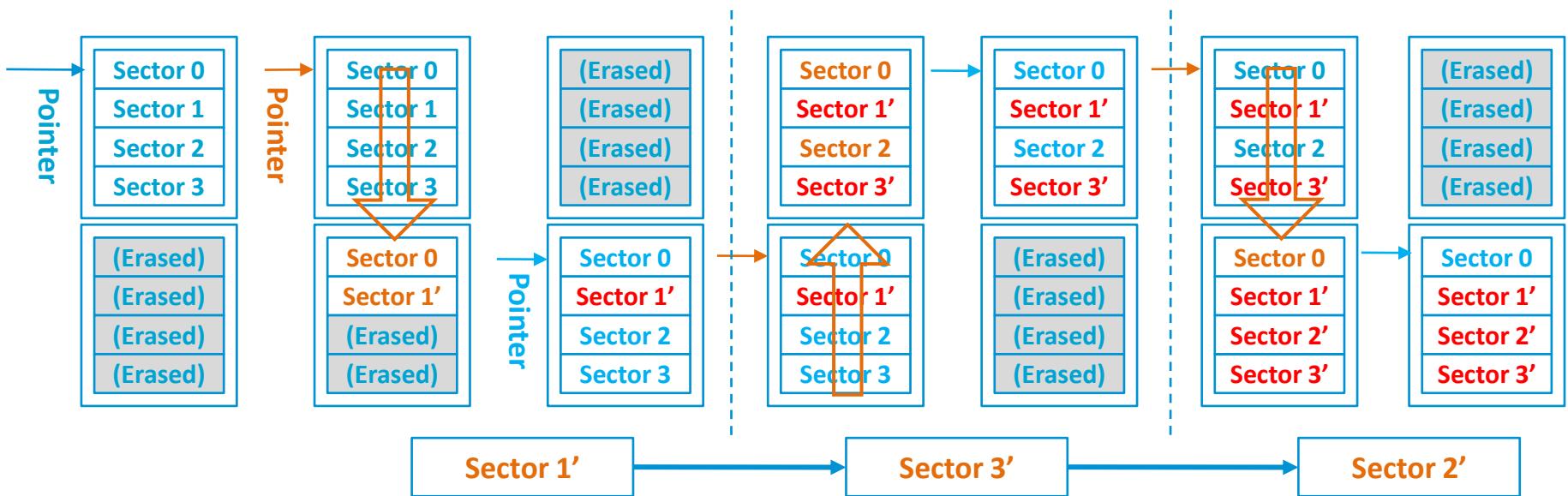
BLOCK MANAGED WRITE

Random Write Requests



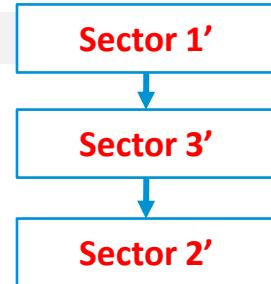
WAF consideration

- Random writes cannot be combined
- WAF depends on physical block size and is $>> 1$
(can be up to 100)



