# EE 382M.8
# VLSI–II
# Non-Volatile Memory Design
# Spring 2017

**Syed M. Alam (salam@alum.mit.edu)**
**Everspin Technologies, Inc.**

# EE 382M VLSI - II

## Lecture Objectives

1) Understand technology, design, and application interaction
2) Learn fundamentals to pursue self-study and research

## Lecture Outline

- NVM Technology Overview
- NVM Chip Block Diagram
- Key Circuit Concepts Differentiating from SRAM
  - Array Architecture
  - Voltage Boosting and Charge Pump Techniques
  - Write Driver Design
  - Sensing Schemes and Circuit
- NVM Interface and Applications

1st half ↑ / 2nd half ↓

References are listed in the back for further study.

# Semiconductor Memory Types

Semiconductor Memory

Volatile

- RAM
  - DRAM
  - SRAM

Non-Volatile

- Floating Gate
- Nitride
- Emerging
  - Ferro-electric
  - Magnetic
  - Phase Change
  - Polymer
- ROM & Fuse

"NV Ram"

"Unified Mem"

Courtesy Intel

# Non-Volatile Memory History

MRAM, Phase Change, Polymer
2000's

Ferro-electric 1988

MLC NOR Flash
1995

NOR Flash
1988

AND, DiNOR Flash
1990's

EPROM
1971

MLC NAND Flash
1996

NAND Flash
1985

EEPROM
1980

Bipolar ROMS/PROMS
→Late 60's

1970 1980 1990 2000

Courtesy Intel

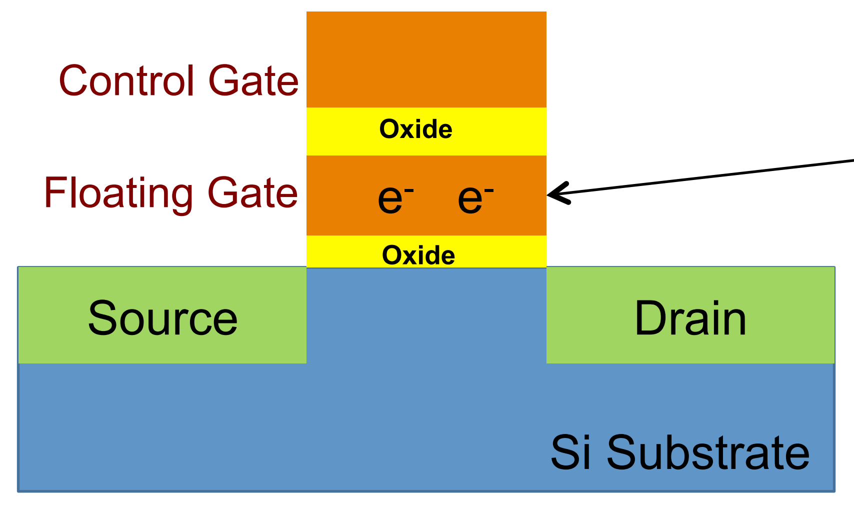
MLC = Multi-Level Cell

| Features | SRAM | PCRAM | STT-RAM | RRAM |
|---|---|---|---|---|
| Nonvolatility | No | Yes | Yes | Yes |
| Memory Cell Factor ($F^2$) | 50-120 | 6-12 | 4-20 | <1 |
| Read Time (ns) | 1 | 20-50 | 2-20 | <50 |
| Write/Erase Time (ns) | 1 | 50-120 | 2-20 | <100 |
| Number of Rewrites | $10^{16}$ | $10^{10}$ | $10^{15}$ | $10^{15}$ |
| Power Consumption – Read/Write | Low | Low | Low | Low |
| Power Consumption – Other than R/W | Leakage Current | None | None | None |

**Key Properties:**

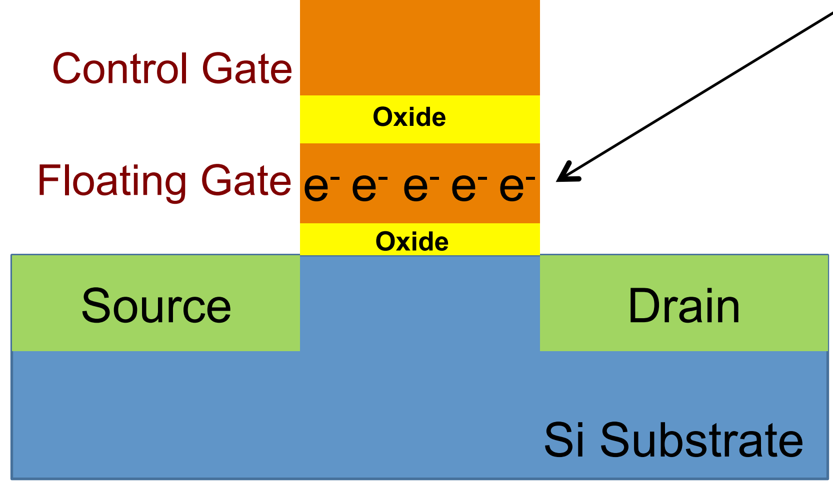
- **Cell size:** Area defined in units of minimum feature (F)
- **Access time:** Read / write latency. Also impacts bandwidth (bits/s)
- **Access mode:** Random byte or block access. Impacts bandwidth.
- **Endurance:** # of writes, program, erase, or read cycles.
- **Retention:** Length of time memory bit holds the state.
- **Power/Energy:** Write energy can be a concern.
- **Cost per bit:** Has many facets including area, technology, test…
- **Non-volatility:** Non-volatility and underlying storage mechanism
- **Scalability:** Key to large density memory and cost reduction.

# Flash Memory Storage

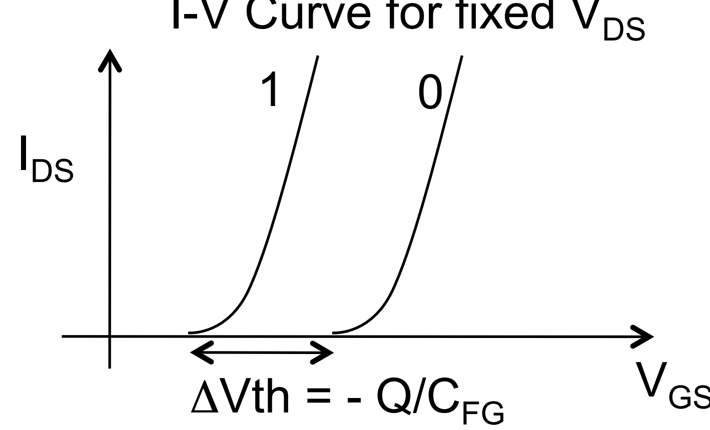



MOSFET device with floating gate

Very low density of trapped electrons
MOSFET $V_{threshold}$ low
Memory state: Logic 1 or high
Erased cell




Very high density of trapped electrons
MOSFET $V_{threshold}$ high
Memory state: Logic 0 or low
Programmed cell




Ref. 1, 3

# Floating Gate Electrons vs. Lithography

Total # Electrons Stored

NAND
NOR

Number of Electrons

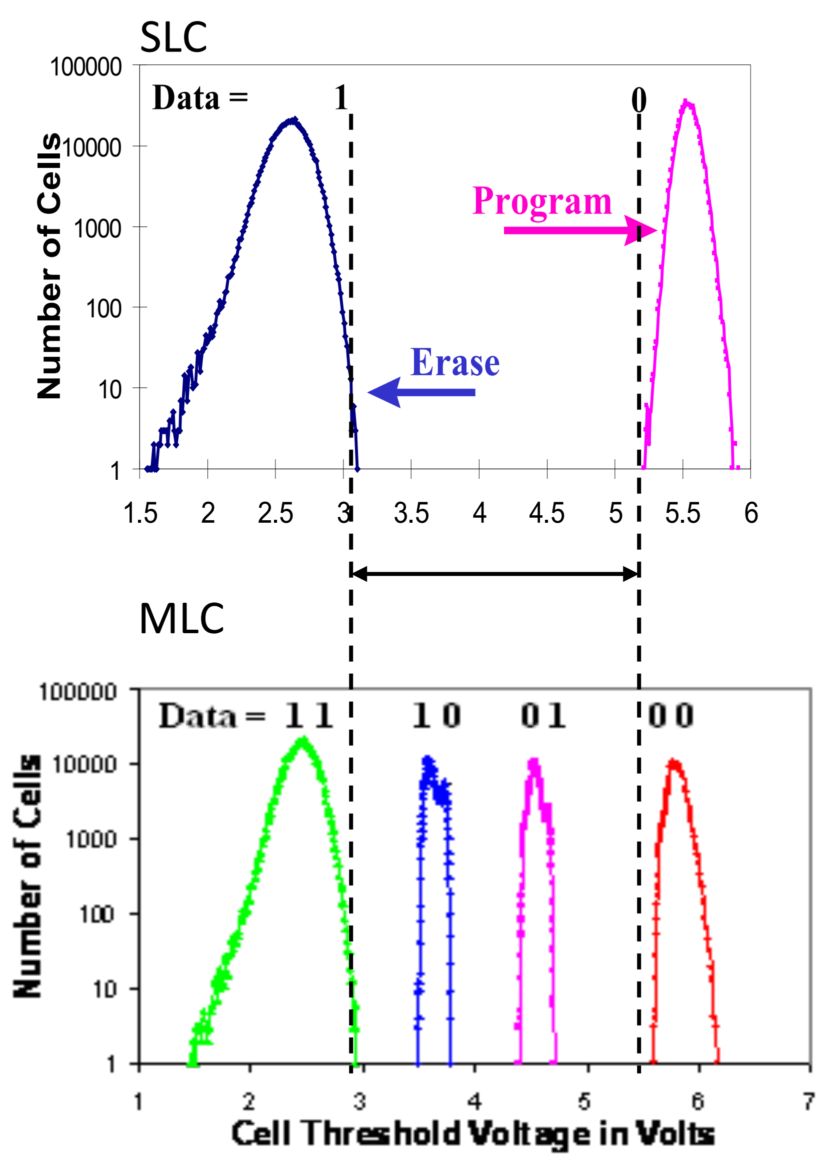
10000
1000
100
10

1
0.1
0.01

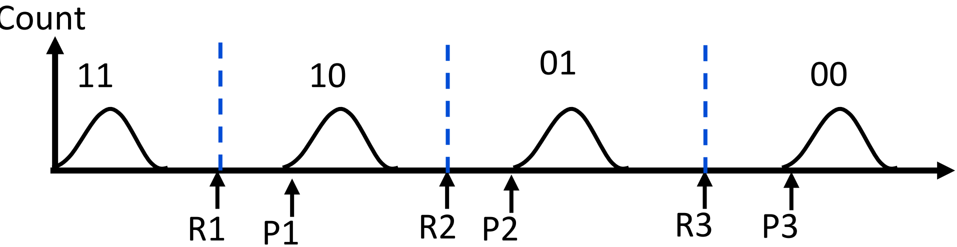
Litho Node

Data Source:
NAND: Kim, et.al., "The Future Prospect of Nonvolatile Memory", VLSI-TSA 2005
NOR: Intel

# Single-Level Cell (SLC) vs. Multi-Level Cell (MLC)




- Take advantage of the threshold voltage difference between the erased and programmed states of the single-level-cell case
  - Two levels = 1 bit/cell
  - Four levels = 2 bits/cell
  - In general: n bits/cell = $\log_2$(#levels)
- Need additional reference cells for program / read
  - One read reference cell for 1 bit/cell
  - Three read reference cells for 2bits/cell
  - N-1 reference cells for n bits/cell
  - Corresponding reference cells for program

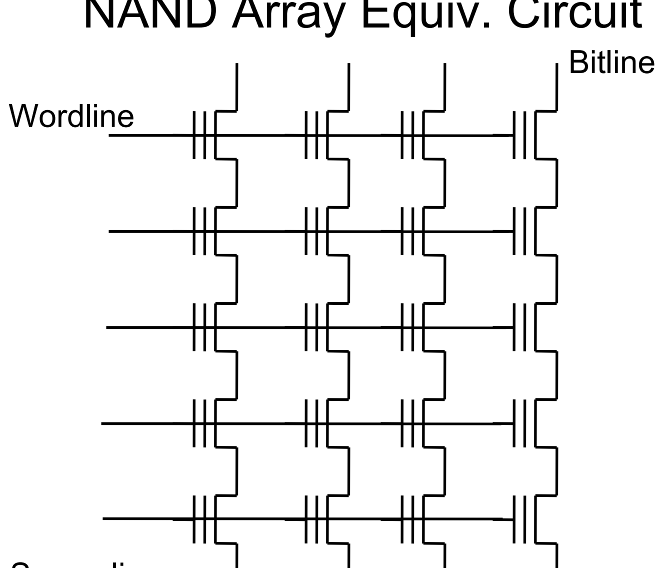



Courtesy Intel

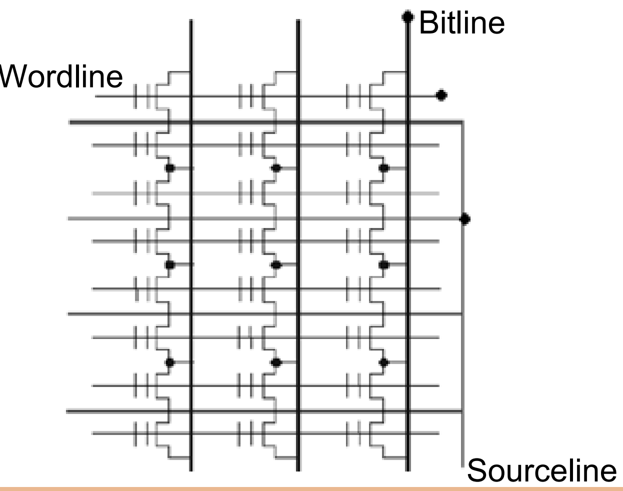
# NAND vs. NOR Flash Array

| | NAND | NOR |
|---|---|---|
| Cell Array | Contact<br>Word line<br>Unit Cell<br>Bit line<br>Source line | Bit line<br>Contact<br>Word line<br>Unit Cell<br>Bit line<br>Source line |
| Layout | 2F<br>2F | 2F<br>5F |
| Cross Section | | |
| Cell Size | $4F^2$ | $10F^2$ |

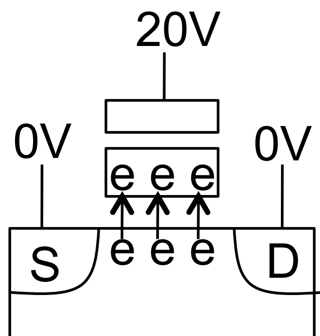
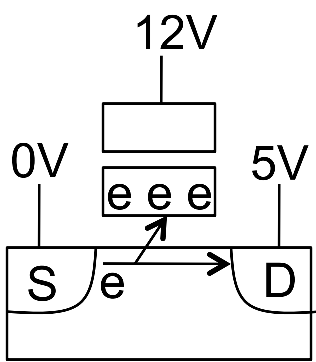
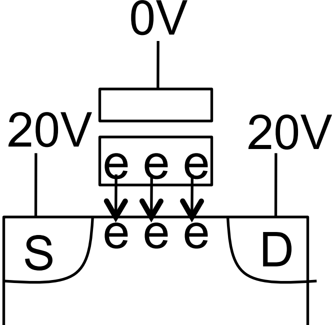
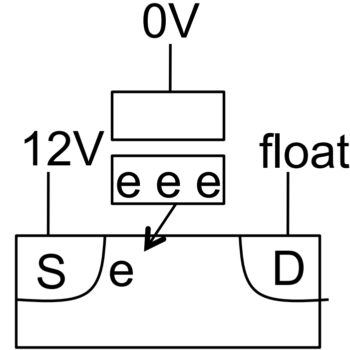
## NAND Array Equiv. Circuit




## NOR Array Equiv. Circuit



# NAND vs. NOR Program and Erase Examples

| | NAND | NOR |
|---|---|---|
| **Program** |  <br>Fowler-Nordheim (FN) Tunneling (low current)<br>e.g. 1 cell = 5nA |  <br>Channel Hot Electron<br>High current<br>e.g. 1 cell = 0.1mA |
| **Erase** |  <br>Uniform FN Tunneling<br>(can also float S,D and raise B to ~20V) |  <br>Local FN Tunneling |

