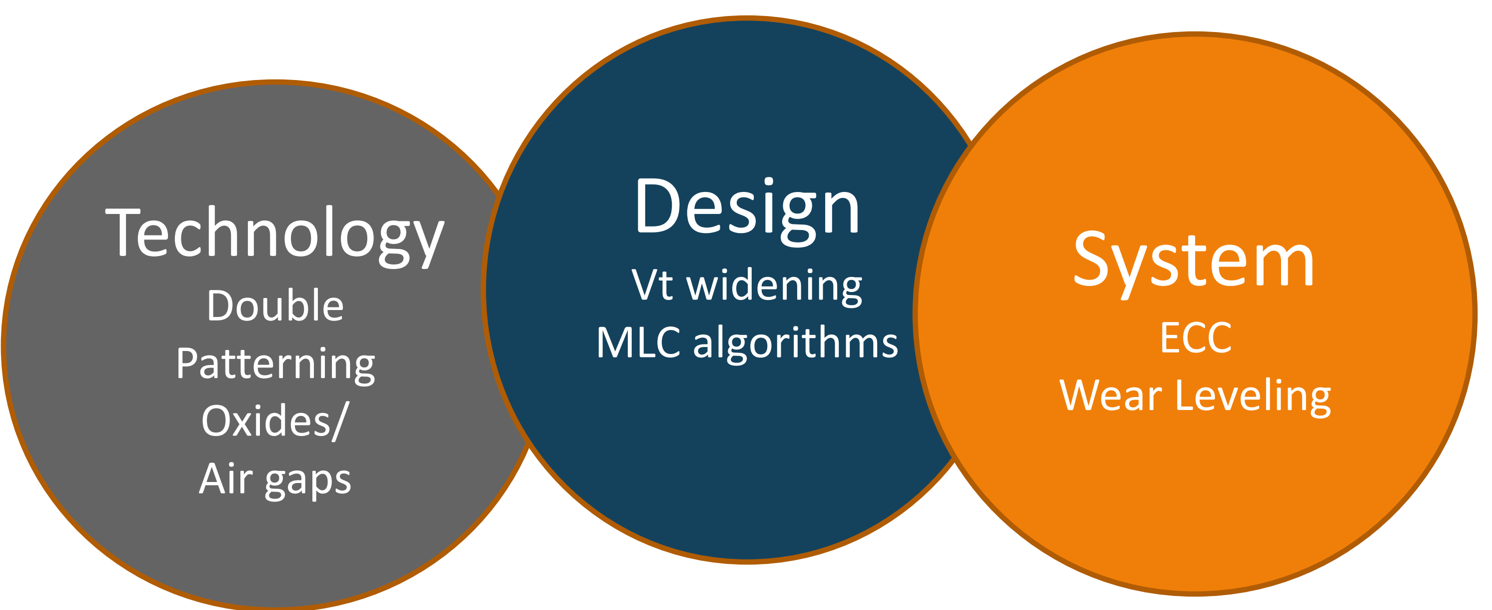
Conductive Metal Oxide Arrays for Terabit Memories

Workshop on Innovative Memory Technologies

June 29, 2011

What made NAND Flash work?



It takes developments at all levels to achieve a commercially viable product. This presentation concentrates on the design aspects of RRAM.

Outline

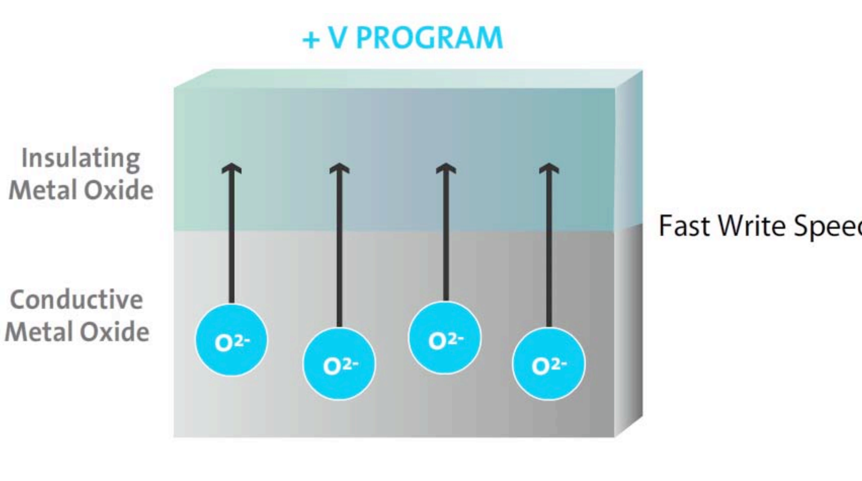
- CMOx technology
- Array construction
- WL length
- BL length
- Disturb
- Power and Speed
- Results

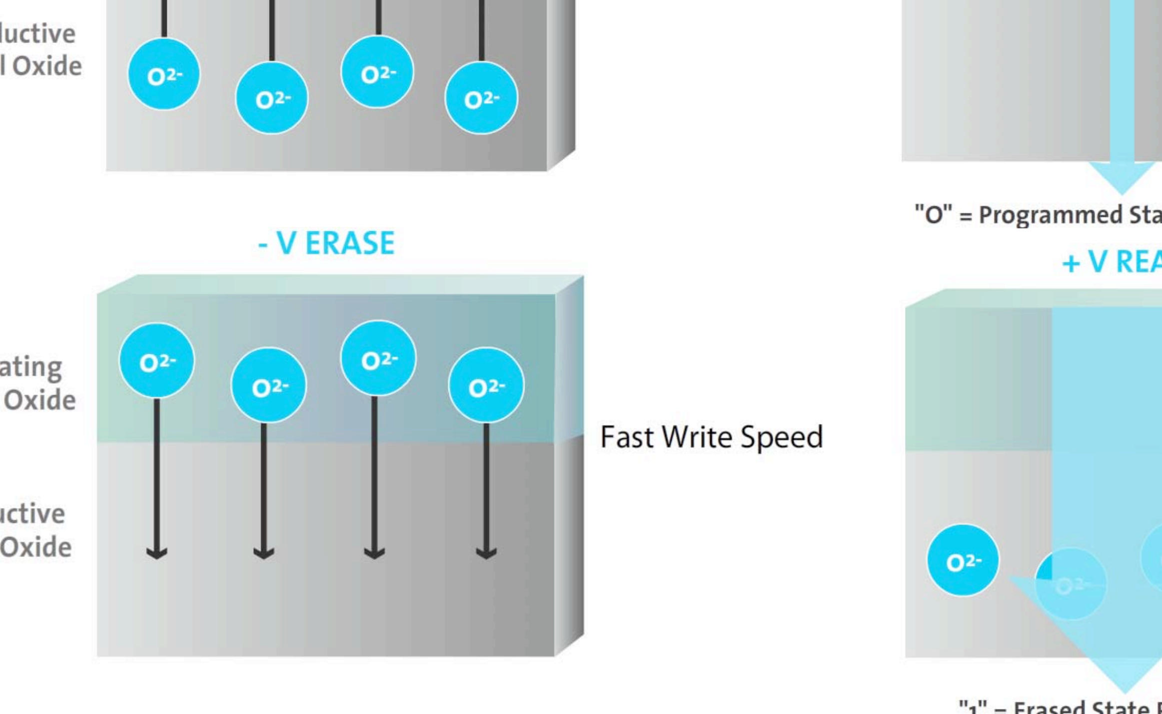
© Unity Semiconductor Corporation

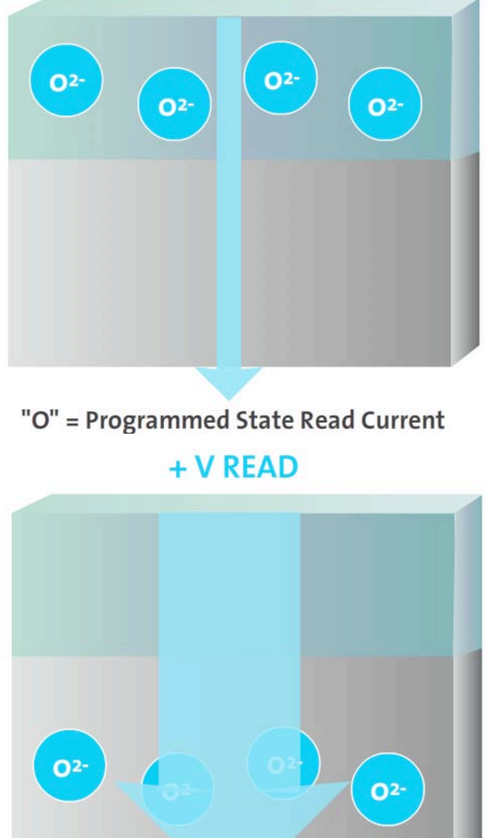
Outline

- CMOx technology
- Array construction
- WL length
- BL length
- Disturb
- Power and Speed
- Results