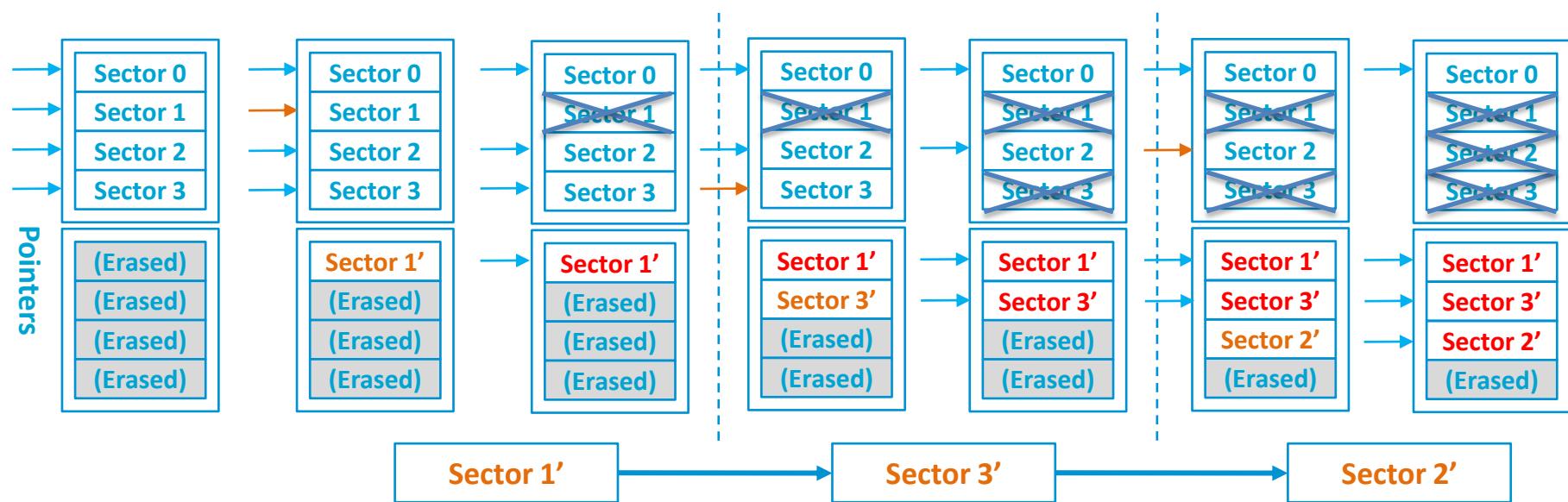
PAGE MANAGED WRITE

Random Write Requests



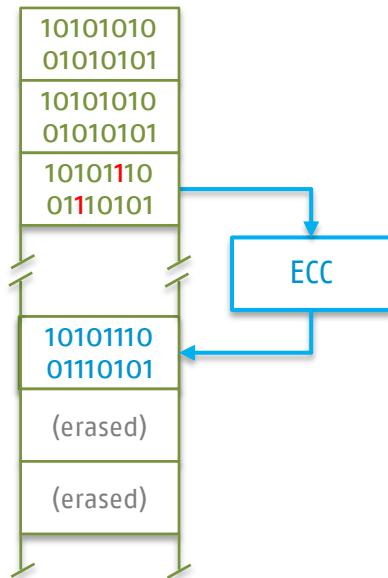
WAF consideration

- Random writes can be combined in new block
- WAF depends on garbage collection = ~ 1

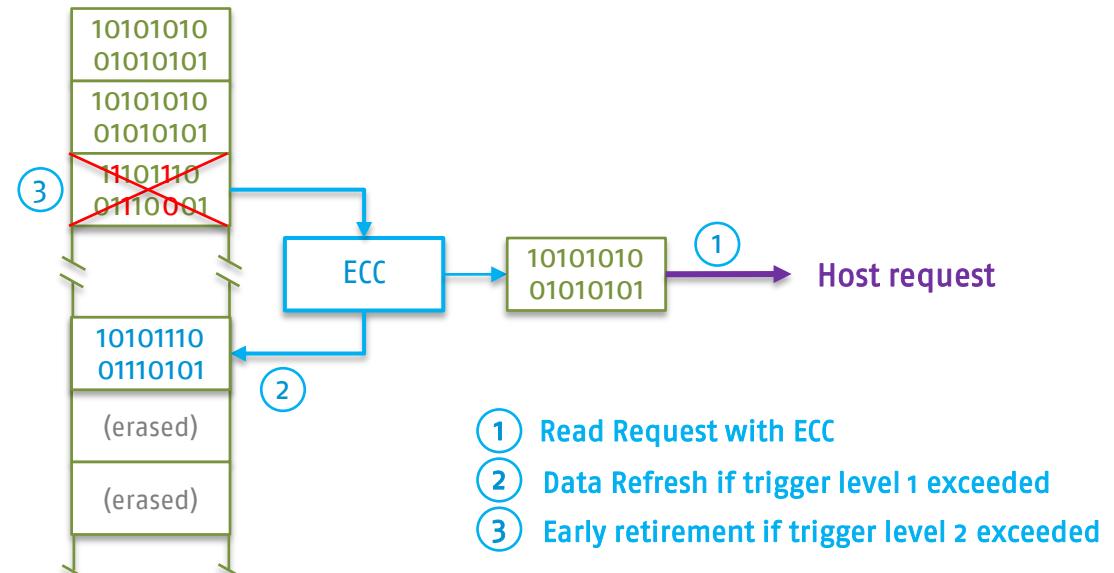


DATA CARE MANAGEMENT

NAND Flash die shrinks pose a greater challenge for the controller manufacturers to insure the same level of performance, endurance and reliability.



Autonomous Background
Media Scan (Passive)



ECC, Data Refresh & Early
Retirement (Active)