# Flash Key Memory Parameters

Example Micron Flash Memory Products

| Characteristic | NAND Flash: MT29F2G08A | NOR Flash: TE28F128J3 |
|---|---|---|
| Random access READ | 25µs (first byte)<br>0.025µs each for remaining 2111 bytes | 0.075µs |
| Sustained READ speed (sector basis) | 26 MB/s (x8) or 41 MB/s (x16) | 31 MB/s (x8) or 62 MB/s (x16) |
| Random WRITE speed | ≈ 220µs/2112 bytes | 128µs/32 bytes |
| Sustained WRITE speed (sector basis) | 7.5 MB/s | 0.250 MB/s |
| Erase block size | 128KB | 128KB |
| ERASE time per block (TYP) | 500µs | 1 sec |

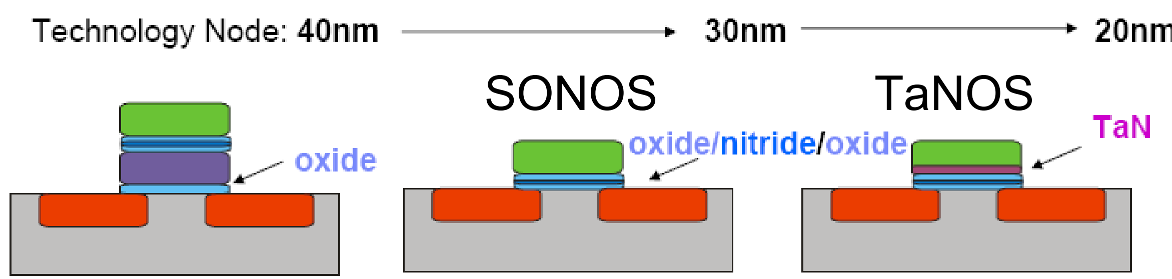
**Access Time**

**Access Mode**

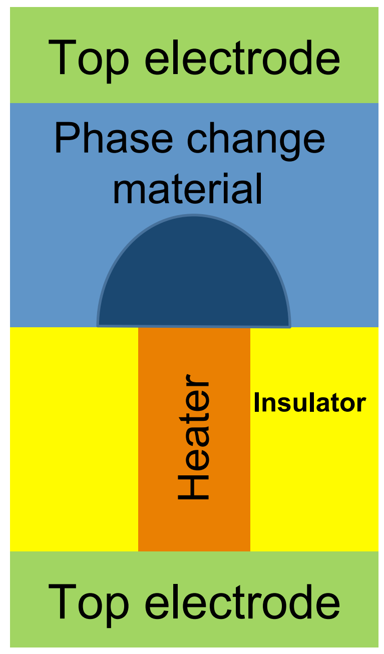
**Retention:** Bit flipping due to leakage, radiation.

**Endurance:** Fail mechanism is oxide forming electron traps with cycling. ~$10^6$ cycles. Apply wear leveling and ECC to increase lifetime.

**Scalability:** Tunnel oxide thickness scaling is the limit. Multi-level cell and new materials for scaling.



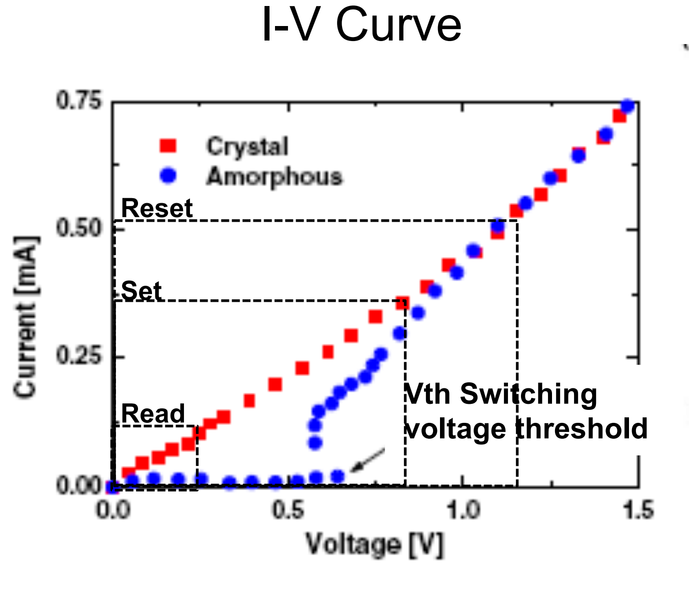
# Phase Change Memory (PCM)



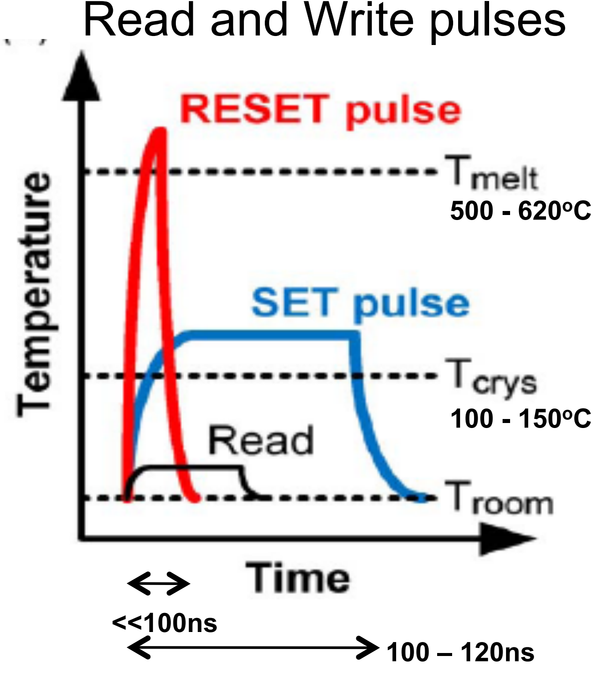

| Phases: | **Crystalline** | **Amorphous** |
|---|---|---|
| Resistance: | Low (kΩ) | High (MΩ) |
| Logic state: | Logic 0 | Logic 1 |
| Programming Methods: | Set current pulse<br>Anneal $T_{crys}<T<T_{melt}$ | Reset current pulse<br>Melt->cool fast |

Chalcogenide sandwich
$Ge_2Sb_2Te_5$ : GST
Germanium Antimony
Tellurium

I-V Curve




Read and Write pulses




Ref. 4, 5

# PCM Key Memory Parameters

**Endurance:** Stuck open or short due to stress and void formation at bottom interface upon cycling. $10^9$--$10^{12}$ cycles of endurance

**Retention:** Bit flipping due to thermal disturbance. 10 years retention of crystalline state requires operation below 110°C.

**Scalability:**

Write current reduces with smaller geometry !!

$\Delta T = P \times R_{th}$ where $P = V_{GST} \times I_{GST}$

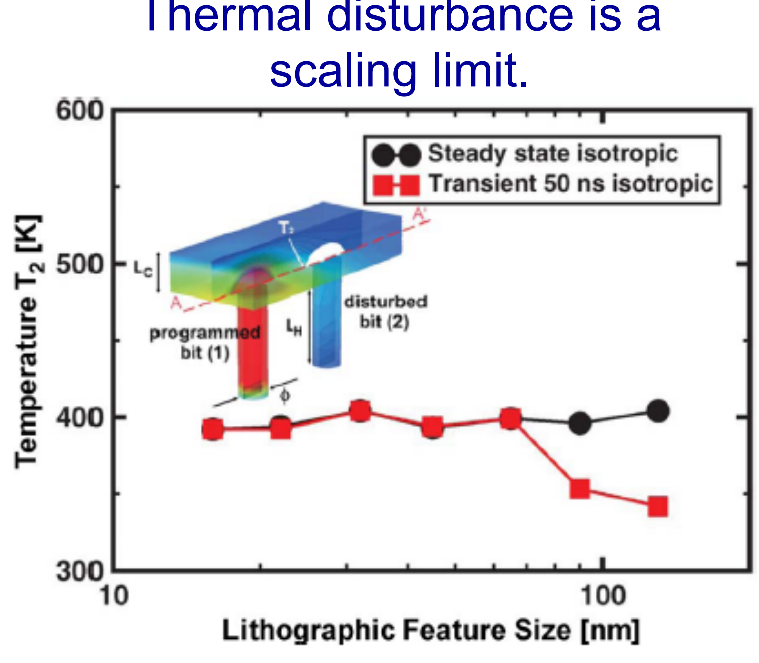
Therefore $I_{GST} = \Delta T \; / (V_{GST} \times R_{th})$

$R_{th} \; \alpha$ *contact area*

$\Delta T$ temperature rise for writing
$P$ joule heating power
$R_{th}$ thermal R to metal sink
$V_{GST}$ voltage across GST cell
$I_{GST}$ current through GST cell

$\Delta T$ and $V_{GST}$ are physical quantities that do not scale.

Thermal disturbance is a scaling limit.




Constant $V_{GST}$ can be a CMOS reliability limit.

# Resistive Random Access Memory (RRAM)

Metal Insulator Metal
MIM structure

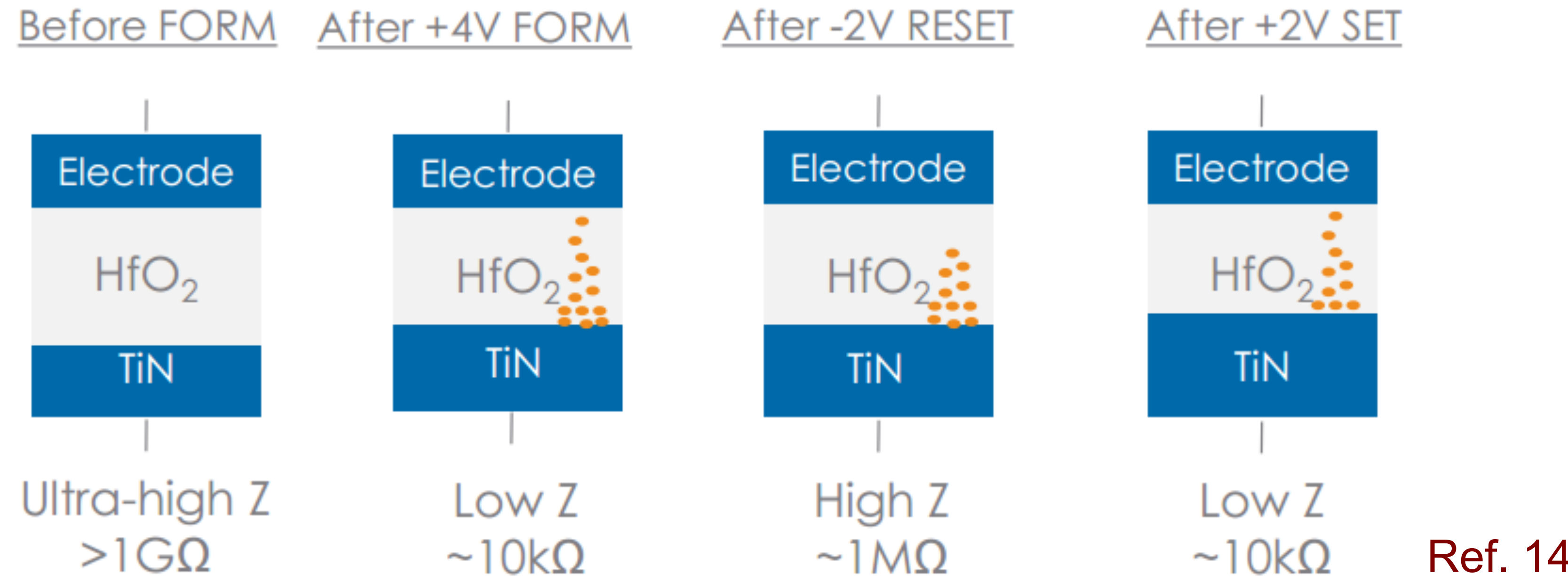
| Top electrode |
|---|
| Transition Metal Oxide |
| Bottom electrode |

| | Examples |
|---|---|
| Top electrode | Pt, TiN/Ti, TiN, Ru, Ni … |
| Transition Metal Oxide | $TiO_x$, $NiO_x$, $HfO_x$, $WO_x$, $TaO_x$, $VO_x$, $CuO_x$ , … |
| Bottom Electrode | TiN, TaN, W, Pt, … |

## Basic Idea:

- Normally insulating dielectric (metal oxide) is made to conduct through a filament formed by applying a sufficiently high voltage.
- Filament is reset or broken for high resistance
- Filament is set or formed for low resistance

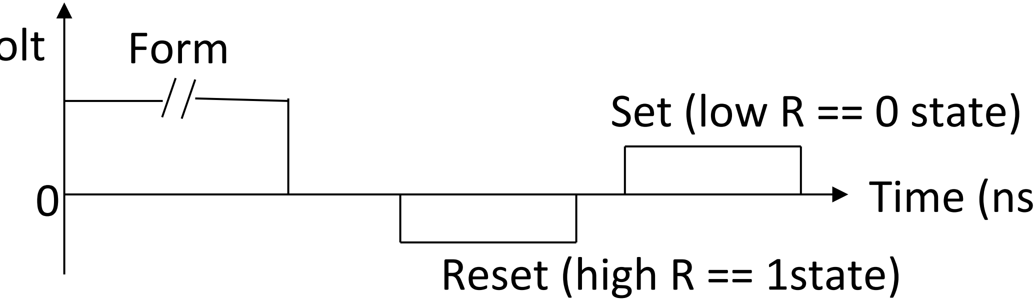
# RRAM Switching Mechanism Example




Ref. 14

Oxygen vacancies are created by hot electrons. Conductive filament is made of Oxygen vacancies.

Bipolar switching
Set / Reset pulse widths are 5-100ns



# RRAM Key Parameters

| Bipolar RRAM Devices | Repeatable Results |
|---|---|
| Write Voltage | <2.5V |
| Write Current | ~20-100uA |
| Switching Time | <10ns |
| Endurance | $10^{6}$ ($10^{10}$ reported in IEDM) |
| Data Retention at 85C | 1 year with 20-50uA<br>10 years with 100uA |

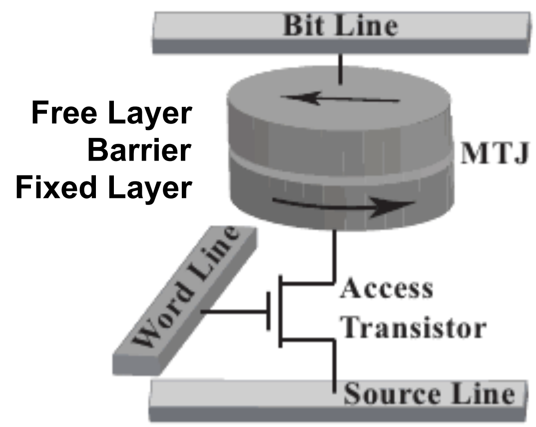
Rambus RRAM Status [14]:

# RRAM Device Specs from reported literature

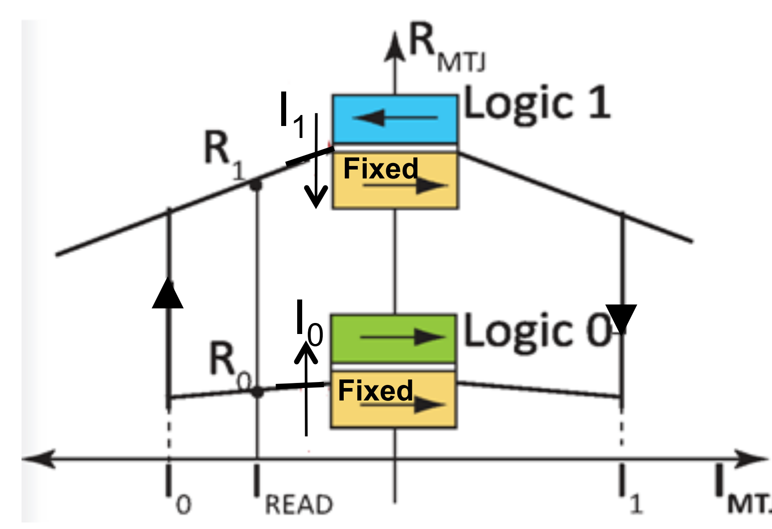
| | ITRI, IEDM 2008 | NEC, VLSI 2010 | Panasonic, IEDM 2008 | Univ. + IMEC, IMW 2010 | Fujitsu, IEDM 2007 |
|---|---|---|---|---|---|
| Device | TiN/Ti/$HfO_x$/TiN | Ru/$TiO_x$/$TaO_x$/Ru | Pt/$TaO_x$/Pt | Au/$NiO_x$/TiN | Pt/Ti-doped NiO/Pt |
| Test chip | 1T-1R | 1T-1R | 1T-1R | 1T-1R | 1T-1R |
| Polarity | Bipolar | Unipolar | Bipolar | Unipolar | Unipolar |
| Reset | 2V, 25uA | 0.65V, 200uA | 1.5V, 100uA | 0.5V DC, 9.5uA | 1.9V, 100uA |
| Set | 2.3V | 2.8V | 2V | 2.7V DC | 2.8V |
| Form Voltage | 3V | ? | ? | 3.7V DC | 3V |
| Switching Time | <10ns | <1us | <100ns | NA | 10ns |
| On/off ratio | ~100x | 100x | 10x | 5x-10x | 90x |
| Endurance, Data Retention | $10^6$, 10 years | $10^5$, 10 years | $10^9$, 10 years | 130 cycles, ? | 100, 10 years |
| Comments | Typical data | Worst case data | Typical data | Typical data | Typical |

# Magnetoresistive Random Access Memory (MRAM)

MRAM Bitcell




R-I Curve




- Sandwich of magnetic (CoFeB ) fixed and free layers with oxide (AlO or MgO) barrier in between
- Magnetic Tunnel Junction (MTJ) with
  Anti-parallel: Logic 1 – high resistance
  Parallel: Logic 0 – Low resistance
- Spin Transfer Torque (STT) current based switching between logic states
  - Direct torque transfer (Logic 1 to 0 switch)
  - Reflective torque transfer (Logic 0 to 1)
- TMR (tunnel magnetoresistance) read

$$TMR = \frac{R_1 - R_0}{R_0}$$

Typical TMR with MgO barrier is >100%

# MRAM Bitcell

Metal 5 Cu
Metal 4 Cu
Via 3
Metal 3
Via 2
Metal 2
Via 1
Metal 1
Contact
Bit cell

MRAM module

Magnetic Tunnel Junction (MTJ)

Metal 5
MTJ
Metal 4

No Impact on CMOS !

Cross-sectional view

BL1
$R_M$
DL
WL
GND

Circuit

MTJ Layers

$R_{LOW}$
Storage Layer
Tunnel Barrier
Reference Layer
$R_{HIGH}$
Parallel = Low Resistance
Anti-Parallel = High Resistance
(a)

MTJ Resistance
Reference Layer
MTJ Voltage
(b)

MTJ R vs. V

$R_{LOW}$
Parallel
MTJ Resistance
Magneto-resistance Ratio (MR)
Low State
MTJ Voltage
(b)

# Spin-Torque MRAM – Next Generation MRAM

Current Generation MRAM uses a magnetic field for switching

*Limits scaling due to constant magnetic field*

**Next generation MRAM enables scaling to Gb densities**

**Toggle Write**

H-field

write current

Free layer

Tunnel barrier

Fixed layer

H-field

write current

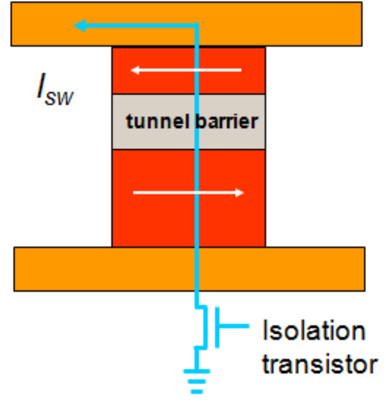
**Spin-Torque Write**

Free layer

Tunnel barrier

Fixed layer

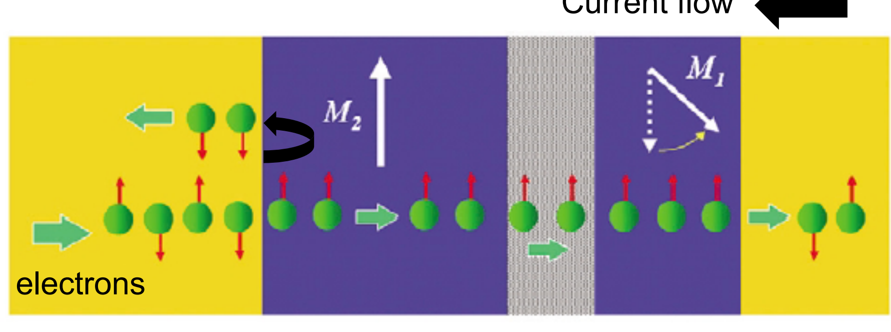
# Writing in Spin-Torque MRAM



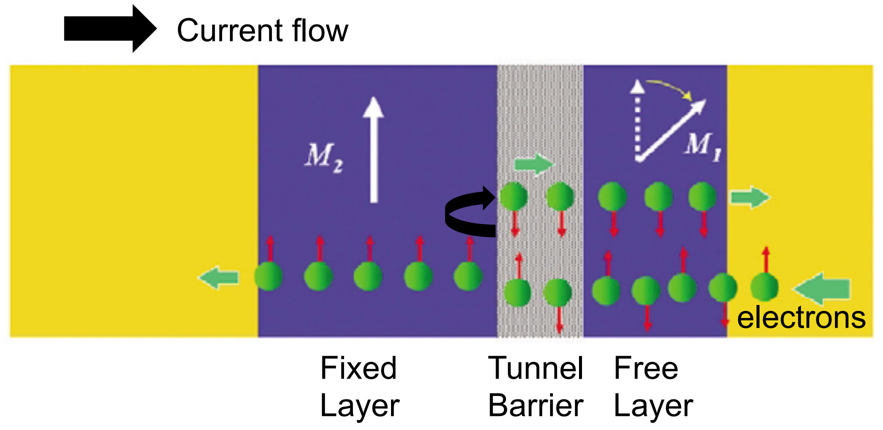

Bi-directional current through the MTJ

Advantages:

- Switching current scales with smaller cell size.
- Lower write current (100s uA instead of >10mA per axis in Toggle MRAM)




AP-to-P Switching (Write 0)
Direct Spin-torque transfer



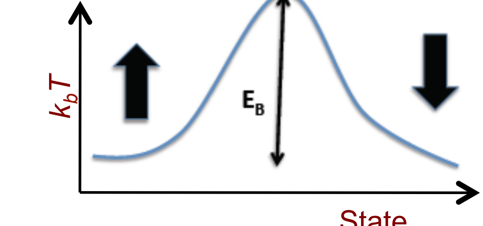

P-to-AP Switching (Write 1)
Reflected electron
Spin-torque transfer

Zhu et al. materialstoday 9, 36 (2006)

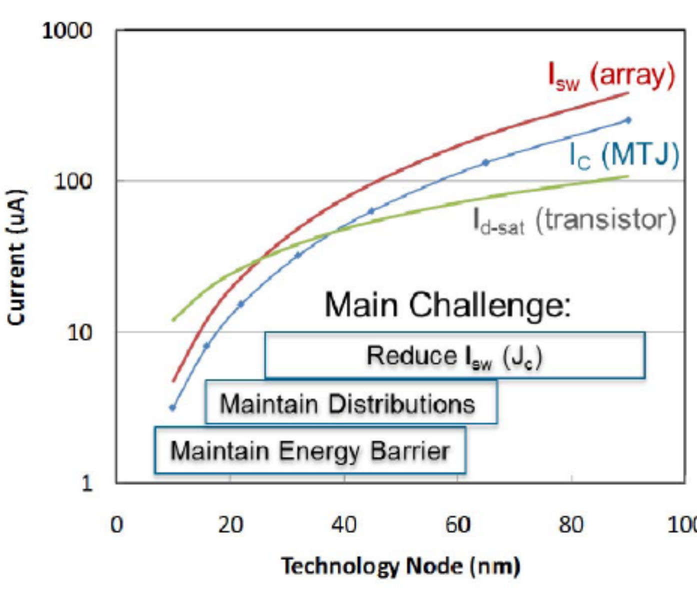
# MRAM Key Memory Parameters

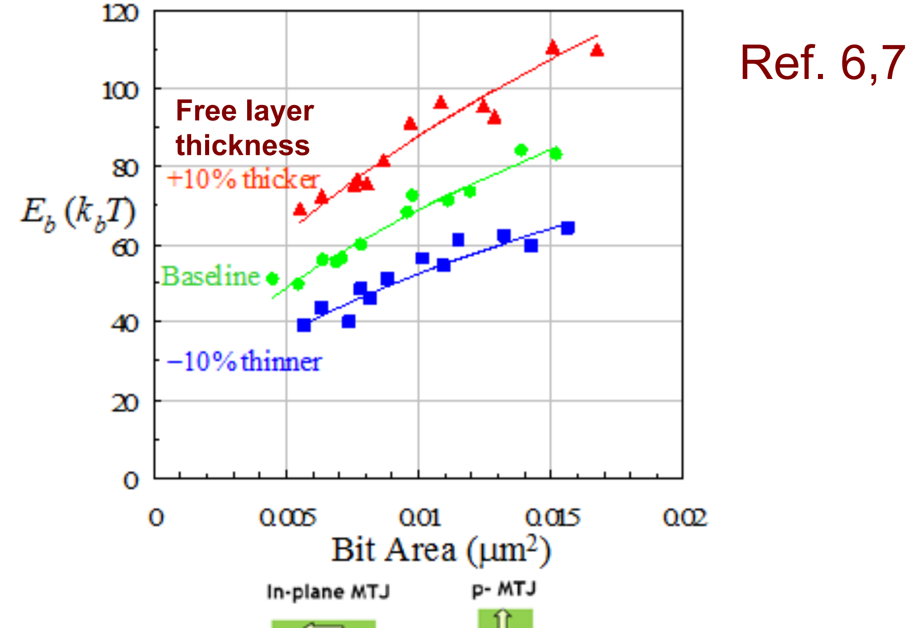
**Endurance**: Stuck-at low due to time-dependent dielectric breakdown. $>10^{12}$ cycles of endurance expected. Need separation between *Vc* and *Vbd distributions*

**Retention:** Bit flipping due to low energy barrier to switching, $E_b$.




**Scalability:** Smaller $A$ → lower $V_c$ → lower $E_b$ → wider $R_{MTJ}$ distributions




Ref. 6,7

6F² cell is possible with scaling

Research to for increased $E_b$

# Storage Capacity Trend

Non-Volatile Memory Trend

Storage Capacity[Mb]

1000000
100000
10000
1000
100
10
1

2000 2002 2004 2006 2008 2010 2012 2014

Cond. 1Mb
ISSC, VLSI Circuits, ASSCC, IEDM,
VLSI Tech.

NAND Flash
128Gb
32Gb
8Gb
RRAM
PRAM
128Mb
FeRAM
64Mb
MRAM

Storage capacity trend of emerging non-volatile memories.

*Source: ISSCC*

# Summarizing Key Properties from ITRS

| | Volatile | | | Non-volatile | | | | |
|---|---|---|---|---|---|---|---|---|
| | SRAM | DRAM | | Flash | | FeRAM | STT-MRAM | PCM |
| | | Stand-alone | Embedded | NOR Embedded | NAND Stand-alone | | | |
| **Storage mechanism** | Inter-locked logic state | Charge on a capacitor | | Trapped charge in floating gate or gate insulator | | Polarization in a ferroelectric capacitor | Magnetization of ferromagnetic layers | Amorphous and crystalline phases |
| **Cell element** | 6T | 1T1C | | 1T | | 1T1C | 1T1R | 1T1R / 1D1R |
| **Feature size, F, nm** | 45 | 36 | 65 | 90 | 22 | 180 | 65 | 45 |
| **Cell area F^2** | 140 | 6 | 12 - 30 | 10 | 4 | 22 | 20 | 4 |
| **Read time** | 0.2 ns | <10 ns | 2 ns | 15 ns | 0.1 ms | 40 ns | 35 ns | 12 ns |
| **Write / Erase time** | 0.2 ns | <10 ns | 2 ns | 1 us / 10 ms | 1 ms / 0.1 ms | 65 ns | 35 ns | 100 ns |
| **Retention** | - | 64 ms | 4 ms | 10 y | 10 y | 10 y | > 10 y | > 10 y |
| **Endurance cycles** | > 1E16 | > 1E16 | > 1E16 | 1E5 | 1E4 | 1E14 | > 1E12 | 1E9 |
| **Write operating voltage V** | 1 | 2.5 | 2.5 | 10 | 15 | 1.3 - 3.3 | 1.8 | 3 |
| **Read operating voltage V** | 1 | 1.8 | 1.7 | 1.8 | 1.8 | 1.3 - 3.3 | 1.8 | 1.2 |
| **Write energy J/bit** | 5E-16 | 4E-15 | 5E-15 | 1E-10 | 2E-16 | 3E-14 | 2.5E-12 | 6E-12 |

**Emerging technology: ReRAM (resistive switching in metal-insulator-metal)**

Ref. 2

# Lecture Outline

- NVM Technology Overview
- NVM Chip Block Diagram
- Key Circuit Concepts Differentiating from SRAM
  - Array Architecture
  - Voltage Boosting and Charge Pump Techniques
  - Write Driver Design
  - Sensing Schemes and Circuit
- NVM Interface and Applications

# NVM Chip Generic Block Diagram

**Chip is very similar to a System-on-chip (SOC).**

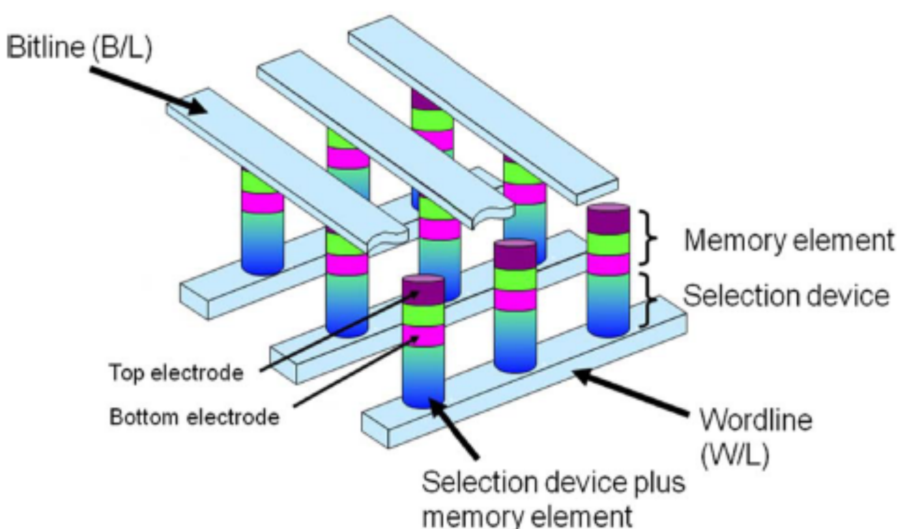


Key Characteristics:

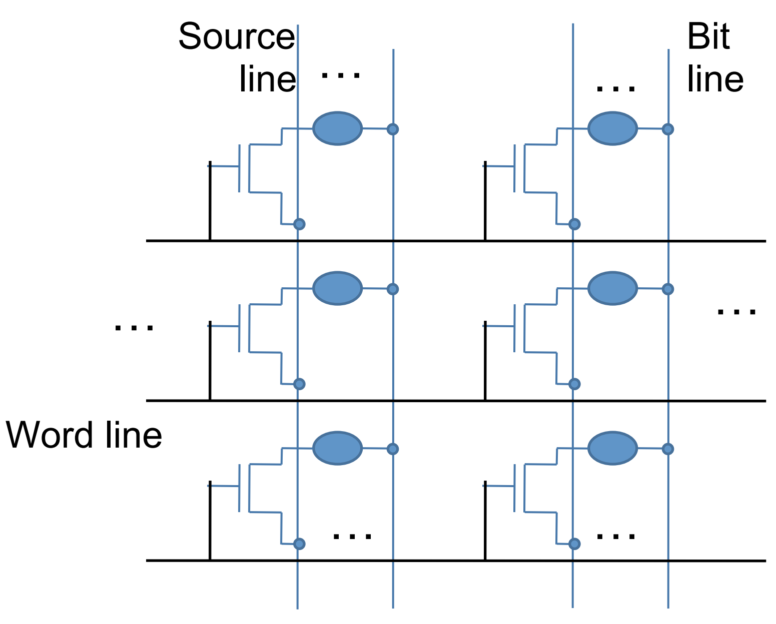
- Multiple voltage domain
- Analog, mixed-signal, and digital
- Regular and area efficient layout