Unity CMOx^{TM}: Movement of Ions



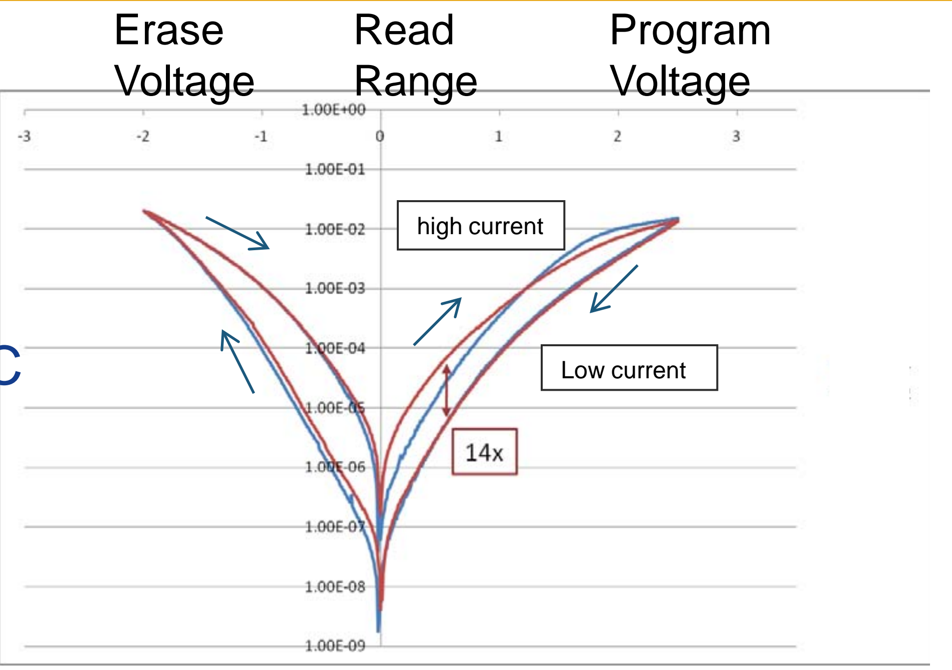


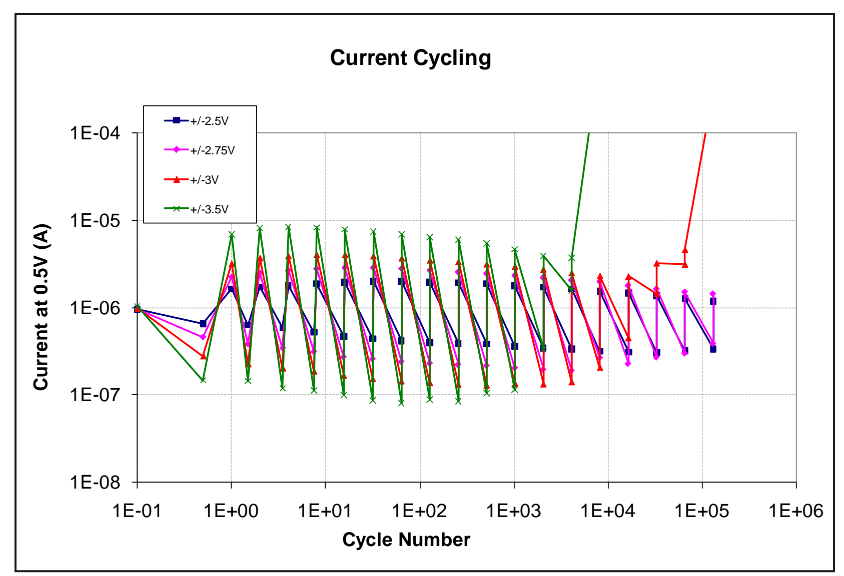


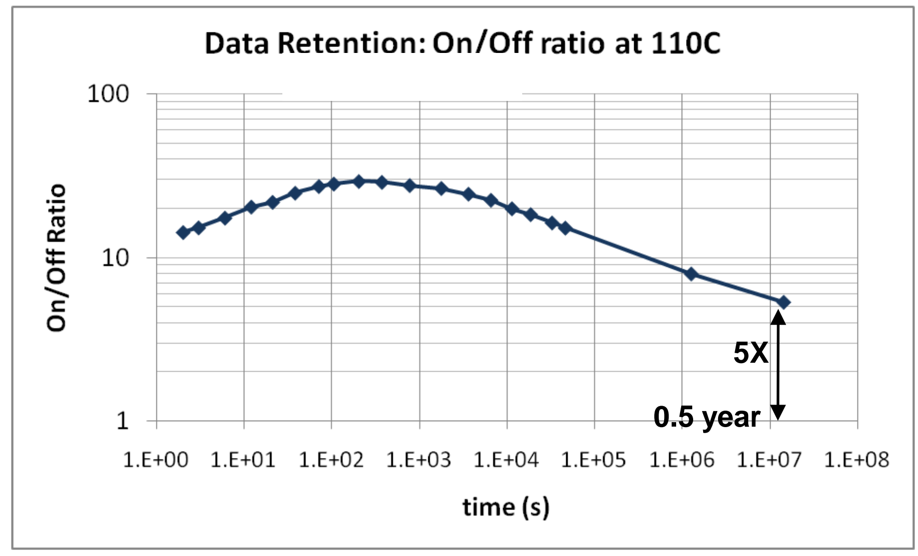
CMOX WORKS BY ION MOVEMENT UNDER ELECTRICAL FIELD

CMOx cell to date:

- IV curve with program/erase
- Cycling to over 100,000 cycles
- 6 months data retention at 110C
- 1 year at 70C spec



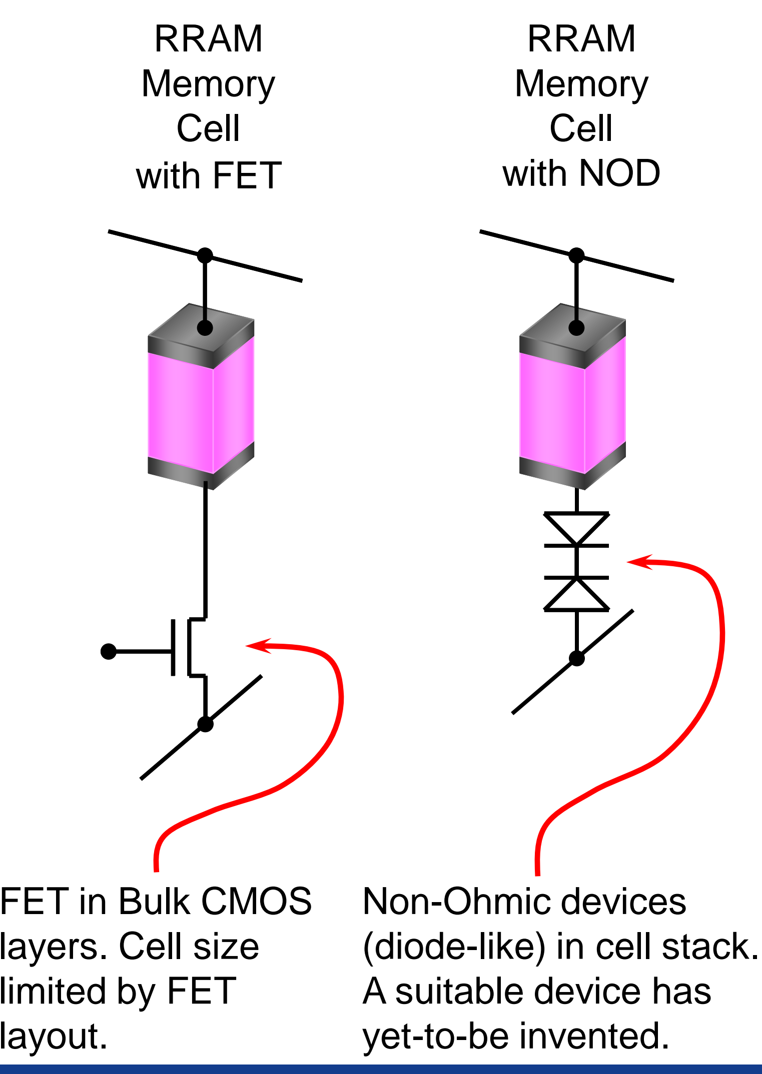




© Unity Semiconductor Corporation

Where is the Select Device?