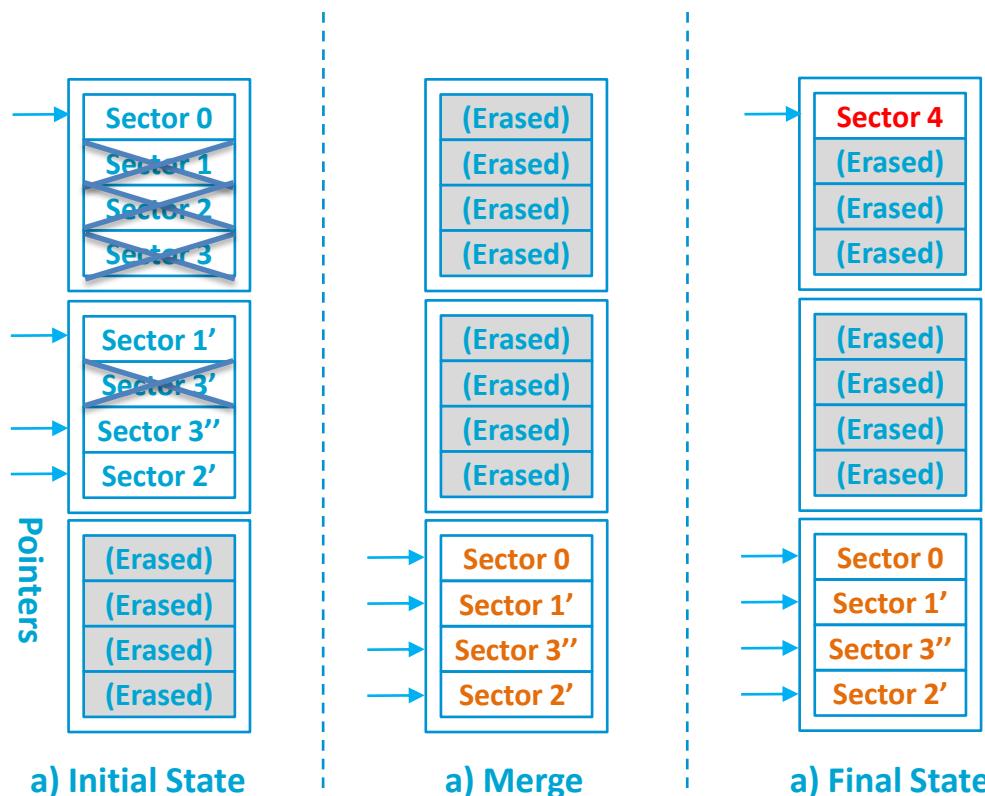
GARBAGE COLLECTION

Random Write Request:

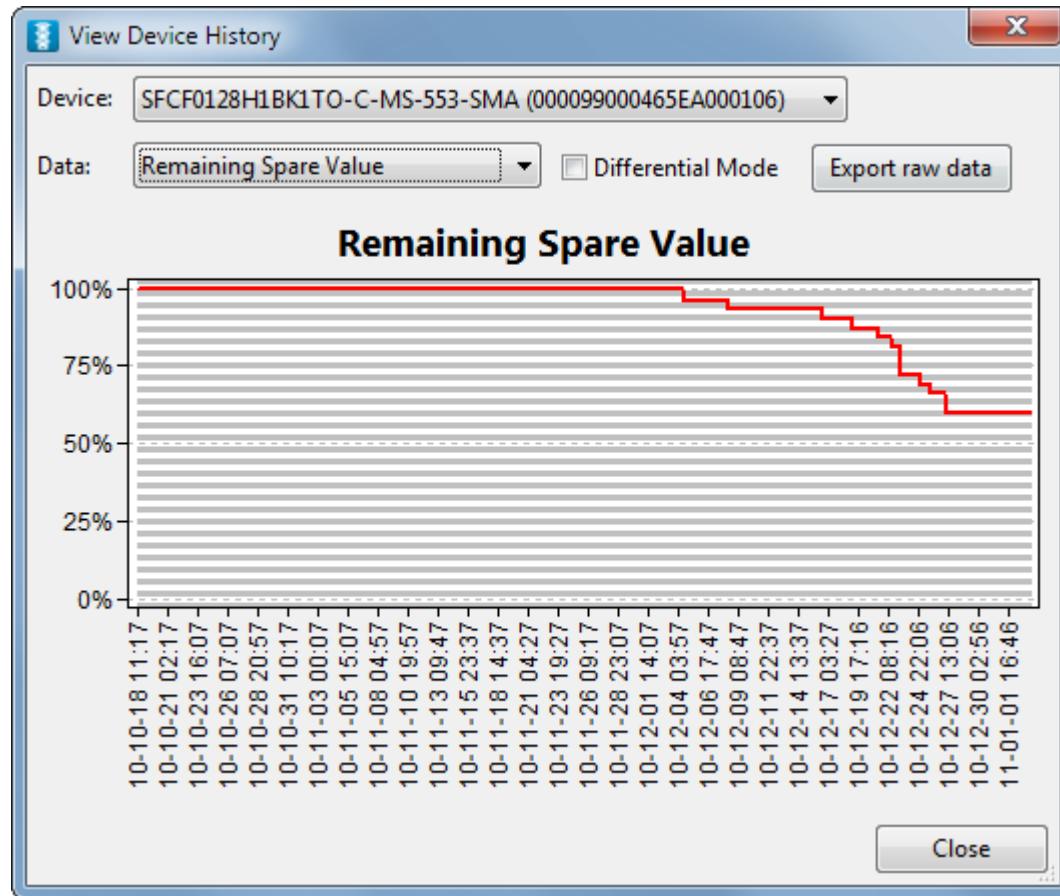
Sector 4

WAF consideration

- Garbage collection adds P/E cycles and increases WAF

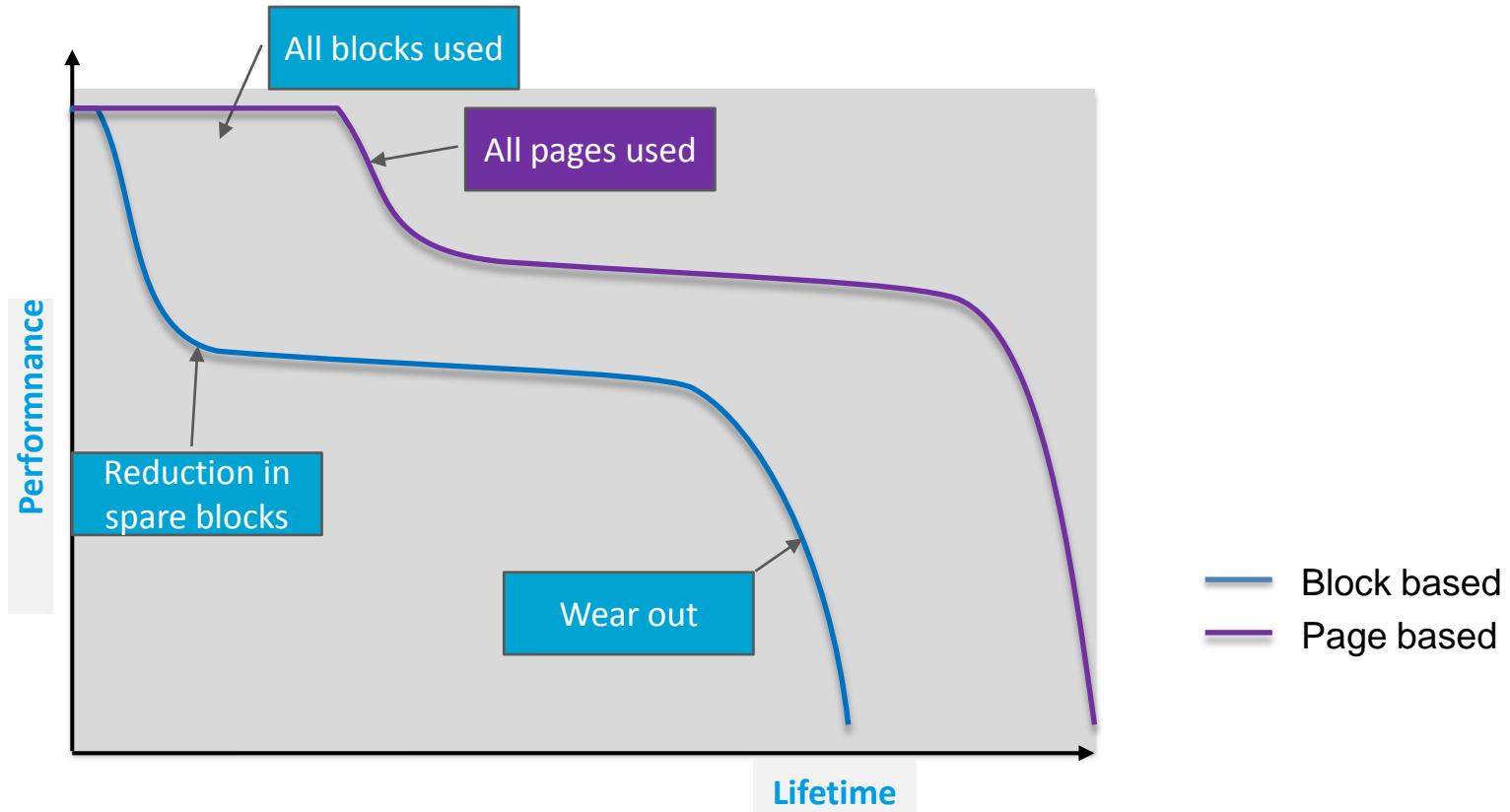


SPARE BLOCK



- Spare blocks and overprovisioning are needed for garbage collection
- Garbage collection has a negative performance impact
- If spare blocks are reduced the performance impact is higher