# Memory Array and Cell Selector Choices

Cross-point Array




MOS/BJT select-device Array

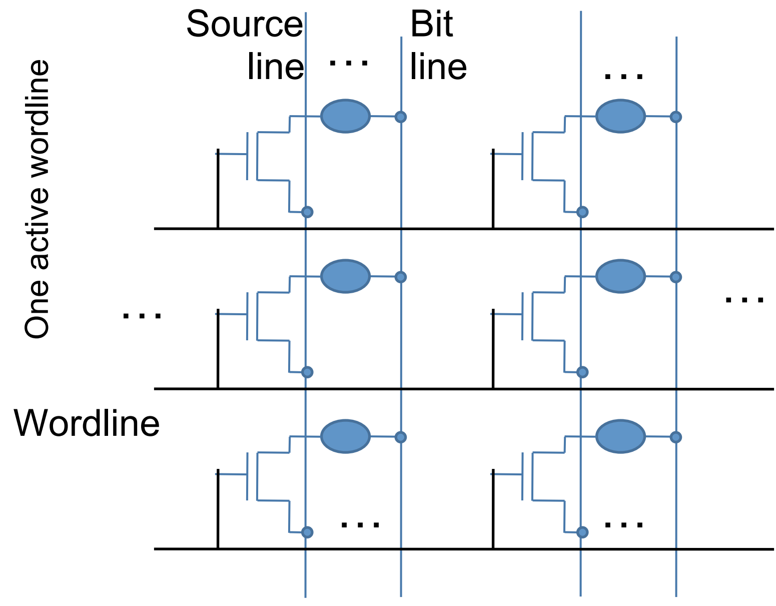



**Array density is mostly determined by memory cell selector size !!**

| | **PN diode** | **MOSFET** | **BJT** |
|---|---|---|---|
| Bidirectional current: | No | Yes | No |
| Current drive: | 8-30 MA/cm$^2$ | 800uA-1.2mA/um-W | 8-30 MA/cm$^2$ |
| Cell size: | ~4F$^2$ | 17-22F$^2$ | 5-8F$^2$ |

# Array Design Considerations

MOS/BJT select-device Array




What determines bitcell size in an array?

- Bit dimension and spacing
- Access transistor drive strength
- Layout spacing rules

How many bitcells on an active bitline / sourceline pair?

- $I_{subthreshold}$ of OFF transistors
- RC delay of the bitline and sourceline

How many bitcells on an active wordline?

- RC delay of the wordline

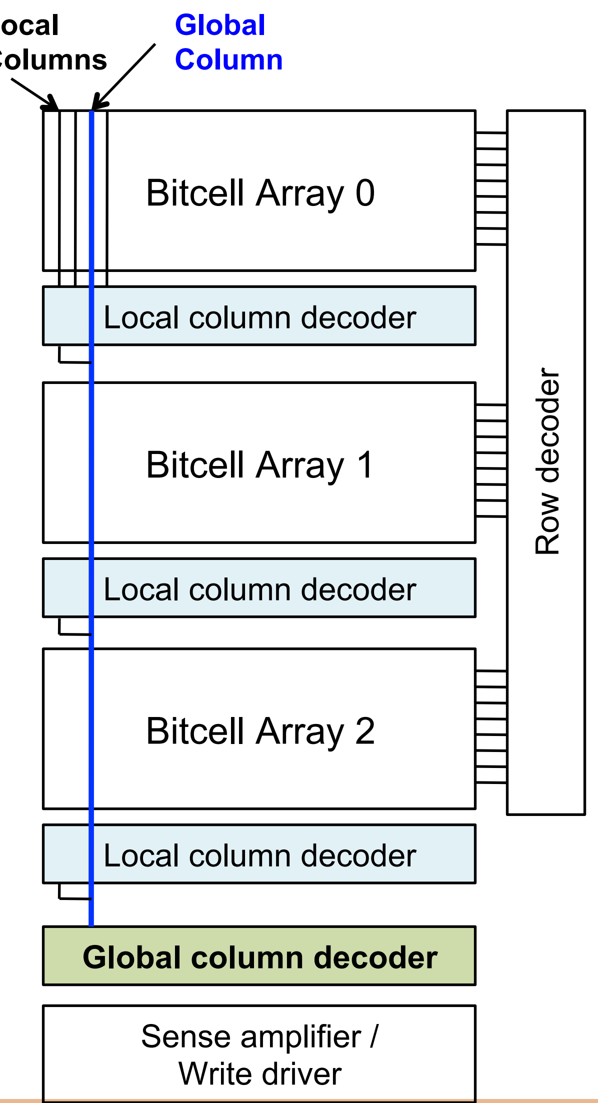
Many array design considerations are independent of storage cell technology.

SRAM array design concepts and general techniques can be applied for:

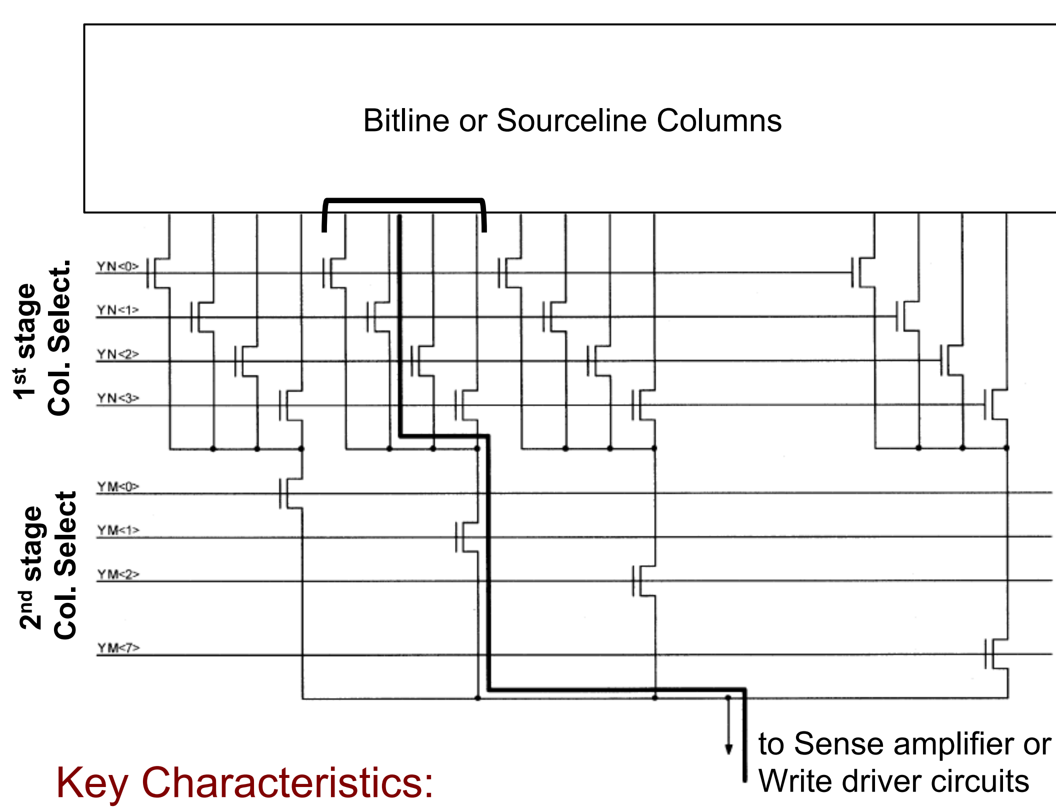
- Physical design and layout (exclusive of the storage cell layout)
- RC delay modeling of wordline, bitline and sourceline.

# Memory Column Selection Scheme

Global-local scheme increases bitcell array to circuit area ratio




Schematic of a two-stage Column Selection Circuit

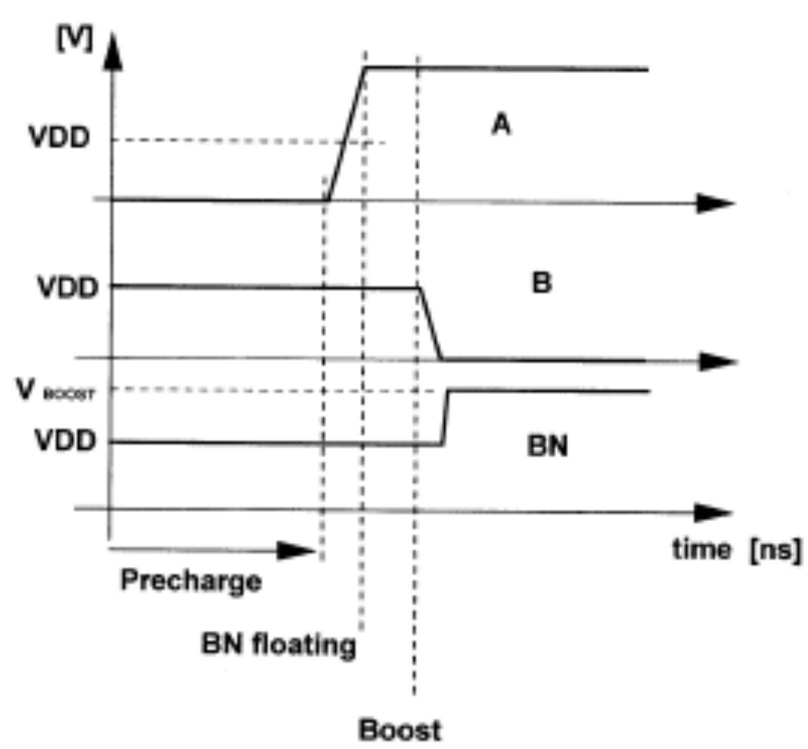


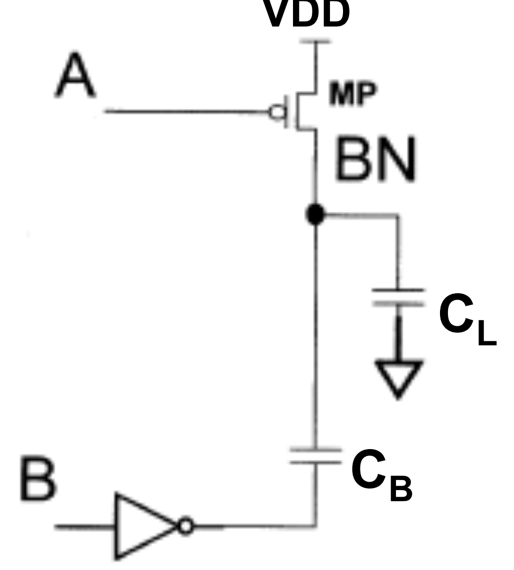

Key Characteristics:

- Two-stage reduces capacitive load of column decode signals and column circuit
- Pitch-matched layout with bitcell columns
- Prefer NMOS transistor . Why?

Drawback: higher resistance in column path

# Voltage Boost Technique: Basic Principal




Why need

$$V_A \geq V_{BN}$$

during boost?

Due to charge conservation
Initial charge = final charge

$$Q_i = Q_f$$

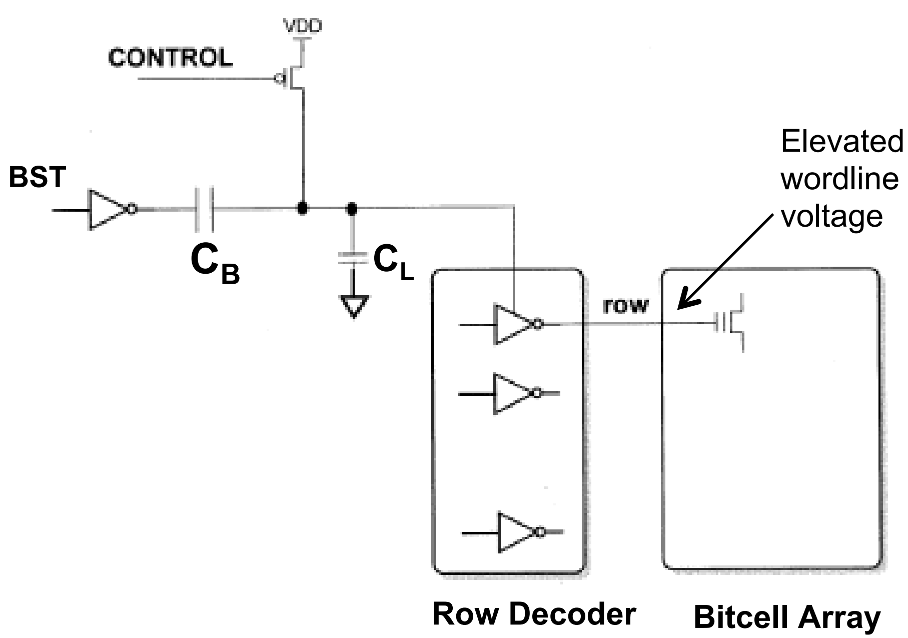
$$Q_i = (C_B + C_L) \cdot VDD$$

$$Q_f = C_B \cdot (V_{BN} - VDD) + C_L \cdot V_{BN}$$

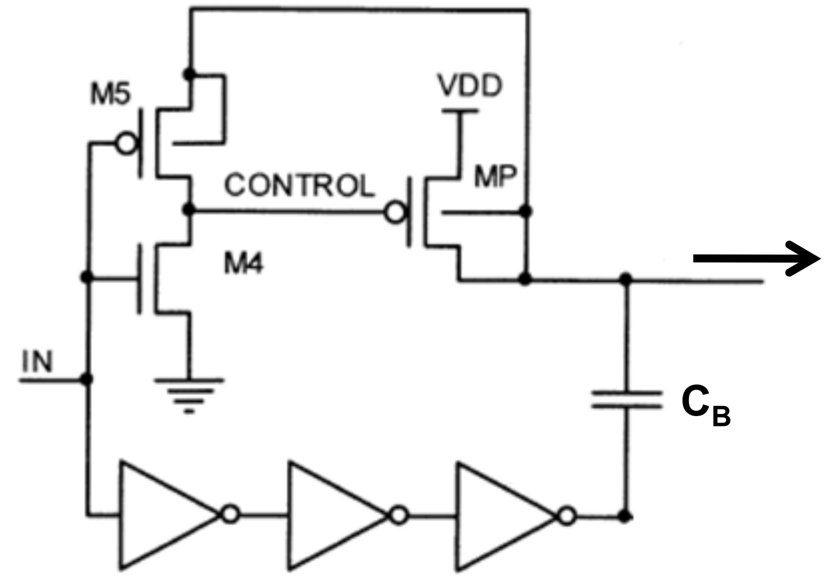
$$V_{BN} = VDD + \frac{C_B}{C_B + C_L} \cdot VDD$$

# Voltage Boost Techniques in Use

Boosted voltage and charge pumped voltage are used in row (word line) decoders for reading and writing.