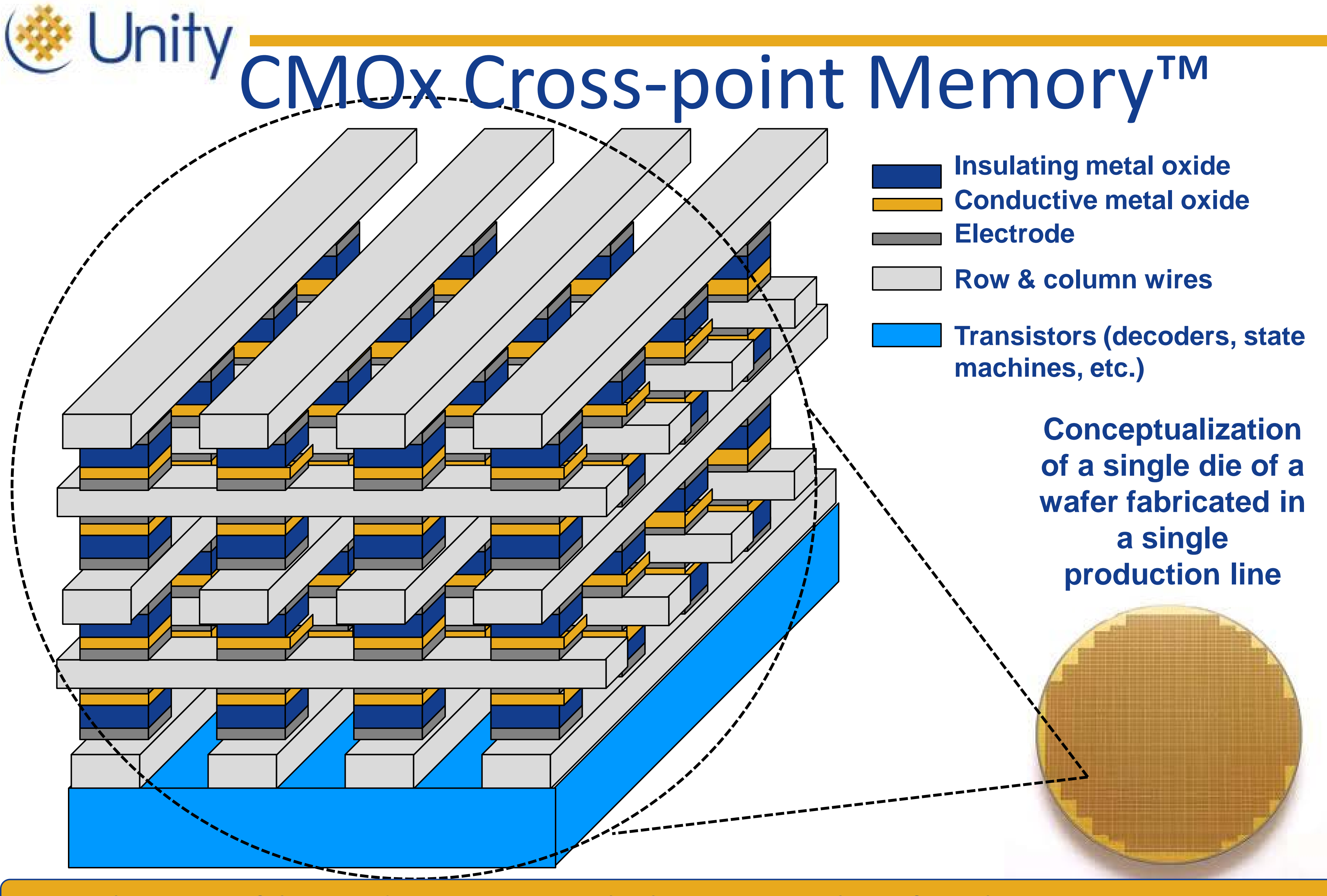
Cross-point array dimensions are severely limited by stray currents unless some sort of select device is introduced



Non-linear electrical characteristics of CMOx element provides "select" function as inherent part of cell operation

© Unity Semiconductor Corporation

CMOx Cross-point Memory™



4 layers are fabricated in approximately the same number of masking steps as NAND

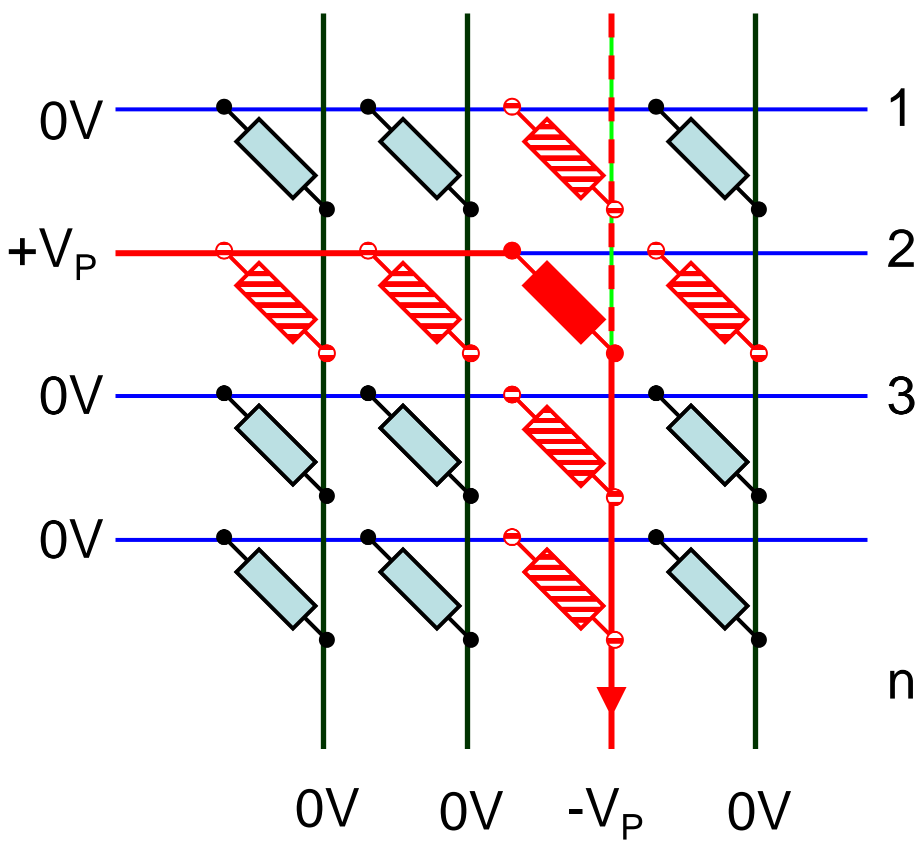
© Unity Semiconductor Corporation

Outline

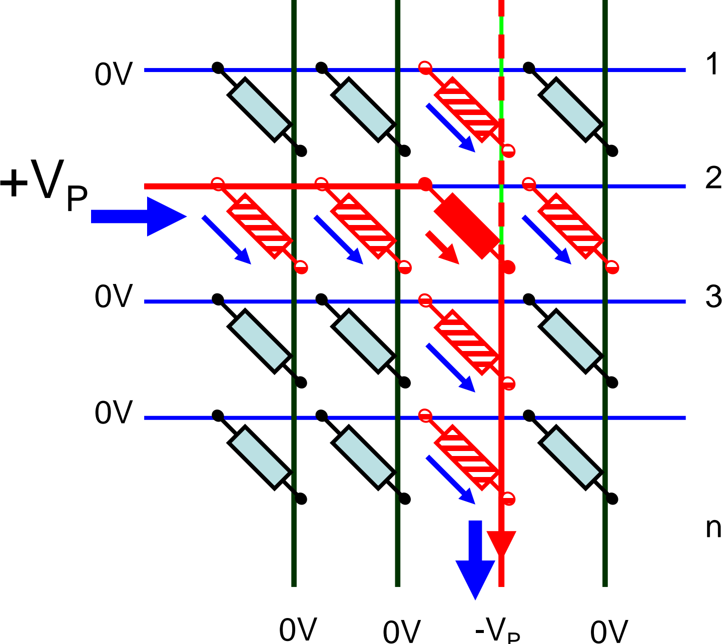
- CMOx technology
- Array construction
- WL length
- BL length
- Sensing
- Disturb