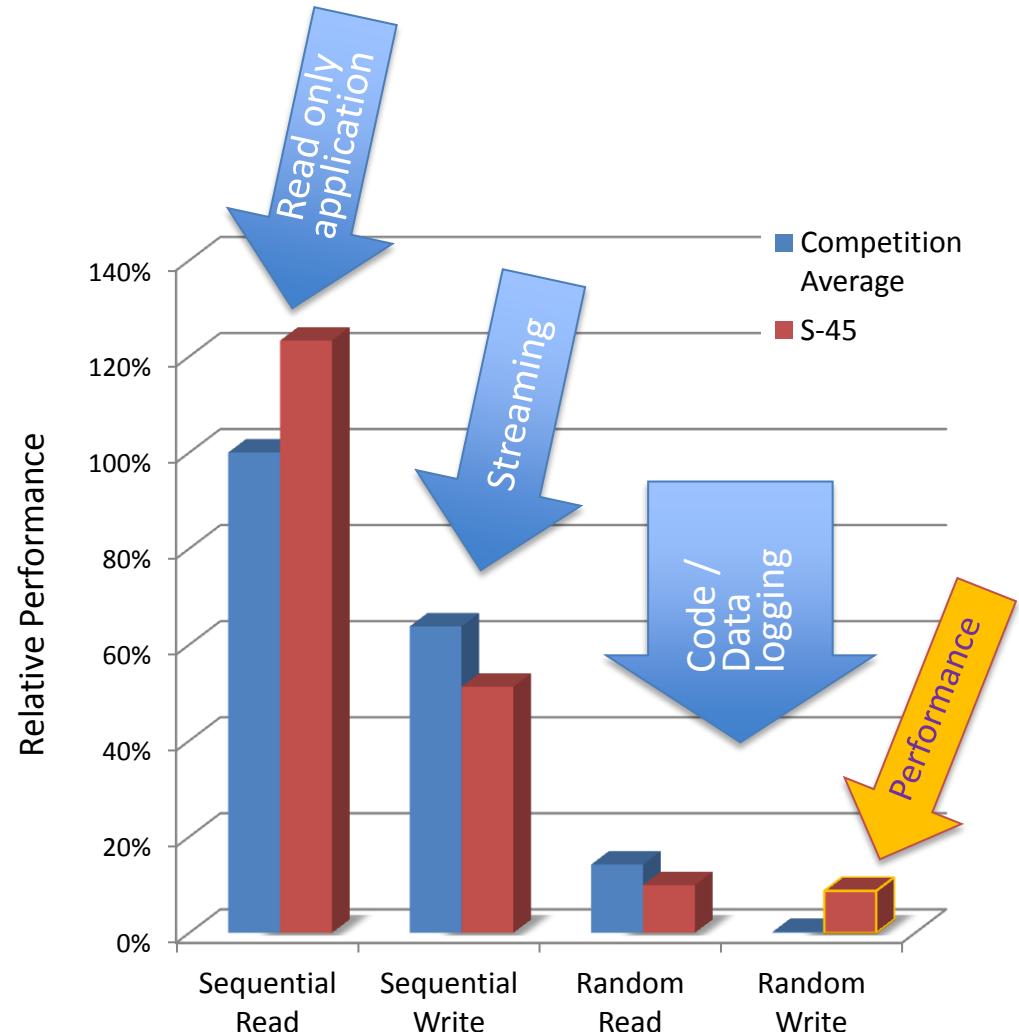
LIFETIME RANDOM PERFORMANCE



- Initial performance (or after TRIM) is high until blocks are used up
- Block combine and garbage collection reduce usable performance
- Page based FTL has more efficient use of free blocks and later wear out