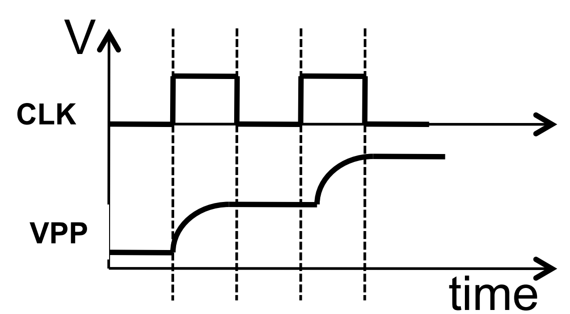



A boosted voltage supply generator




One shot boost circuit for row decoders:

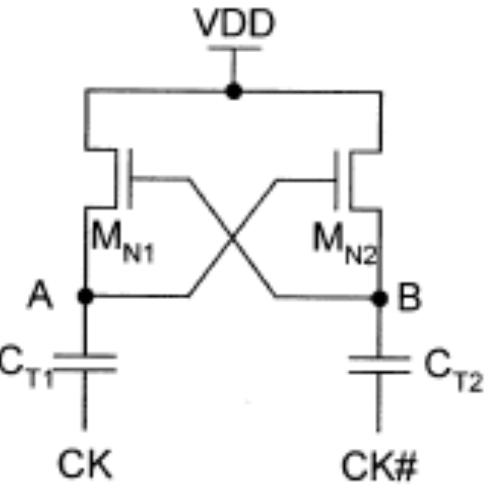
- Find required $C_B$ for $C_L$. $C_B$ can be large!
- Can use local boost circuit per row.

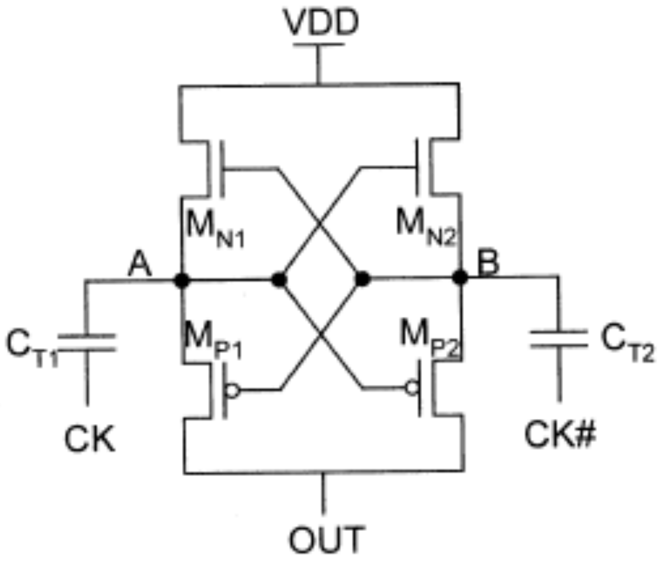


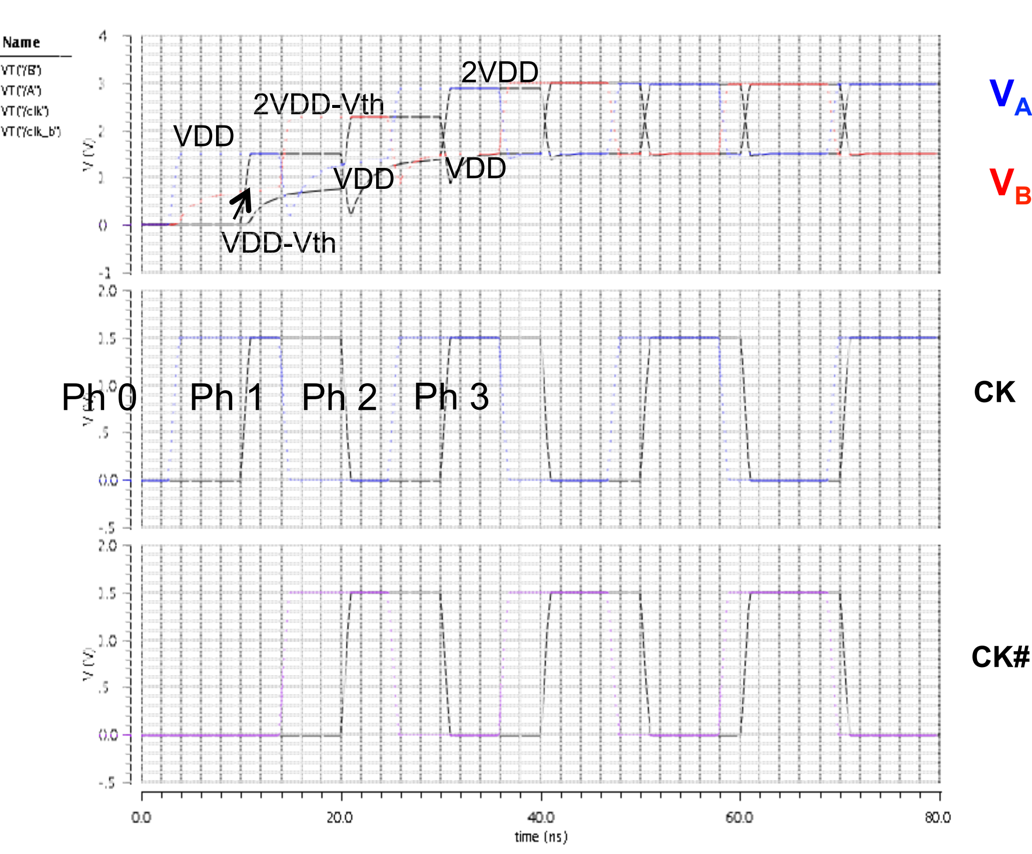

Further study: Ref. 1
Continuous boost circuit techniques

# Voltage Doubler Circuit

Basic Operating Principal:




Voltage doubler with PMOS output stage

Key advantages:

•Cascade to pump to higher supply

•Simple clock phase generators

•Can use “low voltage” transistors. Why?

Further study: Dickson charge pump, Ref. 1

# Lecture Outline

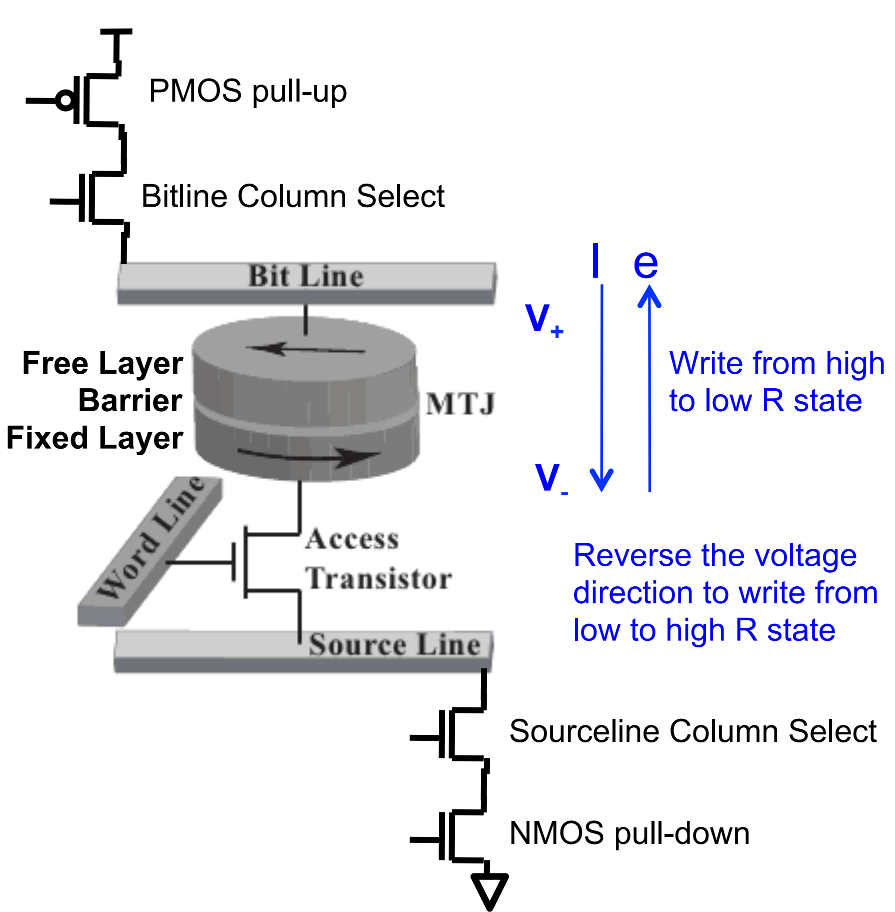
- NVM Technology Overview
- NVM Chip Block Diagram
- Key Circuit Concepts Differentiating from SRAM
  - Array Architecture
  - Voltage Boosting and Charge Pump Techniques
  - **Write Driver Design**
  - Sensing Schemes and Circuit
- NVM Interface and Applications

# Basic Write Driver Scheme

Applies to wide range of resistive memories:

Bi-direction writes in STT-MRAM, ReRAM / Memristor and uni-direction writes in PCM

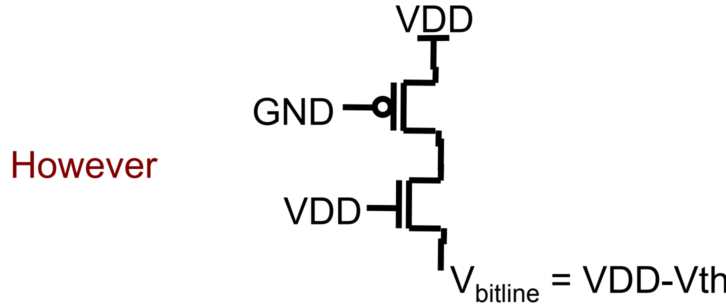
Example Write Driver for STT-MRAM




Bitline Column Select Drive strength:

- Want $V_{bitline}$ as close to VDD

However



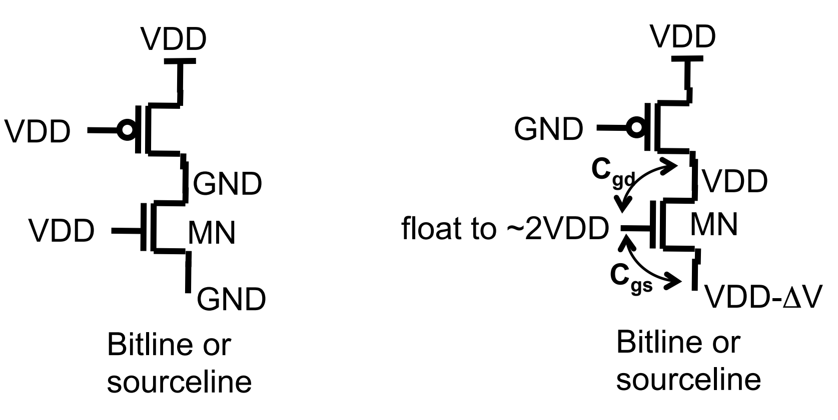

Can use PMOS for uni-direction writes.

Need a solution for bi-direction writes

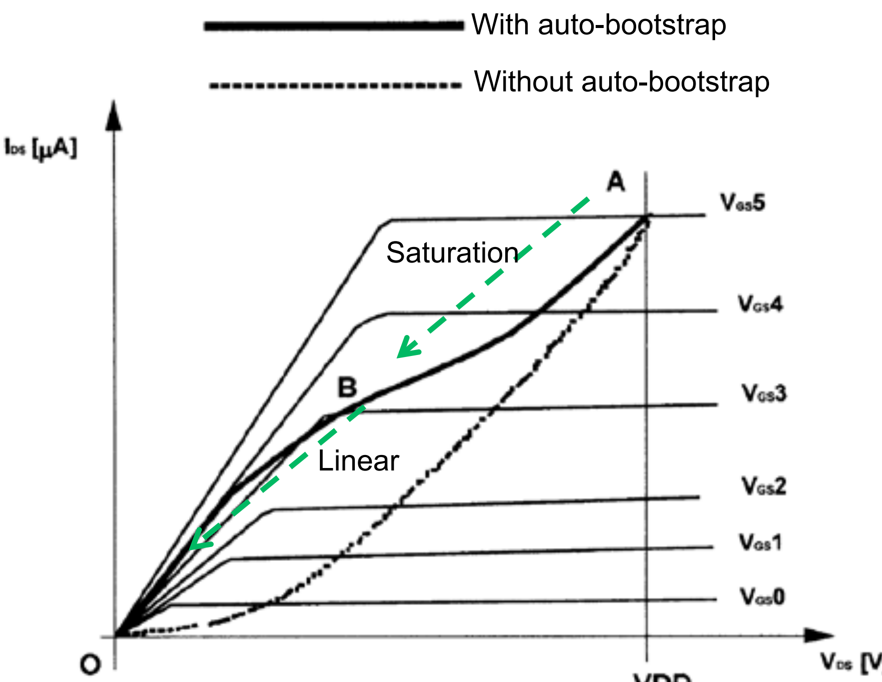
- Charged pumped Column Select signals.
- Auto-bootstrap

# Auto-Bootstrapped Column Selection

2-step $V_{gate}$ during column selection:
VDD → Float




Column Select transistor MN I-V Curve:



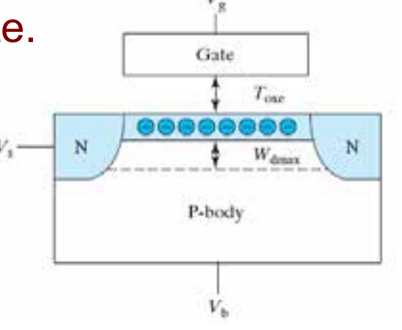

Can apply in single stage column selection.

Why not in 2-stage column selection?

- Additional $C_{load}$ to fixed nodes at the gate.

Still do not follow the ideal I-V curve. Why?

- $V_{threshold}$ increases due to body effect.




$$V_{th}(V_{sb}) = V_{th0} + \frac{C_{dep}}{C_{oxe}} V_{sb}$$

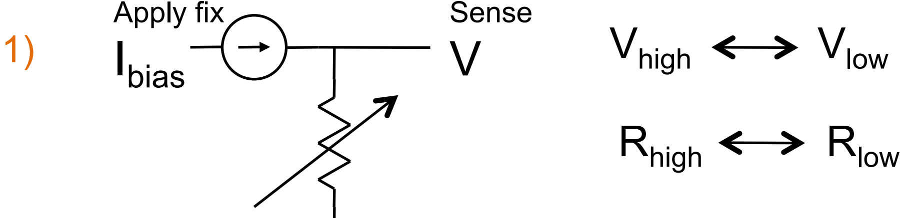
Ref. 8

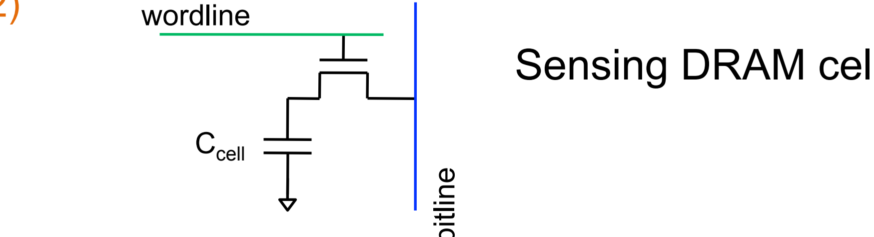
Further study: Ref. 1
- General bootstrap concept
- Bootstrapped Output Drivers

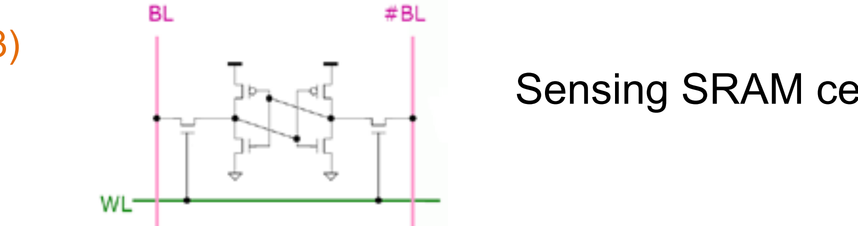
# Fundamental Approach 1: Voltage Sensing

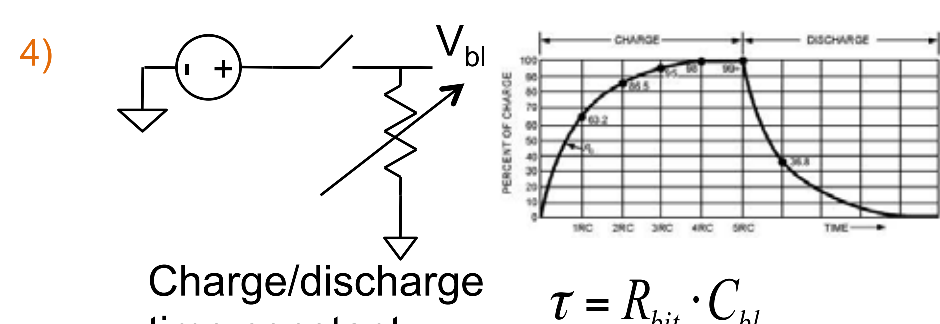
Basic operating principal: Sense the voltage difference from any one of