© Unity Semiconductor Corporation

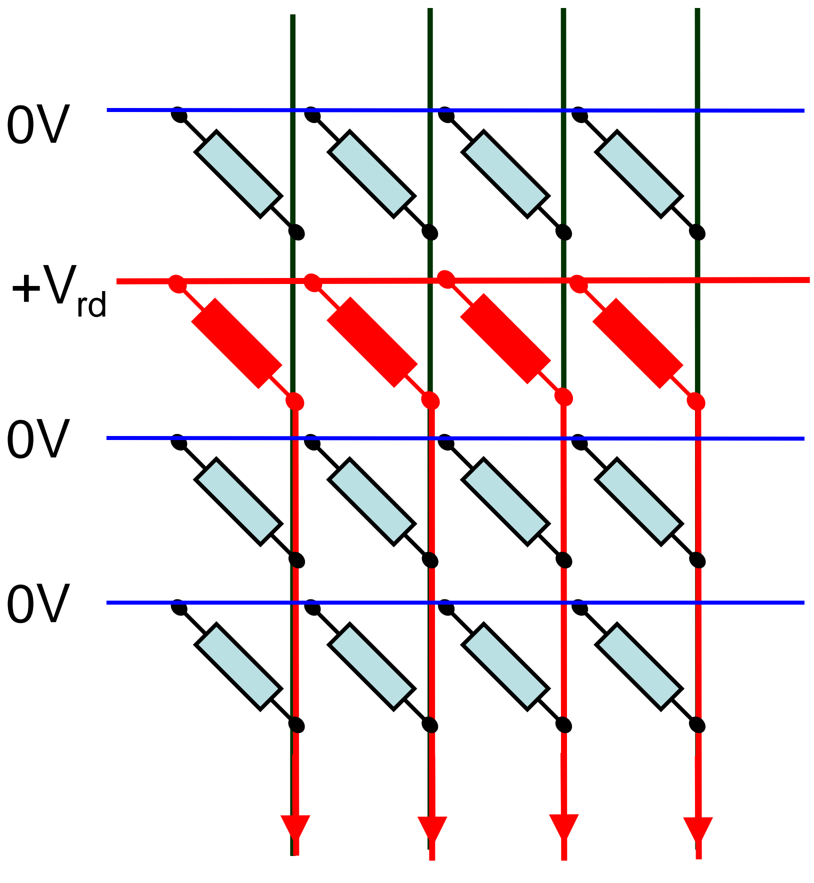
Cross Point Array Half-Voltage Selection



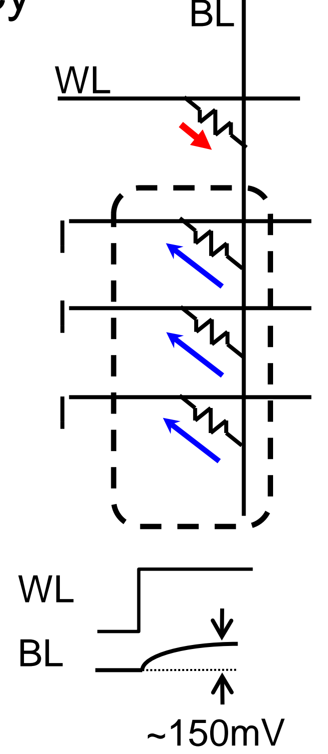
- half-selected cells do not program, but draw current



Read Selection



- Low current cell
- Longer latency



Read

© Unity Semiconductor Corporation

Cross Point Array Limitations

- Array Lines:
  - Total current (electro migration limit)
  - IR drop during program/erase
  - Disturb of unselected cells
  - Loss of sensing margin (on Bit Lines)

- Maximum Array size = 4096 x 128 x 4 layers
  - Row length (4096) limited by IR drop and EM during program and erase
  - Column length (128 x 2 layers) limited by cumulative stray currents thru non-selected cells during read

© Unity Semiconductor Corporation

Cross Point Array with Local Bit Lines

Memory Physical Organization

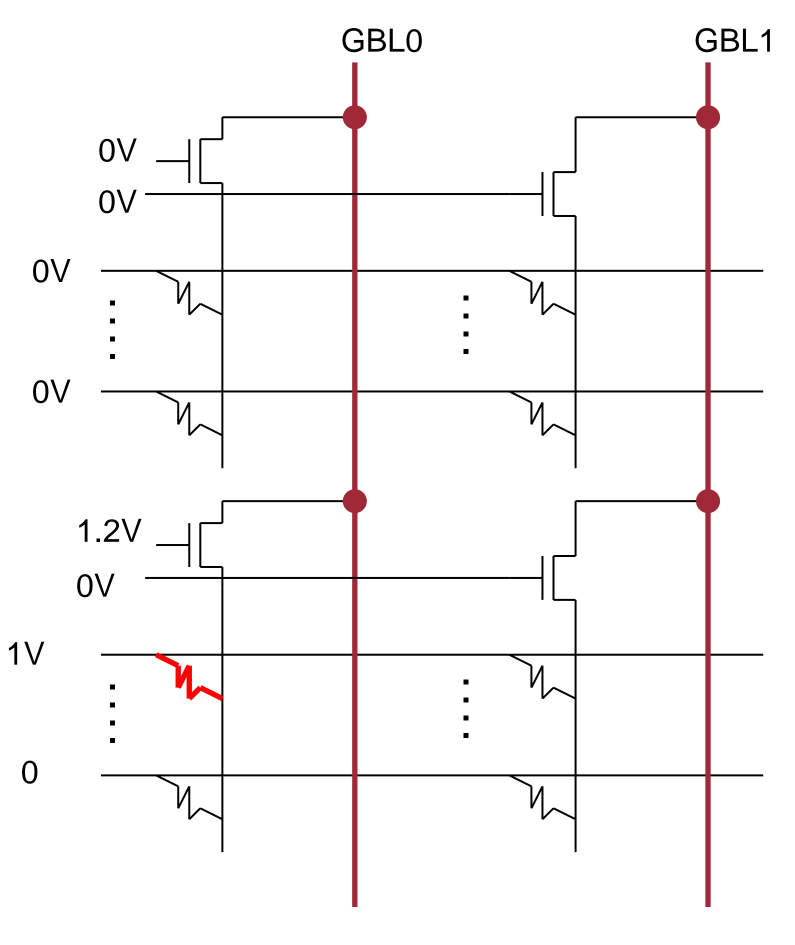
128x
Row Decode and Drivers
located along edge of Array
Layer 3
Layer 2
Layer 1
Layer 0
4096 Columns
128 Rows
Row
Drivers
X2
Block
4096 Columns
128 Rows
Row
Drivers
Y2
X2
Block
Y1
X0
Y0
Block bitline 8:1 select FETs
located underneath Block
16:1
512b
Logic
Column Decode, Sense
Amps, Write Drivers
located along edge of Array