BLOCK BASED VS. PAGE BASED

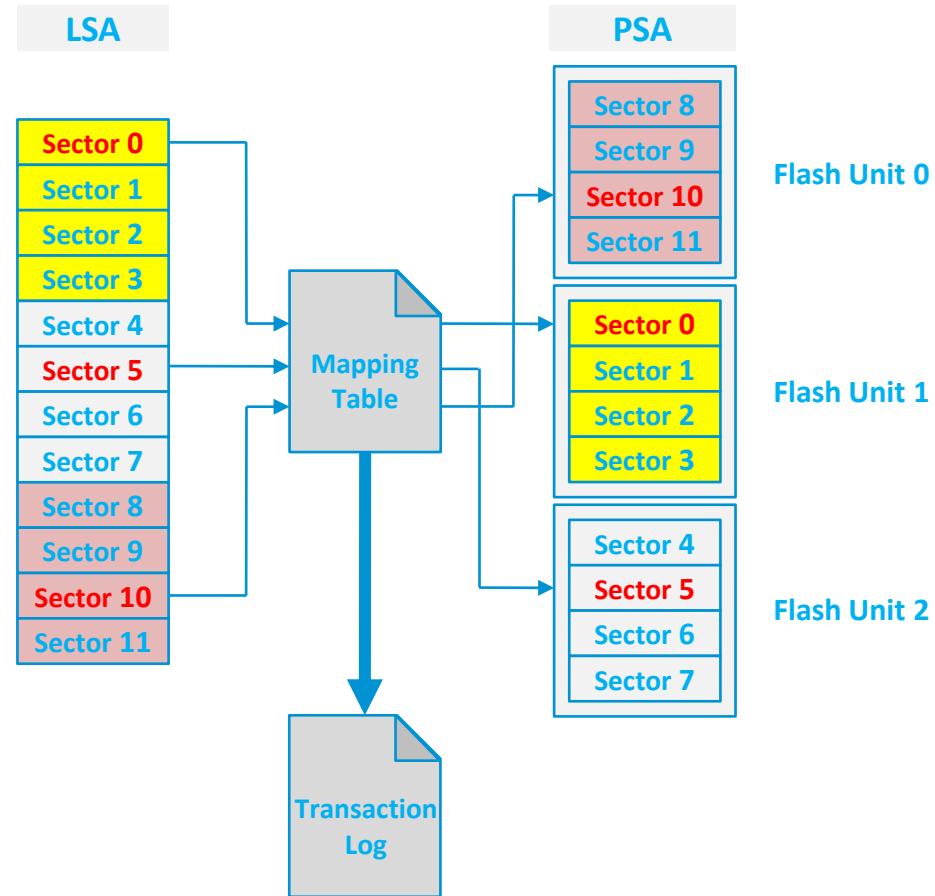
Topic	Block based	Page based
FW complexity	low	high
Sequential performance	high	small reduction
Random write performance	Very small	high
WAF	High (~100)	Low (~1)
Endurance	Normal	10-100x
Retention	Normal	Normal
Busy time	Long	Short



POWER FAIL MANAGEMENT

Actions at sudden power loss

- Reset of controller and immediate write protection of flash
- Keep log of recent flash transactions
- If the last data of this log is corrupt, the controller will recover the latest valid entry
- If a write operation is active at power loss this data might be lost
- As the FW keeps the original data in a so-called twin of the active block it can always recover the last valid state
- Recovery can be done by scanning the transaction log and the block info
- Disable fast acknowledge and write cache prevents loss of user data



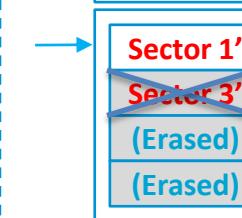
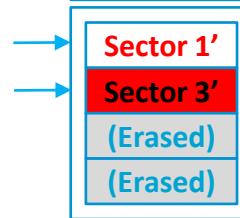
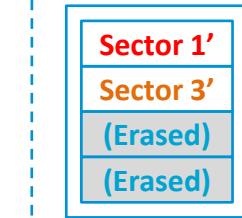
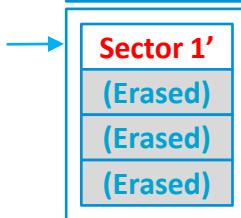
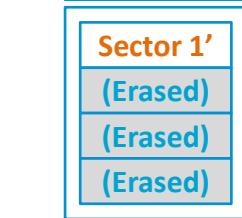
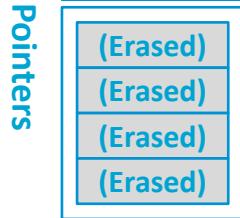
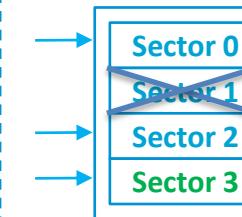
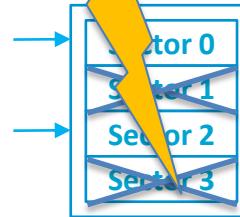
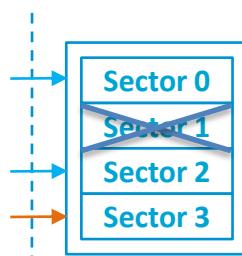
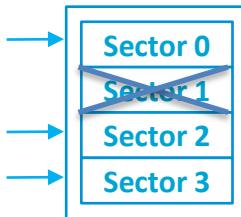
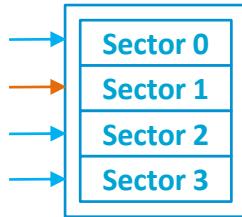
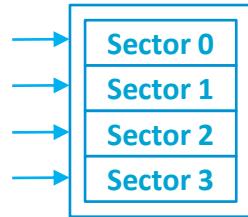
POWER FAIL MANAGEMENT

Random Write Requests

Sector 1'

Sector 3'