1) a fixed current bias through different bit states in for e.g. resistive memories.
2) voltage swing due to charge sharing in for e.g. DRAM.
3) differential voltage swing in bitline and bitline_b from active driver in for e.g. SRAM.
4) RC time constants of charging or discharging bitcell in for e.g. resistive memories



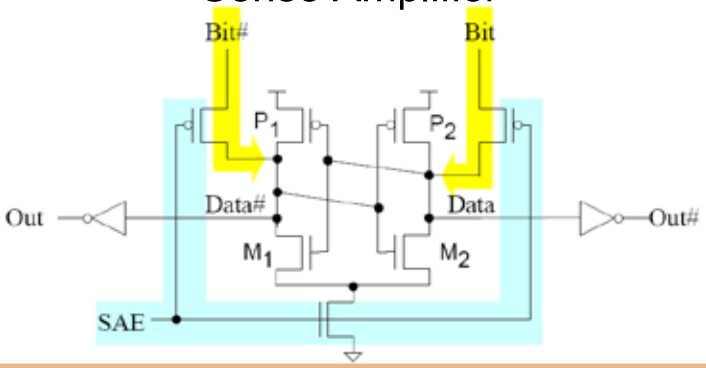




Charge/discharge time constant, $\tau = R_{bit} \cdot C_{bl}$

Example Differential Voltage Sense Amplifier



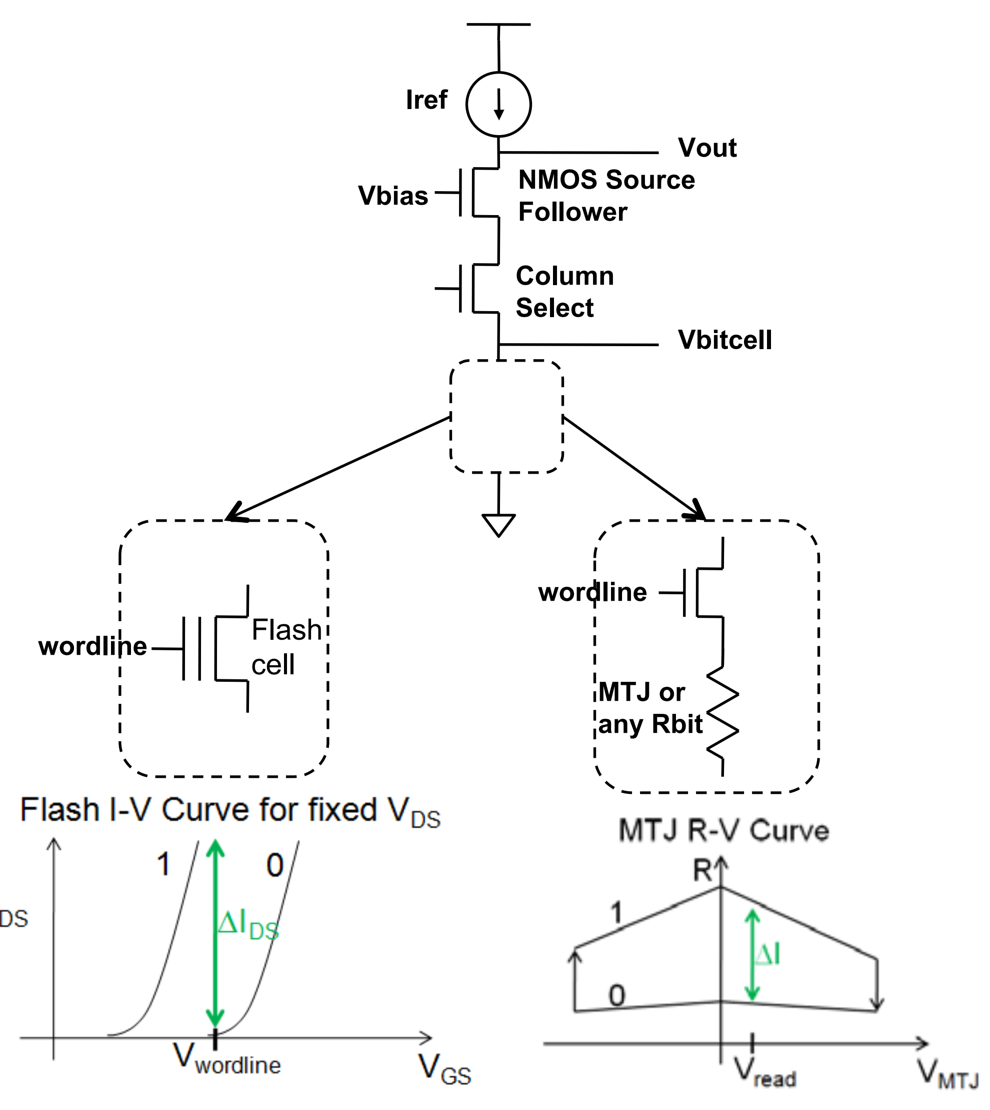
# Fundamental Approach 2: Current Sensing

Basic operating principal:

- Apply a fixed voltage bias across the bit.
- Sense the resulting current.
- Resolve the bit state by comparison with reference current.

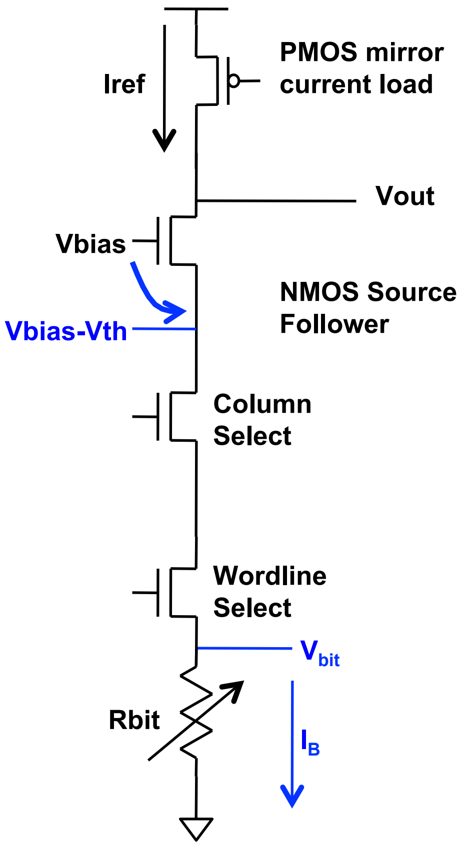
Used for sensing in:

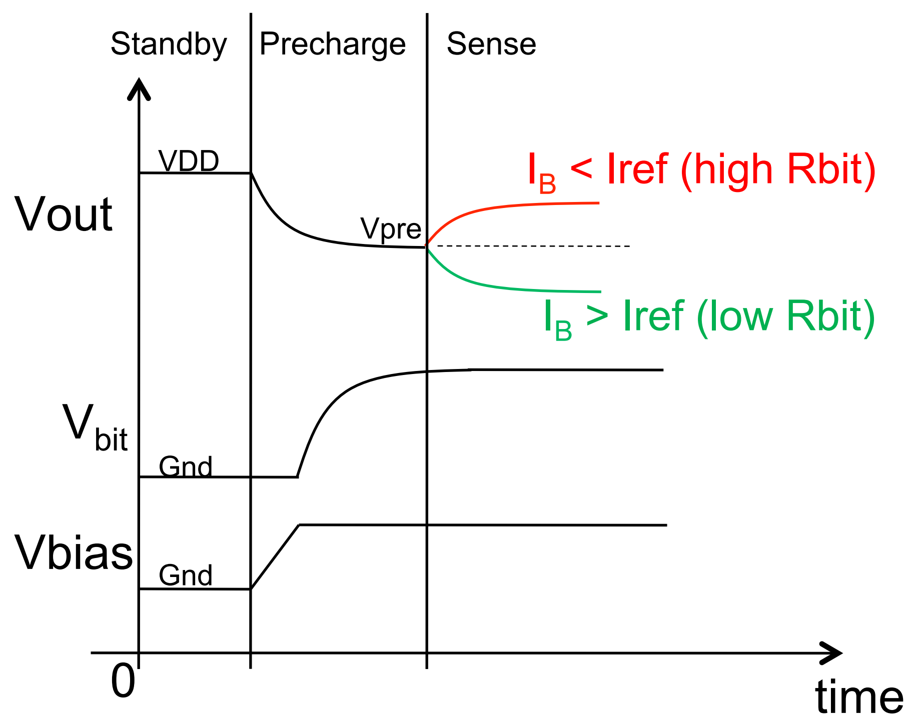
- Flash memory: fixed drain bias for flash cells of different Vth.
- MRAM: Control Magnetic Tunnel Junction (MTJ) bias for high MR and to avoid bit flipping.
- Any resistive memory for controlled bias across the bit to avoid stress and bit flipping.



# Current Sensing Amplifier Circuit Approach

Current Sensing Amplifier Circuit




With very long sensing time, Iref and $I_B$ equalizes due to Vout stabling to final value.

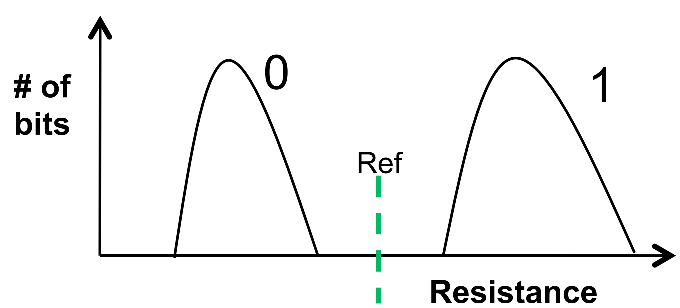
Design requirements for best sensing margin:

- Vpre is the DC voltage for Iref.
- Iref is the mid-point of $I_{high}$ and $I_{low}$.

Ref. 1

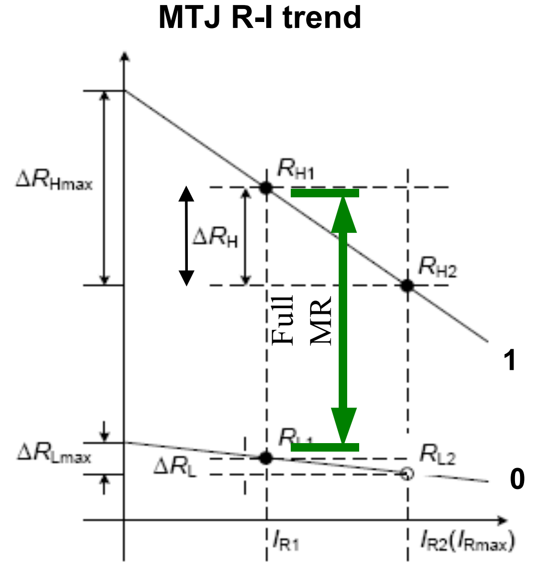
# Generating Reference for Sensing

- **Referenced Read Scheme**
  - Generate a mid-point reference from high and low bits
  - Need separation between high and low states
    - High and low states have distributions due to shape and thickness variation. Increases with smaller geometry.
  - Used in MRAM, PCM, and Flash memories

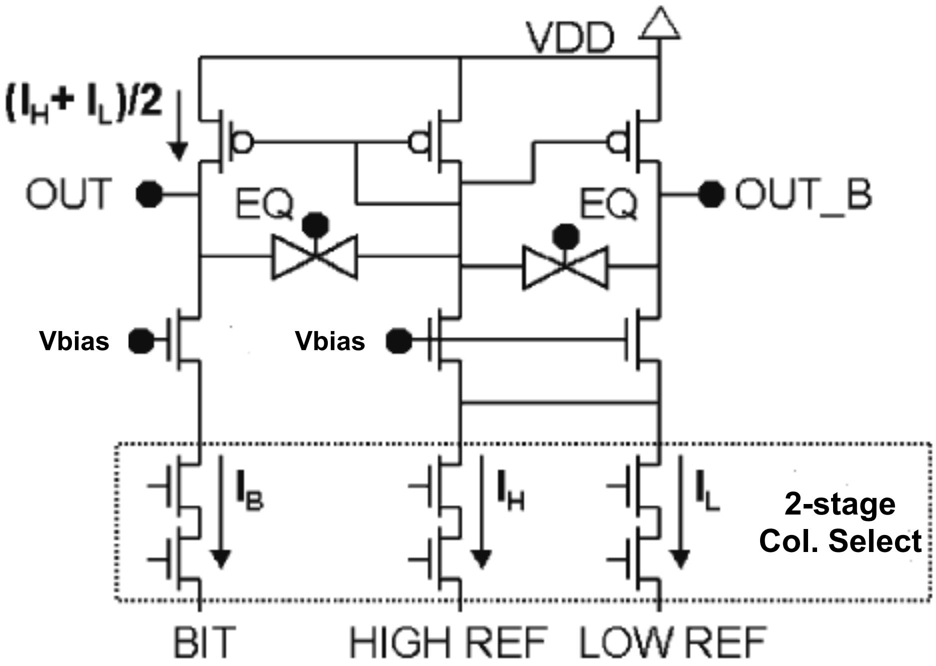



- **Self-referenced Read Scheme** due to overlapping distribution of high and low states
  - Destructive read:
    a) Sample and store the bit state variable as voltage on cap
    b) Switch the bit to opposite or a known state
    c) Resample the bit state variable and compare with a)
  - Destructive read used in FeRAM and DRAM
  - Non-destructive read scheme: Sample at two different bias voltages and compare the resulting bit state variable (e.g. voltage stored on a capacitor). Ref. 10.
  - Non-destructive read can be used in STT-MRAM due to R-I slope differences in high and low bit states.

MTJ R-I trend



# A Current Sense Amplifier Circuit



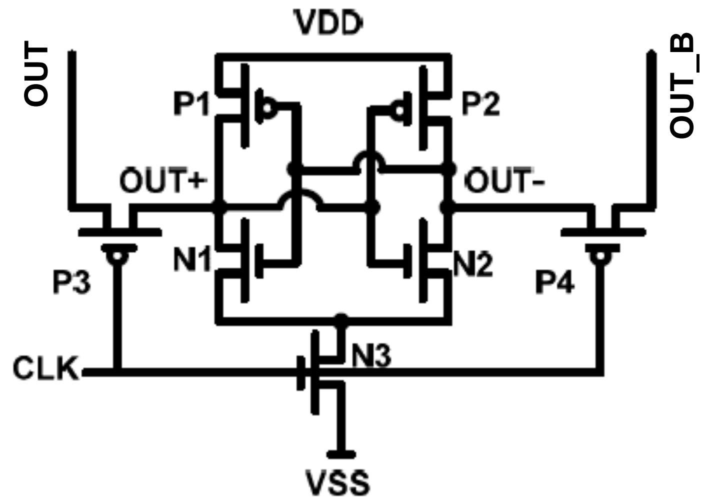

Circuit operation and key characteristics:

- Mid-point current reference from High and Low reference bits
- Self-biasing current load
- Low swing differential output
- If $I_B > (I_H+I_L)/2$
  OUT < OUT_B
  Bit = low (Logic 0)
- If $I_B < (I_H+I_L)/2$
  OUT > OUT_B
  Bit = high (Logic 1)

Further study: Ref. 9.

# Latch Type Sense Amplifier for Final Stage

Simplified LTSA Circuit




Positive feedback system with high gain.

Circuit operation

- Phase 1: CLK low
  OUT+ = OUT, OUT- = OUT_B
- Phase 2: CLK high. Cross-coupled inverter pairs latch output.

LTSA used in SRAM and DRAM sensing.