New Array Advantages

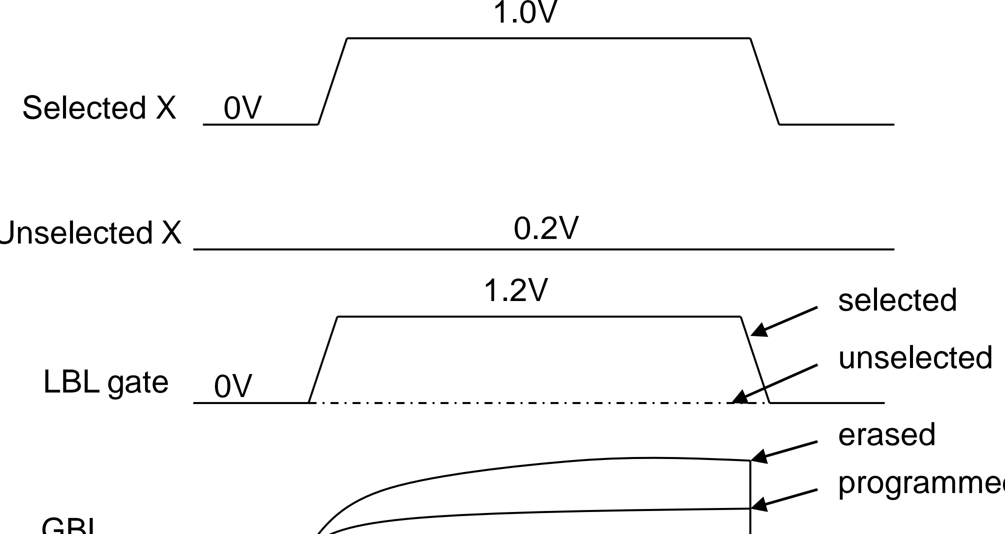
- Short Bit Lines:
  - Controlled read leakage
  - No IR drop issues on BL
  - No Electro Migration issues on BL
  - reduced Program Disturb
  - Some floating bit lines
  - Y decoders are laid out under the memory arrays
- Long Word Lines:
  - IR drop will be cancelled out
  - Limited by electro-migration
  - Limited by program disturb
  - Small total X-decoder area

© Unity Semiconductor Corporation

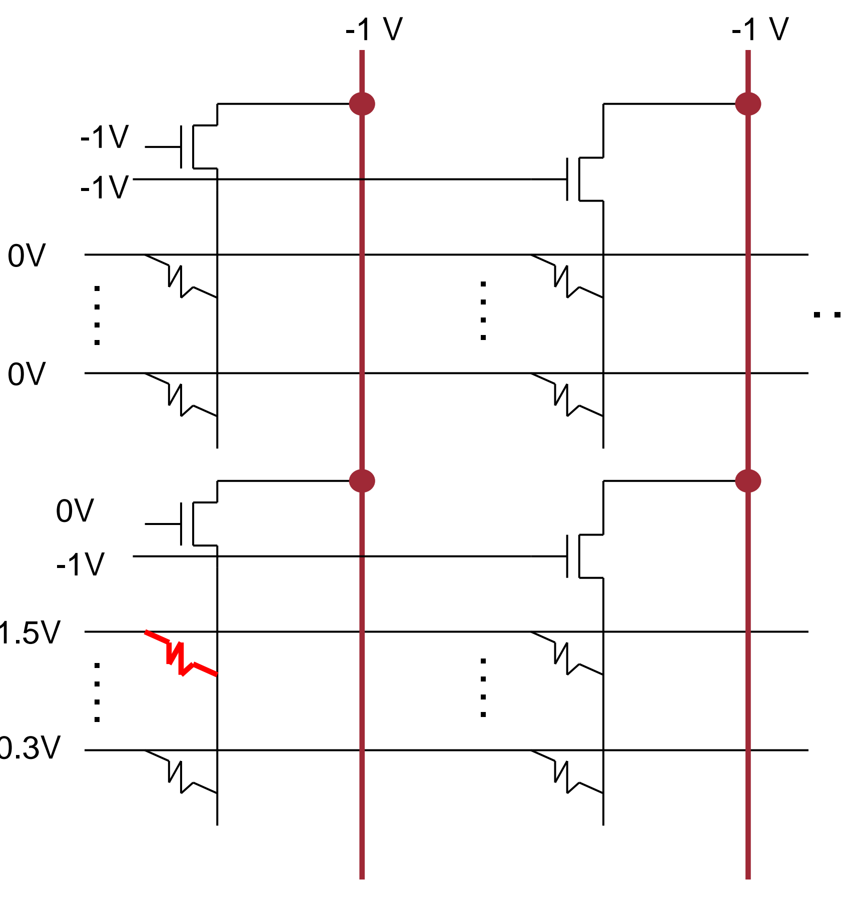
Read



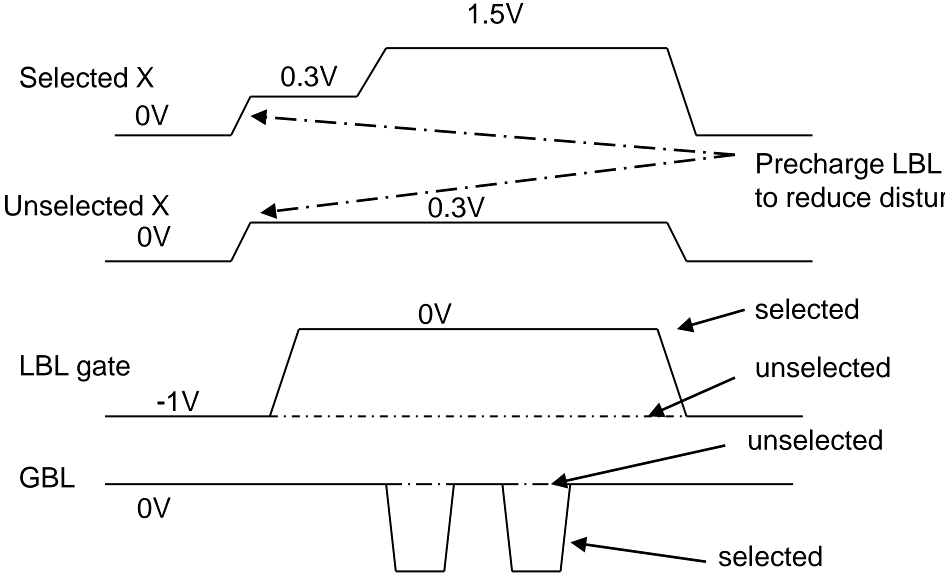
- Selected cells at 0.8V
- Unselected cells at less than 0.2V



Program



- Selected cells at 2.5V
- Unselected cells at 1.5V (BL) or 1V (WL).



Erase

0.3V
0.3V
-0.3V
-0.3V
0V
0V
1.0V
-0.3V
-1.7V
-0.3V

High disturb is a don't care because device will be erased

- Selected cells at 2V
- Unselected cells at 0.6V

Selected X
0V
-1.7 V
Unselected X
0V
1.0V
selected
unselected
LBL gate
0V
G BL
0.3 V

Array with Local Bitlines – Layout

Local Bitline "breaks"

CMOx Cross-point Array

128 Rows = Local Bitline

128 Rows = Local Bitline

VIA drops down to CMOS

NFET select gates underneath

© Unity Semiconductor Corporation

512Gb/1Tb Storage Chip Architecture

(4k x 128 x 4) x 128
= 256Mb per Block

Block

Brick = 128 Blocks

[(4k x 128 x 4) x 128] x 16
= 4Gb per Sub-Plane

Sub-Plane = 16 Bricks

Page is 1 row of cells across 16 Blocks/Bricks

"Erase" Block is 1 row of Blocks across 16 Bricks

Plane 0 (32 Sub-Planes)

Plane 1

Plane 2

Plane 3

PB/$ µC Ana 125MB/s

PB/$ µC Ana 125MB/s

PB/$ µC Ana 125MB/s

PB/$ µC Ana 125MB/s

Timing Master µC I/F

≥500MB/s DDR

(4k x 128 x 4)
x 128
x 16
x 32
= 128Gb per Plane

x 4
= 512Gb (SLC)

© Unity Semiconductor Corporation

Die size estimates

- 20nm cell, 45nm CMOS
- 512Gb 178 mm2

- 15nm cell, 32nm CMOS
- 1Tb 179 mm2

© Unity Semiconductor Corporation

Does it really work?