Power loss at write



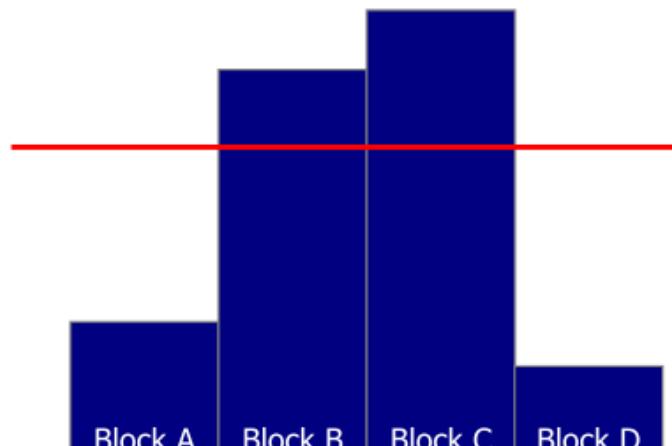
Sector 1'

Sector 3'

Sector 3

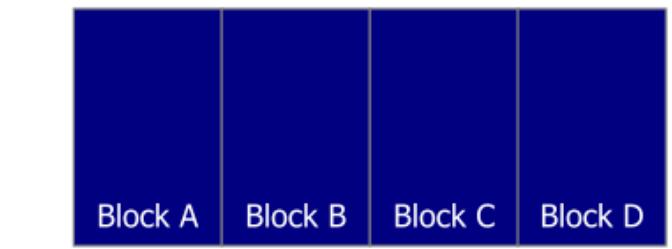
- During recovery from power loss the last not written page is reverted to the previous state

WEAR LEVELING



Without wear leveling

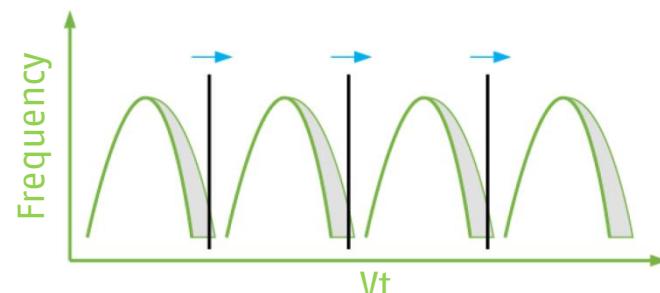
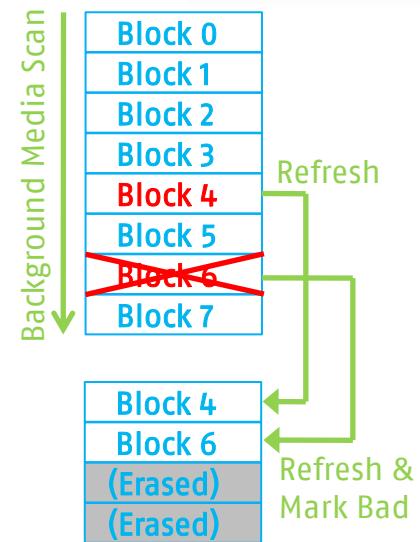
100%
erase
cycles



With wear leveling

FEATURES

- **DATA CARE MANAGEMENT**
- Swissbit S-4x cards introduced a read disturb management to improve data retention for read intensive applications.
- **Autonomous Read Refresh**
- An independent background media scan process checks the health status of each block and rewrites the block if its data content shows signs of reduced retention.
- **Near Miss ECC Read/ Early Retirement**
- On each read operation the health status of the read block is evaluated and refreshed if necessary
- **Read Retry**
- If a read operation to a page shows uncorrectable errors then a sophisticated algorithm retries the read with different internal reference voltage levels. This eliminates typical fails that occur due to temperature changes between writes and reads.



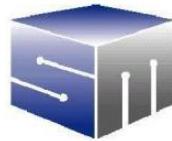
IV. CONTROLLER THE KEY INGREDIENT

CONTROLLER ALIGNMENT

Swissbit leverages 3 primary SSD Controller partners to address the Automotive, Embedded, Industrial and Netcom Infrastructure markets

Controller Partner	Market Focus	Level of Engagement	Product Breadth
 Hyperstone	Embedded/Industrial Applications	Highest	ATA, USB, SD, SATA
 Silicon Motion	Consumer & Embedded/Industrial Applications	Very High	ATA, USB, SD, SATA, PCIe
 Phison	Consumer & Embedded/Industrial Applications	Moderate, but growing	ATA, USB, SD, SATA, PCIe

SSD CONTROLLER ROADMAP



SMI2246EN

SATA III
4-CH, 8CE/CH
66b/1KB ECC
AES 128/256
R/W: 530/420 MBs

SMI2256

SATA III
4-CH, 8CE/CH
LDPC & Flash RAID
AES 128/256
R/W: 560/460 MBs

SMI2262

PCIe Gen2 x 2
AHCI
4-CH, 8CE/CH
BCH ECC
R/W: 750/520 MBs

SMI2260

PCIe Gen 3 x 2, Gen 2 x 4
NVMe + AHCI
8CH, 4CE/CH
LDPC & Flash RAID
R/W: 1300/1000 MB/s