NEED V(OUT) – V(OUT_B) > LTSA offset

Statistical design for yield and speed:

Considering NMOS pair (N1, N2) Vth variation

$$\Delta VthN = V_{th_N1} - V_{th_N2}$$

$$\sigma^2{}_{\Delta VthN} = \sigma^2{}_{Vth_N1} + \sigma^2{}_{Vth_N2}$$

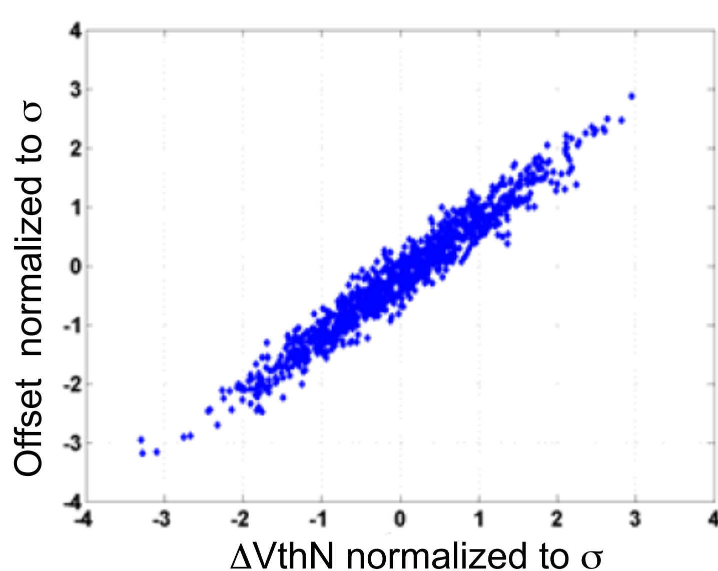
$$\sigma_{\Delta VthN} = \sqrt{2} \cdot \sigma_{Vthn}$$

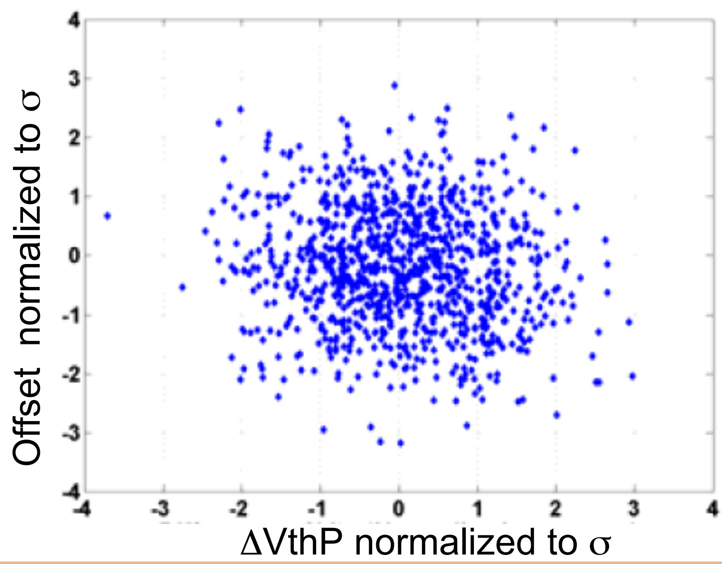
Similarly, for PMOS pair (P1, P2)

$$\sigma_{\Delta VthP} = \sqrt{2} \cdot \sigma_{Vthp}$$

Further study: Ref. 11

# Statistical Design of LTSA




NMOS pair contribution is dominant due to close to VDD bias point of OUT and OUT_B.

First order offset model:

$$offset = a \cdot \Delta VthN + b \cdot \Delta VthP$$

$$\sigma_{offset} = a \cdot \sigma_{\Delta VthN} + b \cdot \sigma_{\Delta VthP}$$

Estimate the sensitivity factors a and b from

- Montecarlo simulations
- Or spice simulations with Vmismatch voltage sources at transistor gates

Increase the most sensitive transistor area for acceptable offset to desired yield level (number of σ)

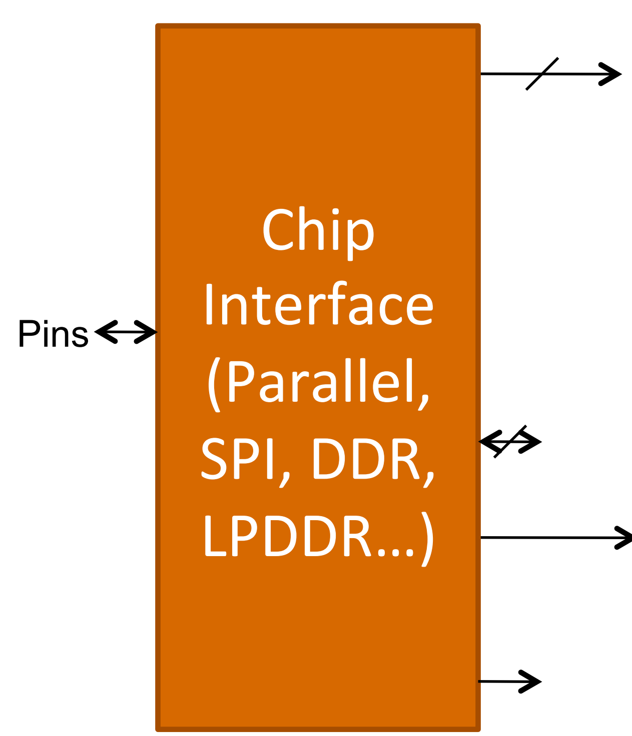
$$\sigma_{Vth} = \frac{A_V}{\sqrt{W \cdot L}}$$

← Tech. parameter
← Transistor area

# NVM Chip Interface and Applications

NVM Chip Interface facilities interaction between application and NVM technology !!

You can exploit technology strength and hide technology weakness by choosing the right memory interface and application.




Asynchronous Interface Pins:

| | |
|---|---|
| WE/ | Write Enable |
| CE/ | Chip Enable |
| OE/ | Output Enable |
| DQ | Data Input/Output |
| A | Address |

Additional Synchronous Interface Pins:

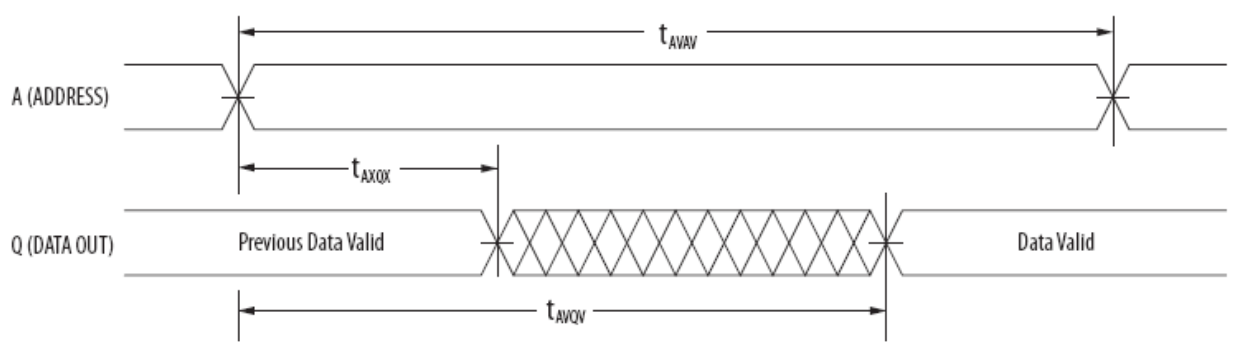
| | |
|---|---|
| RAS/ | Row Access Strobe |
| CAS/ | Column Access Strobe |
| CLK | Clock |

Differential Signals for high speed operation:
CLK, CLK/
DQS, DQS/

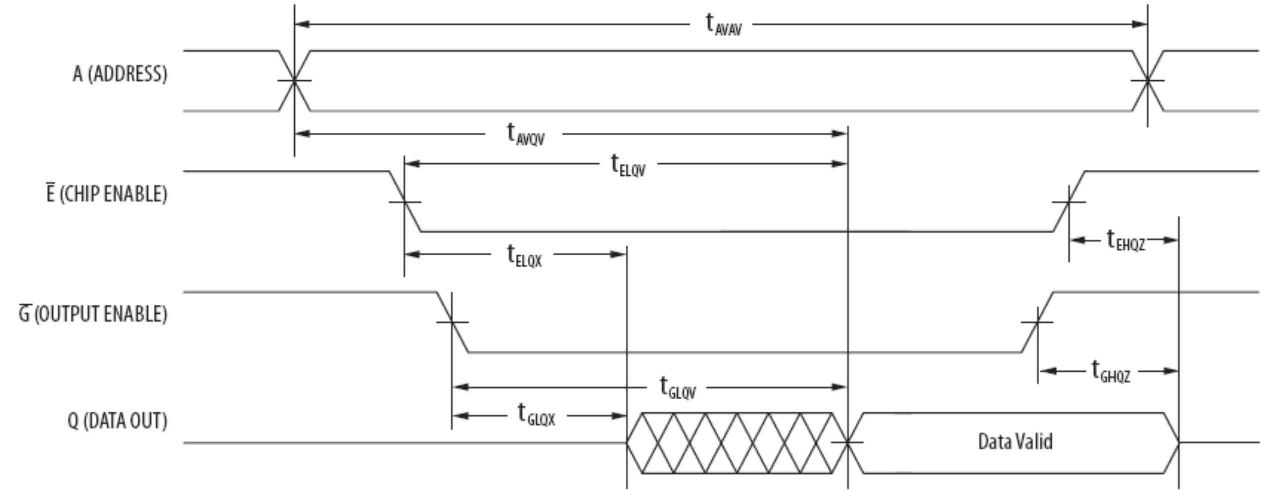
Common NVM Interfaces:
- Asynchronous Parallel interface, Ref. 13
- SPI (Serial Peripheral Interface) synchronous interface, Ref. 13
- DDR (Double data rate) synchronous interface, Ref. 12

# Asynchronous Parallel Read Access

Address-driven Read Cycle when Chip Enabled




Chip Enable Driven Read Cycle



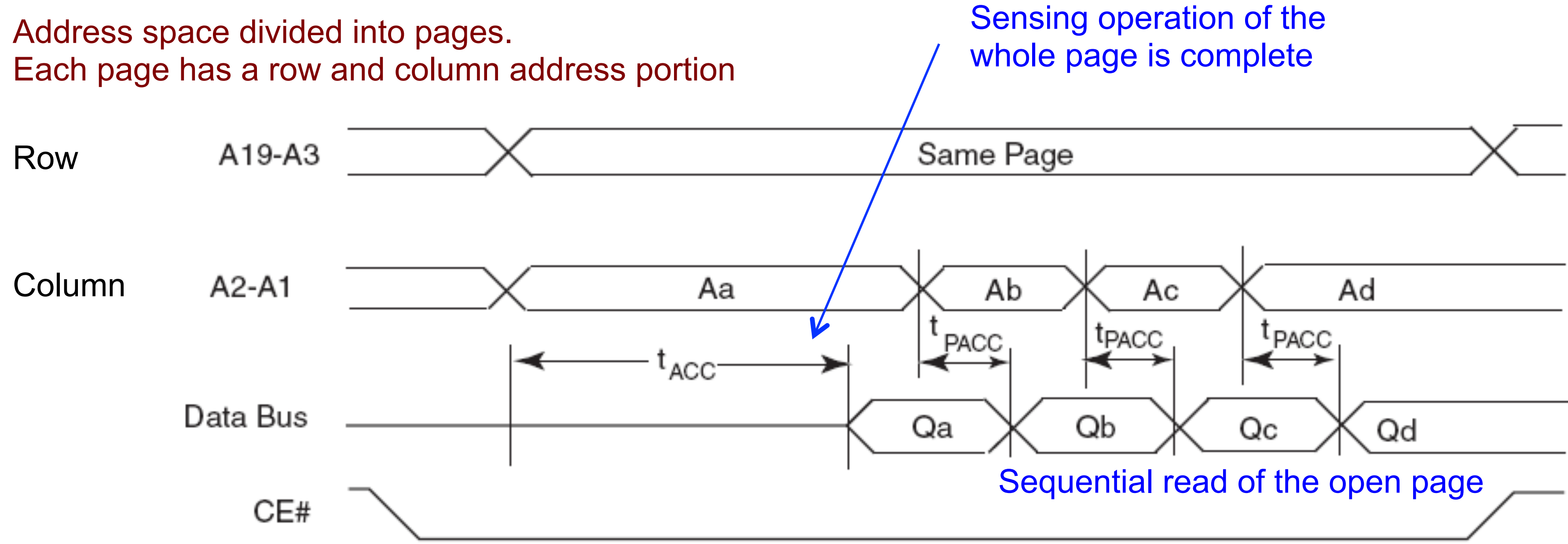

Read 8-bit or 16-bit data (width of DQ bus) in response to address or chip enable.

Data access time would directly translate to memory read time (sensing time)

- Random data access
- Fast access time (for e.g. 10 – 40 ns)

Application: Ideal for code and non-stream data in industry automation, automotive, computing, networking

# Asynchronous Page Read Access

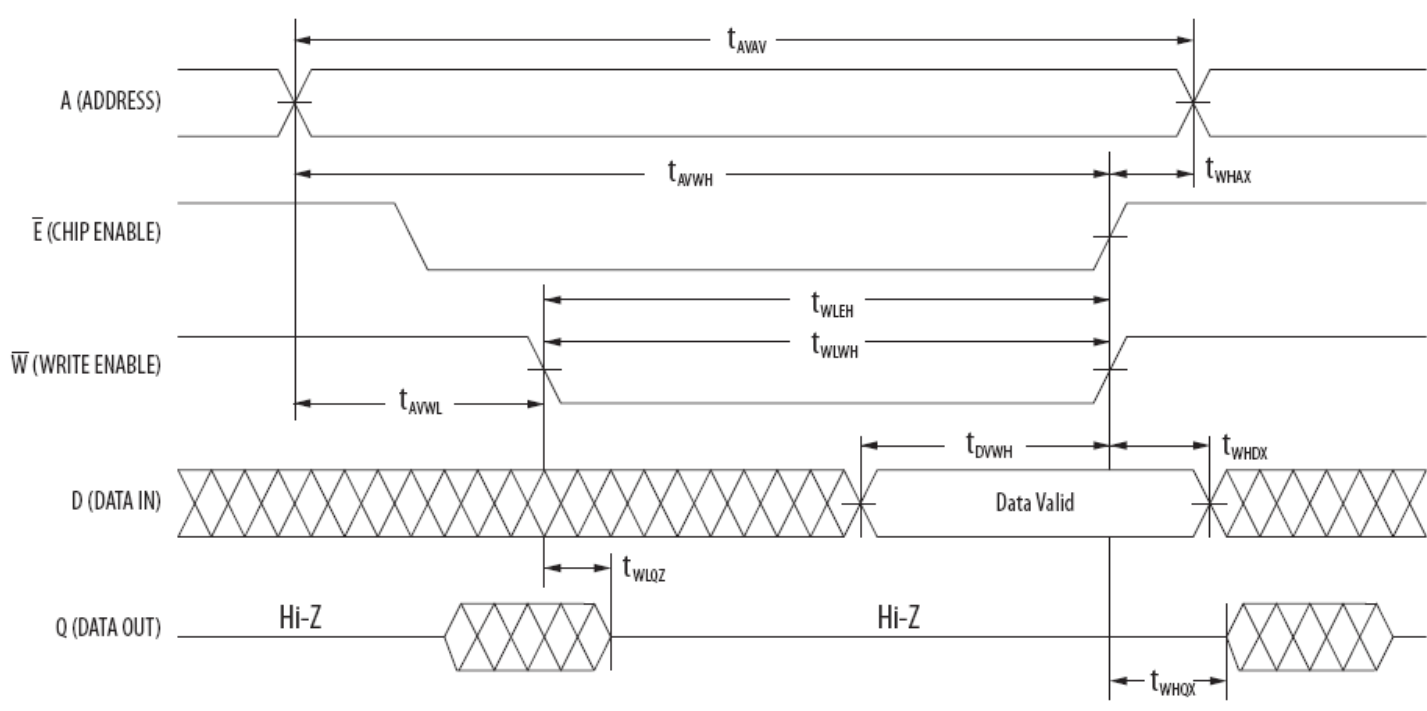



Key Advantages:

- Timing cost of long sensing is amortized with sequential fast accesses. Suitable for memory with long sensing time e.g. NAND Flash.
- Page or block data access w/ narrow data bus: Ideal for code and stream data
- Increased bandwidth. Example bandwidth calculation with $t_{ACC}$=30ns, $t_{PACC}$=5ns and 8-bit wide data bus:

| Random Access | Page Access |
|---|---|
| 4 x 8 bits = 4 x tACC = 120ns | 4 x 8 bits = 1 x tACC + 3 x tPACC = 45ns |
| Bandwidth = 32 bit / 120ns | Bandwidth = 32 bit / 45ns |
| = 266 Mbit/s | = 711 Mbit/s |
| | (almost 3x increase in bandwidth!) |

# Asynchronous Write Access

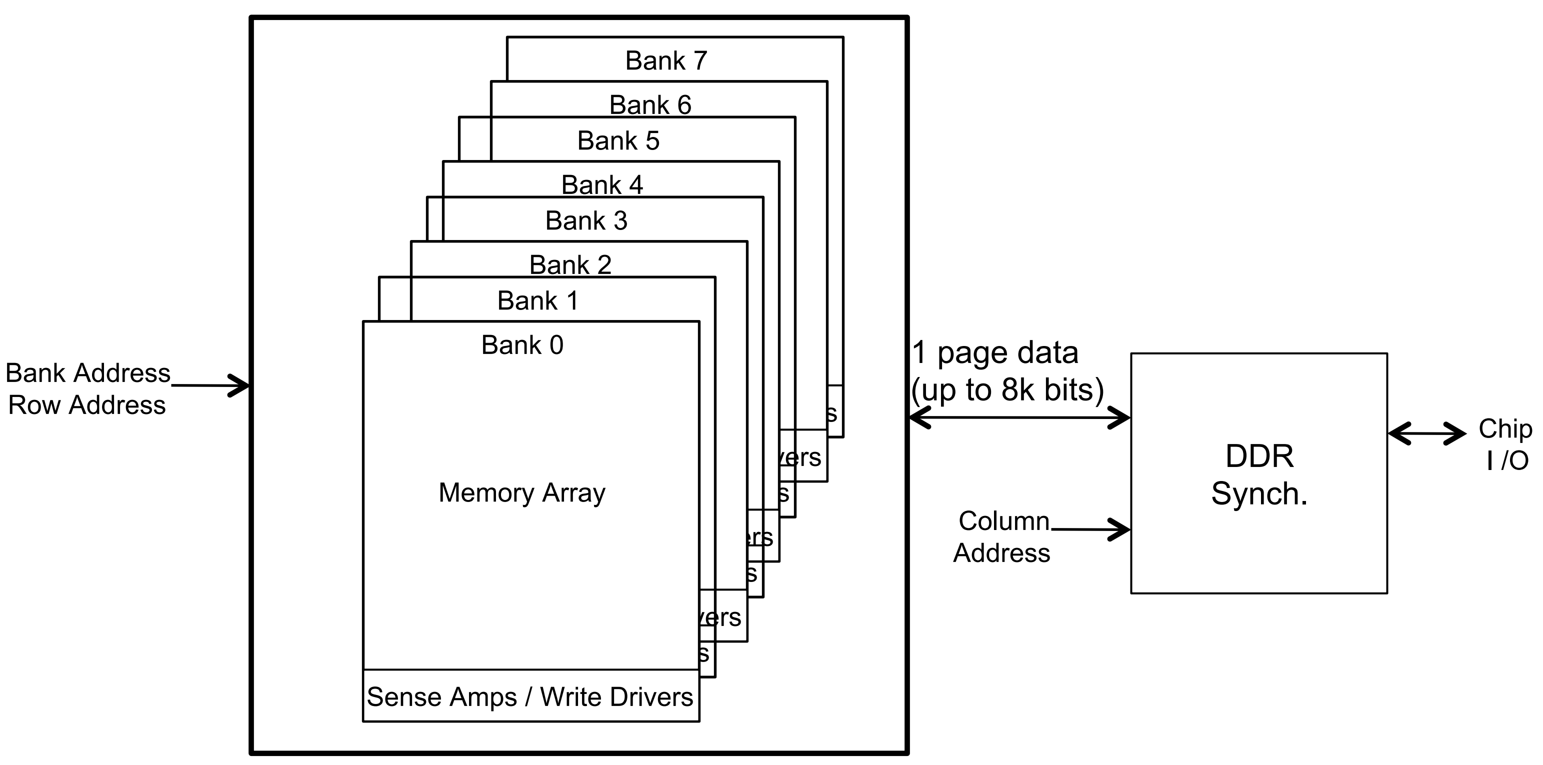



Write 8-bit or 16-bit data (width of DQ bus) in response to write or chip enable.
Data write access time would directly translate to memory write time

- Random data write and fast write time (for e.g. 10 – 40 ns) in MRAM
- Write delay needed for Flash and PCM

Application: Ideal for code updates (cognitive computing, security) and non-stream data

# Prefetch Architecture for DDR Interface



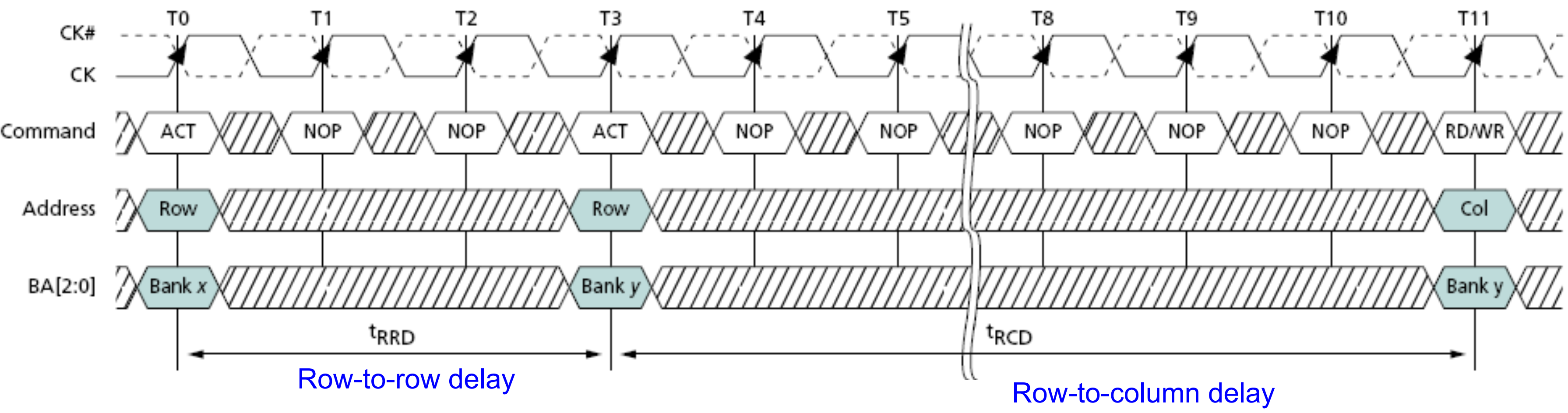

- Prefetch a page (up to 8k) of data from a Bank (8 banks) and store in local cache for fast read and write of cache.
- Parallel bank accesses for high data bandwidth.
- Differential clock (up to 800MHz) and data strobe.
- Data at both rising and falling edges of the clock.

Further study:
Ref. 12

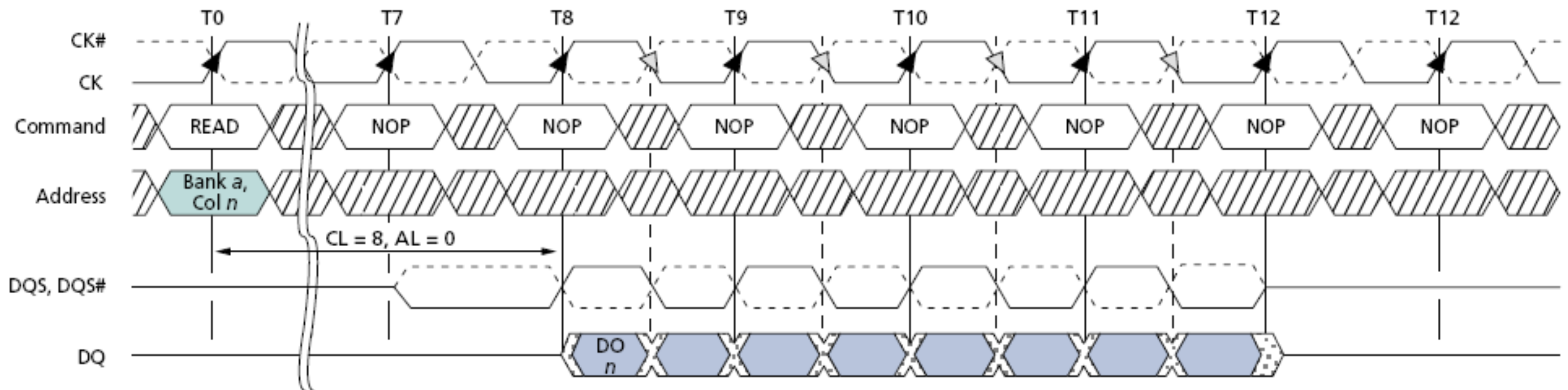
Applications:
- Main memory (DRAM) in computing systems
- High speed buffer (non-volatility application) in storage systems

# High Speed DDR Interface Operations

Activate: Open a bank for read or write (row access)

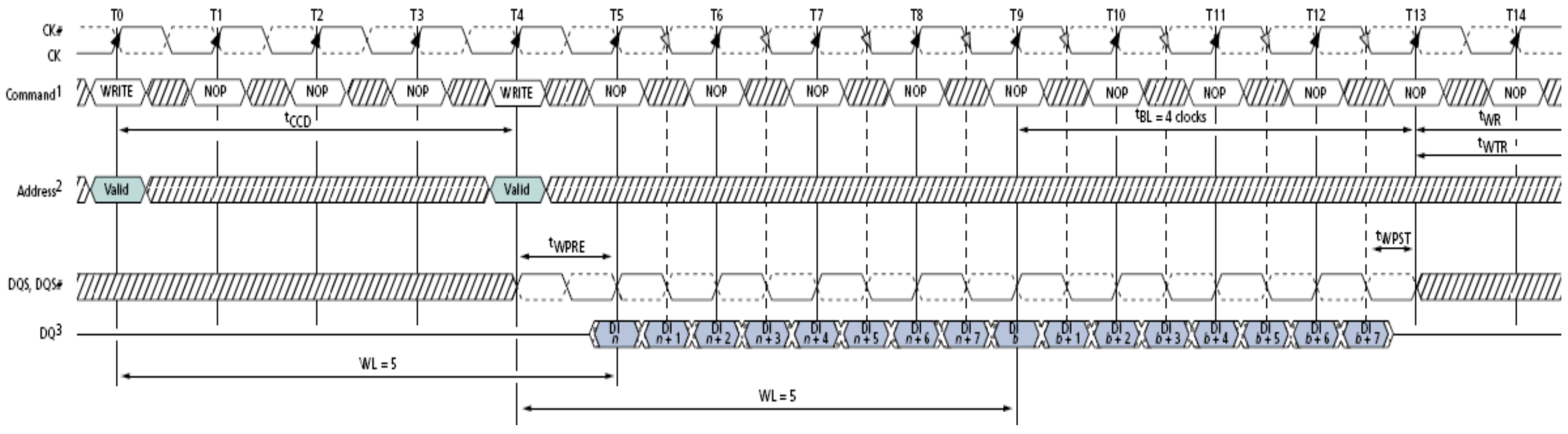



Read: Read local cache bits using column address (column access)

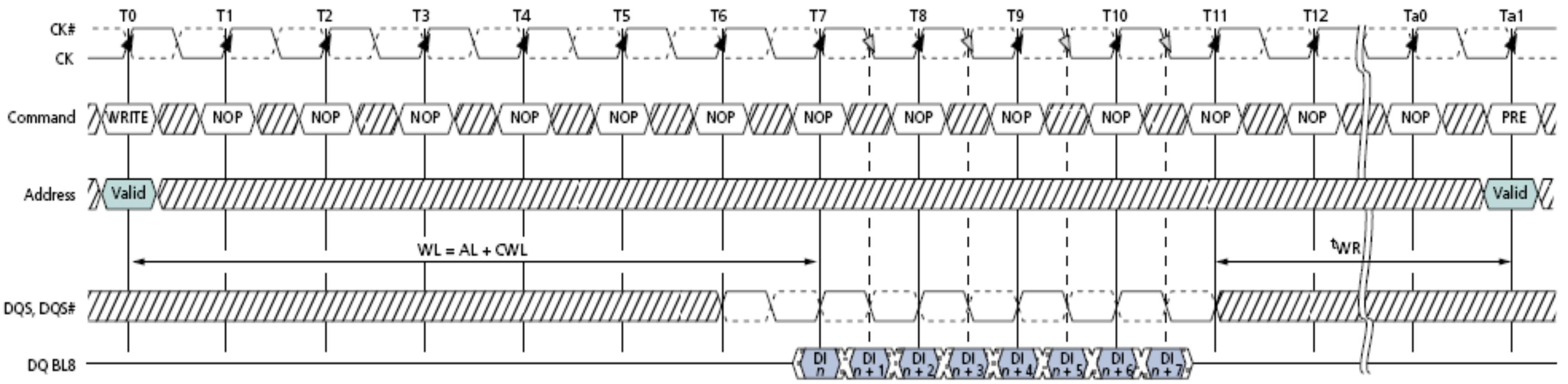


# High Speed DDR Interface Operations

Write: Write to local cache bits using column address (column access)




Precharge: Close a bank using row address (row access)



# Conclusion

- Non-volatile memory designers need extensive interaction and co-design of technology, circuits, and application / market particularly for introducing emerging technologies.
- NVM chip design is similar to SoC design requiring analog, mixed-signal, and digital design skills.
- No one NVM technology (Flash, MRAM, FeRAM, ReRAM, PCM) has the best of all atributes (area, endurance, retention, cost ….) yet.
- Key circuit concepts differentiating from SRAM covered in the lecture are
  - Array Architecture
  - Voltage Boosting and Charge Pump Techniques
  - Write Driver Design
  - Sensing Schemes and Circuit

# References

1. VLSI-Design of Non-Volatile Memories, G. Campardo et. al. Springer; 2005 edition (March 17, 2005)
2. International technology roadmap of semiconductors, ITRS, emerging device technologies, 2011 edition, public.itrs.net.
3. TN-29-19: NAND Flash 101 Introduction, Micron Technical Note. *download.micron.com/pdf/technotes/nand/tn2919.pdf*
4. A. L. Lacaita, "Phase change memories: State-of-the-art, challenges and perspectives," Solid-State Electronics, vol. 50, pp. 24-31, 2006.
5. Wong et. al. "Phase Change Memories,", Proceedings of the IEEE, vol. 98, no. 11, pp. 2201-2227, Dec. 2010.
6. R. Ranjan et. al., "Progress and Outlook for STT-MRAM," ICCAD Tutorial, 2011.
7. J.M. Slaughter et al, "High Density ST-MRAM Technology (Invited)," in *Proceedings* of International Electronic Devices Meeting (IEDM), paper 29.3; San Francisco, CA (2012).
8. C. Hu, Chapter 6: MOS Transistors, www.eecs.berkeley.edu/~hu/Chenming-Hu_ch6.pdf
9. T. Andre et. al., "A 4-Mb 0.18um 1T1MTJ Toggle MRAM With Balanced Three Input Sensing Scheme and Locally Mirrored Unidirectional Write Drivers," IEEE Journal of solid-state circuits, pp. 301-309, vol. 40, no. 1, 2005.
10. Y. Chen et. al, "A Nondestructive Self-Reference Scheme for Spin-Transfer Torque Random Access Memory (STT-RAM)," proceedings of IEEE Design and Test in Europe, 2010.
11. L. Pillegy et. al., "Mismatch Analysis and Statistical Design at 65 nm and Below," proceedings of IEEE Custom Integrated Circuit Design, 2008.
12. DDR3 SDRAM, Micron Technologies, http://www.micron.com/products/dram/ddr3-sdram
13. Everspin Technologies, product family, http://www.everspin.com/products.php
14. D. Sekar, Resistive RAM: Technology Status and Future Opportunities, 2014 IEEE Memory Symposium.