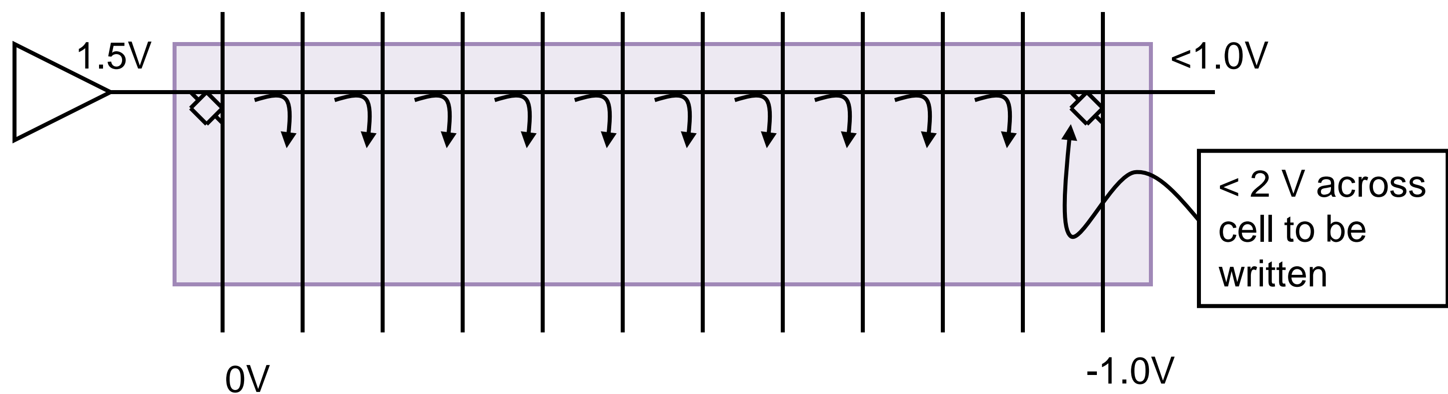
- After ISSCC ‘2010 1Mb arrays presentation:
- “I don’t think you can program a cell at the end of a line without disturbing cells at the beginning of the line”.
- “The cell signal will be lost in the leakage current from other cells”.
- “The total current on the lines will be too high, IR drops will be huge”.

© Unity Semiconductor Corporation

Outline

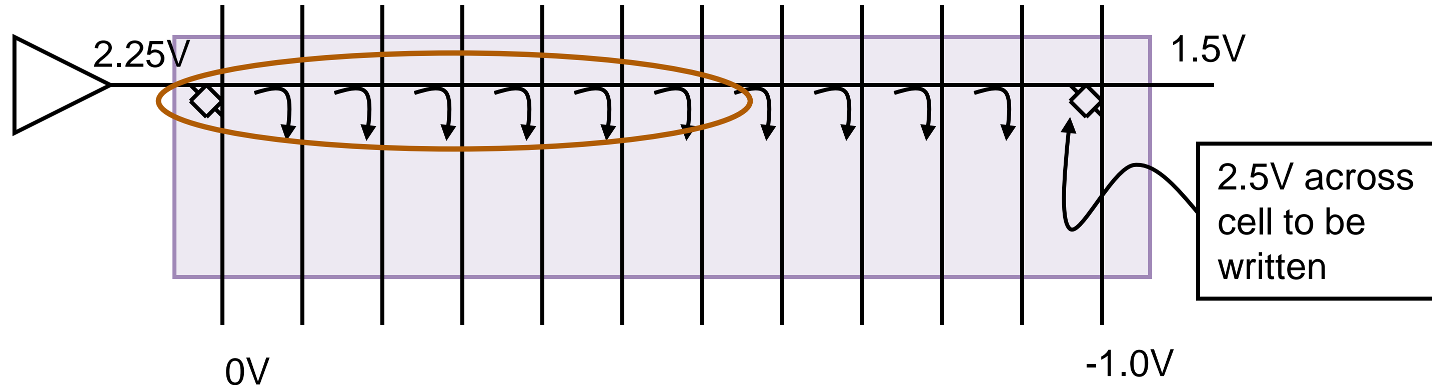
- CMOx technology
- Array construction
- WL length
  - IR drop and unselected WL biasing
- BL length
- Sensing
- Disturb
- Results

Row IR Drop



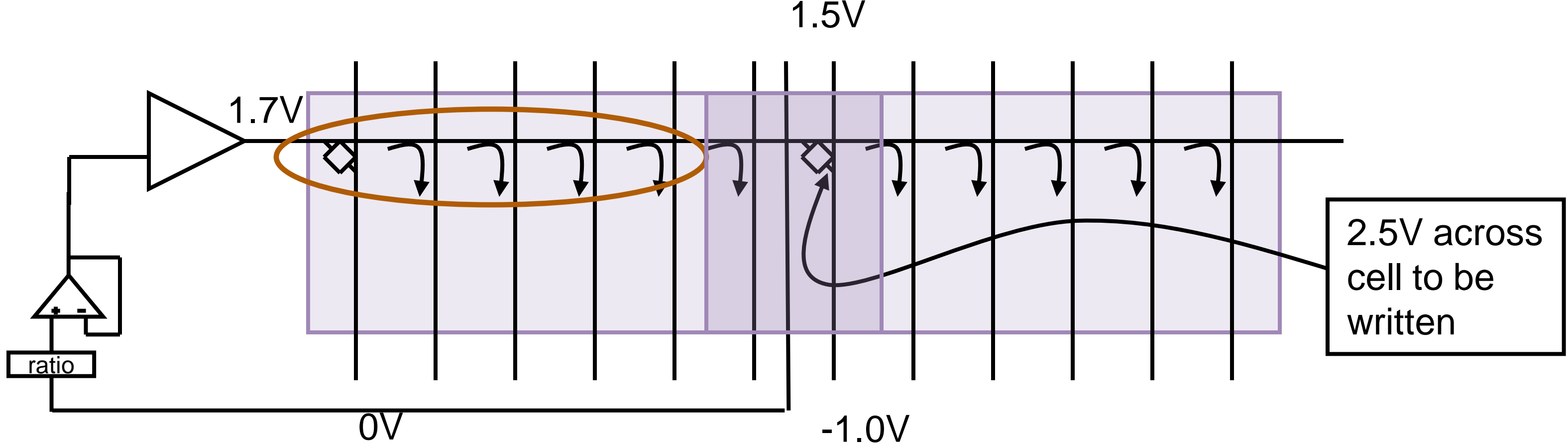
- characteristics
  - 4096 cells on Row
  - 1Ω/   * 8192   = 8.2kΩ
  - I_{prog} = 1uA, I_{half_select} = 26nA @ 1.25v ➔ IR drop over 500 mV at line end
  - Total current at driver is around 200 uA

Row IR Drop with ΔV compensation



– Added voltage creates higher current and parasitic programming in unselected cells