PS3110

SATA III
8-CH, 8CE/CH
66b/1KB ECC
AES 128/256
R/W: 540/480 MBs

PS5007

PCIe Gen3 x 4
AHCI + NVMe
8-CH, 8CE/CH
BCH ECC 75b/2KB
R/W: 2800/1500 MBs

2H CY14

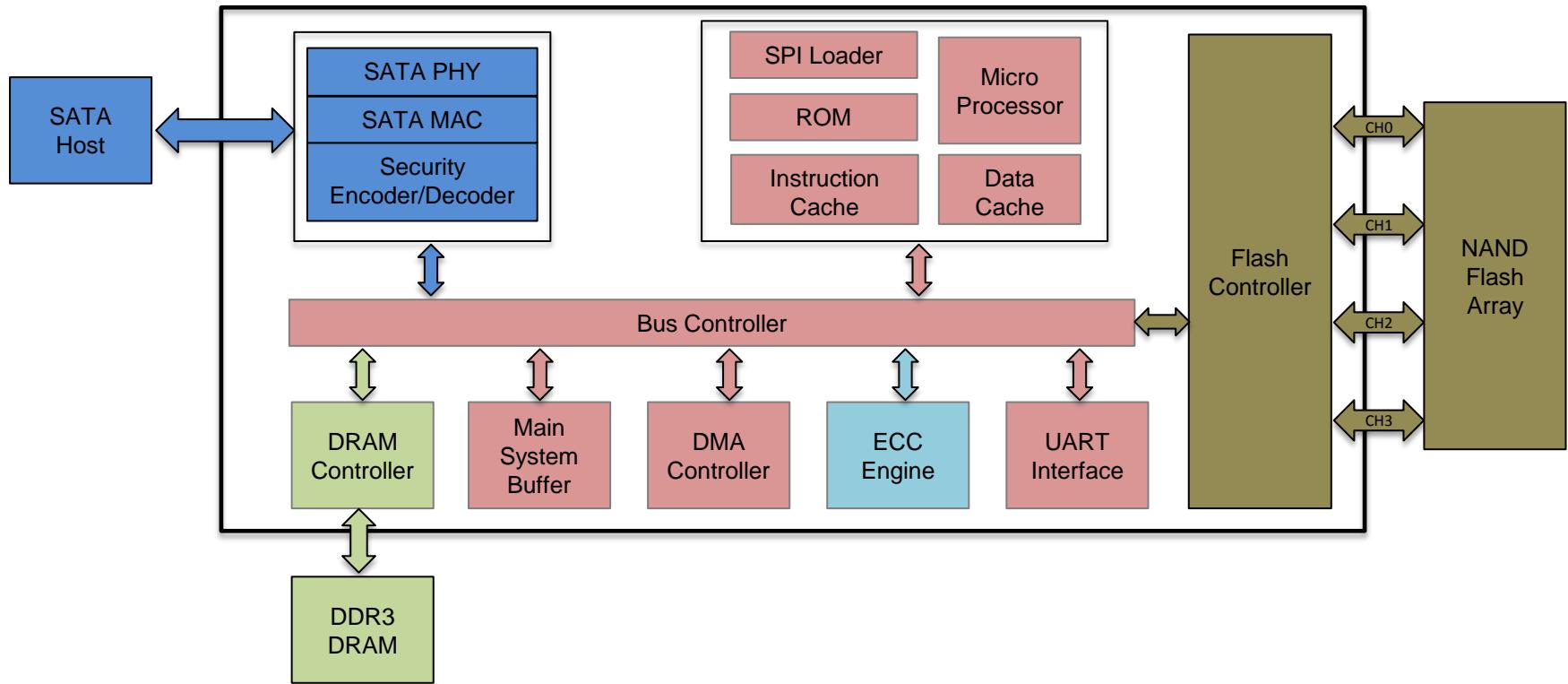
1H CY15

2H CY15

1H CY16

2H CY16

SATA GEN III CONTROLLER



Host Interface
SATA 3.1 (6Gb.s)
NCQ 32
TRIM
S.M.A.R.T.
SATA Device Sleep

DRAM Interface
x16 bus
DDR3
128Mb or 256Mb
(512GB)

System Architecture
32b RISC CPU, up to
370Mhz
Data Buffer, GP SRAM –
520KB
64b Internal Bus

Data Reliability
ECC BCH 66b/IKB
Background Refresh
Read Retry on UECC
ECC Threshold
Detection

NAND Interface
4 Channel x8
ONFI 3.0/Toggle2.0
200Mhz DDR (400MB/s)
8 CE/channel

POWER MANAGEMENT