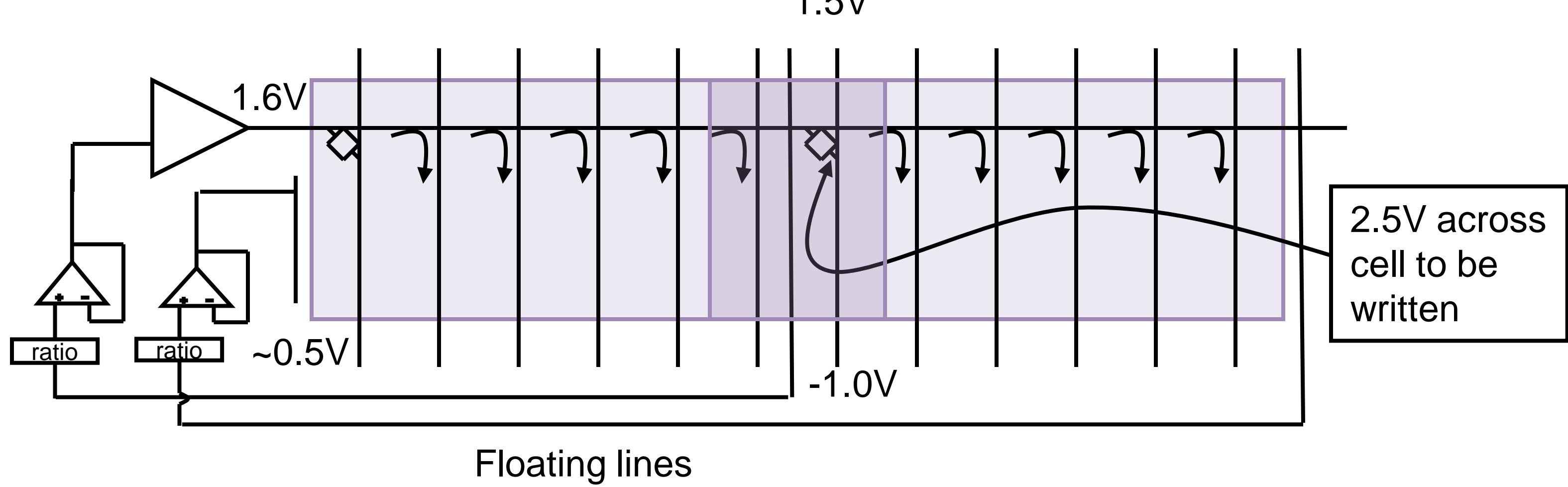
Row IR Drop with ΔV compensation and location adjustment



– Adjust Driver voltage depending on location of cells being programmed
– Added current in cells near driver is still too high

Unity

Row IR Drop with ΔV and location compensation, plus counter bias



- Adjust Driver voltage depending on location of cells being programmed
- Added bias on unselected Word Lines will bias unselected Bit Lines
- Total Word Line current around 100 uA

© Unity Semiconductor Corporation

Simulation on IR drop

/wl_far /wl_near /bl_far /bl_near

IR drop from near to far = 120mV

wl_near

wl_far

bl_near
biased

1.25V
stress
on the
near cell

bl_far
pgm pulse

V (V)

2.0 1.5 1.0 .5 0 -.5 -1.0 -1.5 -2.0

0 5.0 10.0 15.0 20.0

time (us)

5.2738us 294.1mV

Outline

- CMOx technology
- Array construction
- WL length
- BL length
  - Unselected cell leakage and data dependency
- Disturb
- Power and Speed
- Results

CMOx Read with No Select Device

BL
WL

WL
BL
~150mV

CMOx 20nm Cell Model

Selected cell sensing area

Iread = 12nA

Exponential nature of cell IV curve means small “partial select” voltages across unselected cells result in leakage currents orders of magnitude lower than cell read current

Partial select current ~20pA

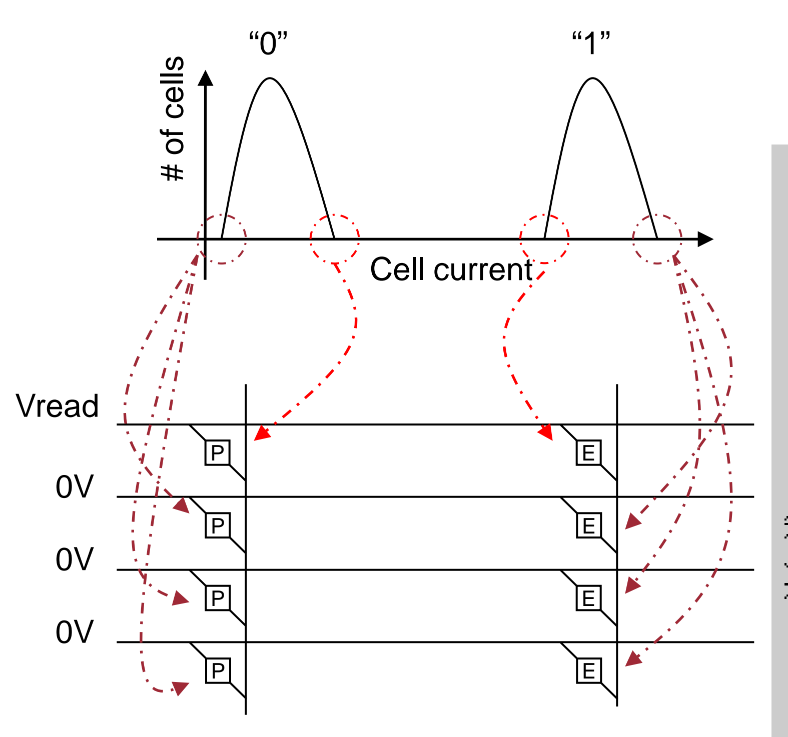
Unselected cell sensing area

Cell Current: 1uA, 100nA, 10nA, 1nA, 100pA, 10pA

Cell Voltage: 0, 0.5, 1, 1.5, 2, 2.5

Measured
Model

Simulation of Data Dependency



Read margin is sufficient, even after all dependencies are accounted for.

Bitline voltage w/ erased cell selected

Margin lost due to data dependency

M0(146.5mV)

M1(60.84mV)

Bitline voltage w/ programmed cell selected

Margin = 85mV

V (mV): 300, 250, 200, 150, 100, 50.0, 0, -50.0

time (us): 0, 25.0, 50.0, 75.0, 100, 125

Outline

- CMOx technology
- Array construction
- WL length
- BL length
- Disturb
- Power and Speed
- Results

© Unity Semiconductor Corporation

Disturb Times during Program and Read

- Array 4096 x 128 x 2 layers
- WL during program: 32 cells programmed at a time (128 write cycles per page)
- 10us pulse x 128 write cycles x 2 layers = 2.6 ms disturb at half write voltage – counter bias
- WL during read: 8 read cycles per page to read all bit lines
- 50 us x 2 x 8 x 100000 read cycles = 80 s disturb at read voltage – 0.2V
- BL during program: 10 us pulse x 255 cells = 2.6ms at ½ write voltage + counter bias
- BL during read: disturb at less than 200 mV is inexistent.

Measured effect of disturbs

Max. Disturbed Cell after Write (WL/BL)

Disturbed Cell after Read

Programmed Cell, gets more programmed by other cells programming
Unselected WL bias can be adjusted to equalize BL/WL disturb

Power and Speed

- Memory + Non Ohmic Select device: 5V -7V operation
- Memory alone: < 3V operation

- No positive charge pump
- Small negative charge pump

➔ much lower total current consumption

- Speed is achieved by parallelism
  - Trading off block size for speed

© Unity Semiconductor Corporation

Single Block Throughput Example

Read 16 rows in 1 Block to fill page buffer

16x50uS = 800µS → 10MB/sec

I/O

8kB Page Buffer

4096b

4096b 4096b 4096b 4096b 4096b 4096b

16x

= 4096b sub-word read from Tile in one 50us sensing cycle

16 Block Throughput Example

Read 1 row in 16 Blocks to fill page buffer
1x50uS = 50uS → 160MB/sec

© Unity Semiconductor Corporation

Outline

- CMOx technology
- Array construction
- WL length
- BL length
- Disturb
- Power and Speed
- Results

Program Mixed Patterns on 4 Kbits arrays without select devices
Total Array Size=32Kbits

IO7 IO0

_Select> - [Ecr1/Dut1/X05 Physical Bitmap]

Window Help X05Menu

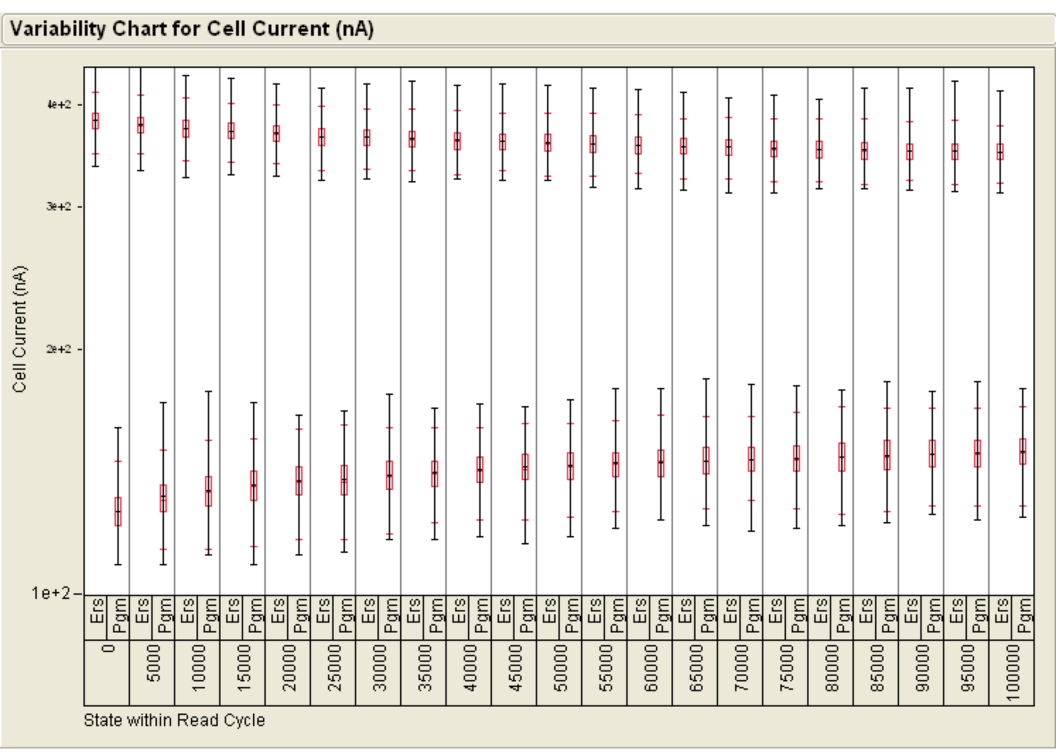
Stop Break User V/I Timing Pattern LVM DBM ECR Bitmap Shmoo LEC Scan Wafermap Help

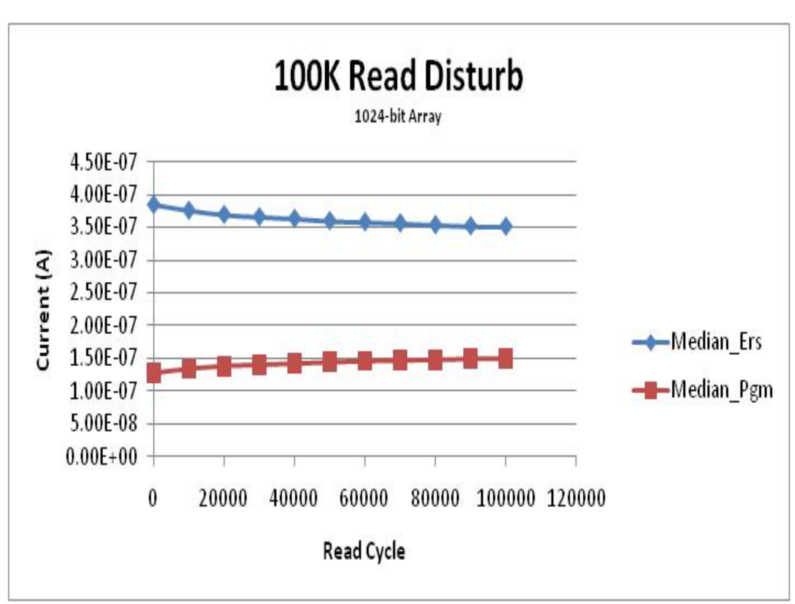
_Select> - [Ecr1/Dut1/X05 Physical Bitmap]

Window Help X05Menu

Stop Break User V/I Timing Pattern LVM DBM ECR Bitmap Shmoo LEC Scan Wafermap Help

CMOx™100Kcycles measured Read Disturb

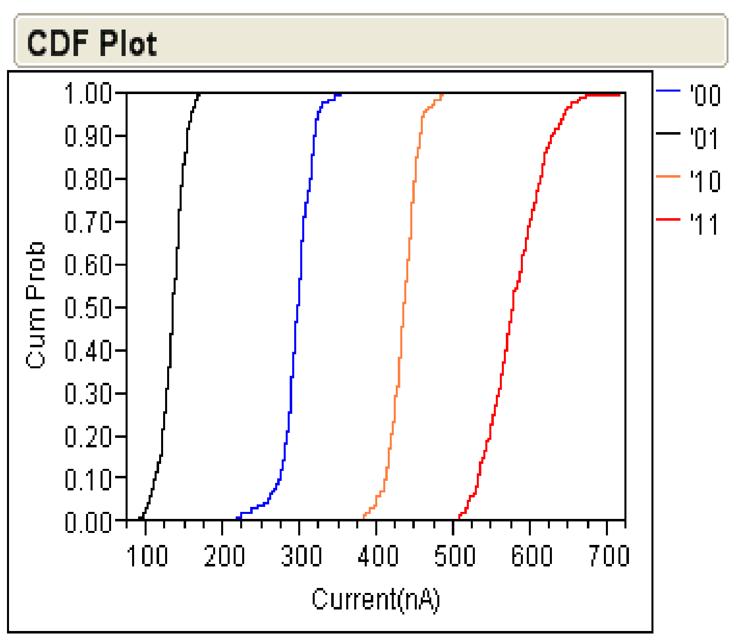




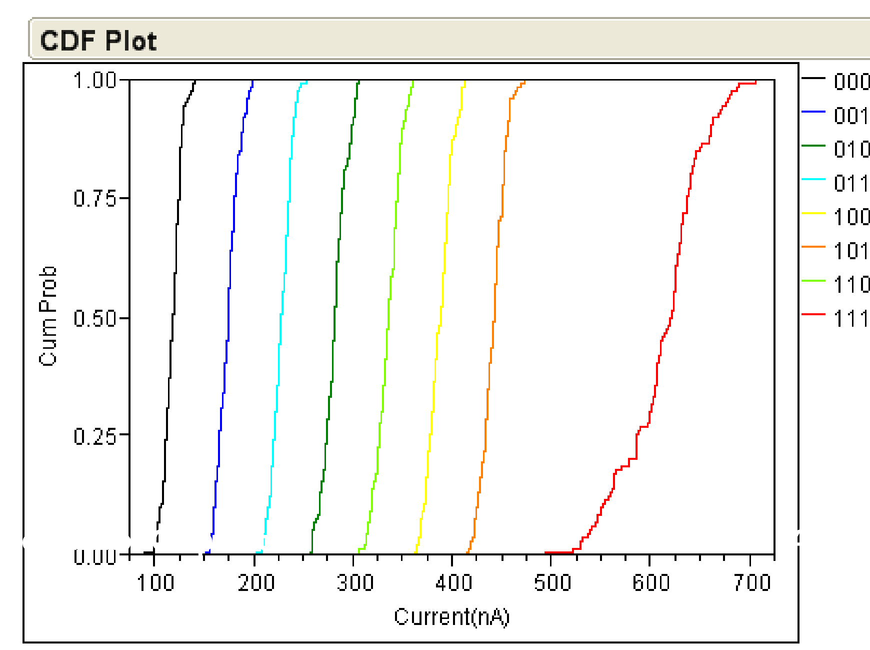
- Read cycle = ~10us/bit @ 1V
- ΔiMedian erase @ 100K reads = -8.8%
- ΔiMedian program @ 100K reads = +18% (includes influence of retention loss)

2bpc & 3bpc Cell Current Distribution

2 bpc



3 bpc



Decoded arrays, bitmaps
after full array program

Why CMOx (and not other RRAMs)?

- Embedded non-linearity allows for cell selection with a 1000:1 cell/leakage ratio, allowing for 256 bit wide arrays
  - Non-linearity in both program and erase levels, unlike RRAM
- Low switching current (< 1uA/cell)
- Progressive, time dependent programming

© Unity Semiconductor Corporation

Conclusions

- An architecture with Local Bit Lines enables CMOx RRAM technology for high density memories.
- The LBL pass gates can be laid out under the arrays with minimal die size increase.
- Effect of IR drops, disturbs and data sensitivities have been taken into account and minimized.
- Increased parallelism due to small arrays enables fast read and write speeds.