5 Survey of memories

5.1 Introduction

History of information storage

How to improve : reliability - general access, copies… :

- (1) Coding : from LANGUAGE to CODE : pictures - pictogrammes – alphabet
- (2) Techniques & Media : from Stone to Paper

Rosetta Stein (196 v. Chr.)

Dichte	~ 25kB
Höhe	114 cm
Breite	72 cm
Dicke	28 cm
Gewicht	762 Kg
Material	Stein

In/on **stone** : drawings, pictograms

- Long retention times
- Labour intensitive write

More flexible writing – but more prone to decay

- **Clay tablets** (cuniforms) : Bronze to Iron age, esp. Middle 3rd Millenium BC
- **Papyrus** (hieroglyphs) : Egypt, from 3rd Millenium BC
- **Parchment** : from 3rd Millenium BC till Middle Ages
- **Pape**r : 2nd century BC in China

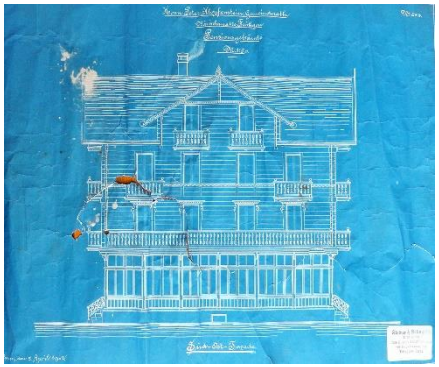
History of information storage

How to improve : reliability - general access, copies… :

- (1) Coding : from LANGUAGE to CODE : pictures - pictogrammes – alphabet
- (2) Techniques & Media : from Stone to Paper
- (3) Reproduction : Hand Copying to Printing

Seals (3000 BC)

Blueprints (19th Century)



Book printing :

Woodblock
- Han dynasty 220BC

Moveable Type
- Bi Chen China (1040 AD)
- Improved by Johannes Gutenberg (1450)

Xerox Copy
(1938)

Different forms of information

Not only coded language, also SIGNALS : e.g. audio / video

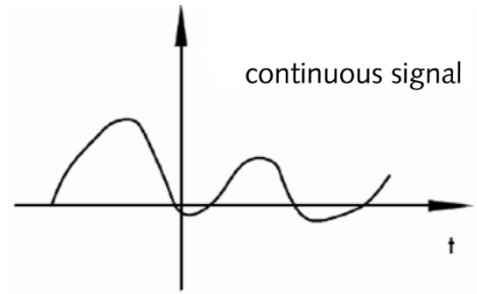
First **audio** recording:
- Phonograph Thomas Edison 1877
- Grammophone Emil Berliner 1887

Magnetic **audio** recording :
- Wire recording 1888
- Magnetic Tape 1928
- Compact Casette Philips 1963

Magnetic **video** recording:
- Vision Electronic Recording Apparatus (VERA) (BBC 1952)
- Portable Video Tape Recorder (Ampex 1954)
- Video Casette Recorder (Sony 1971)

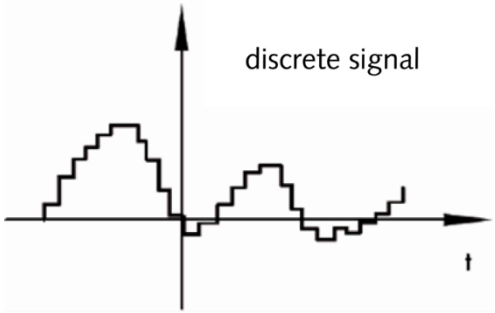
The DIGITAL world

Nature signals: **analog**

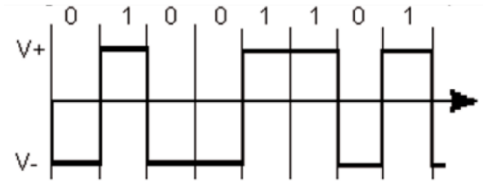


Technical signals: **digital**

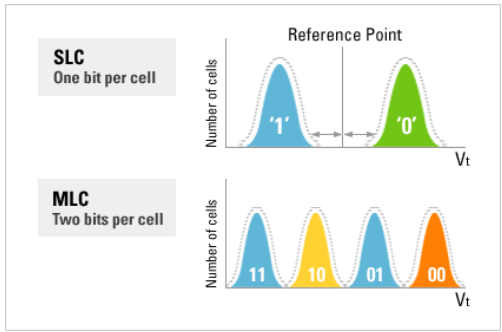
- Noise immunity (unambiguousness)
- No error propagation (realization of complex systems)
- Long distant transmission
- Boolean algebra *
- Simple testing, storing and processing



digital signals:
rectangle pulses

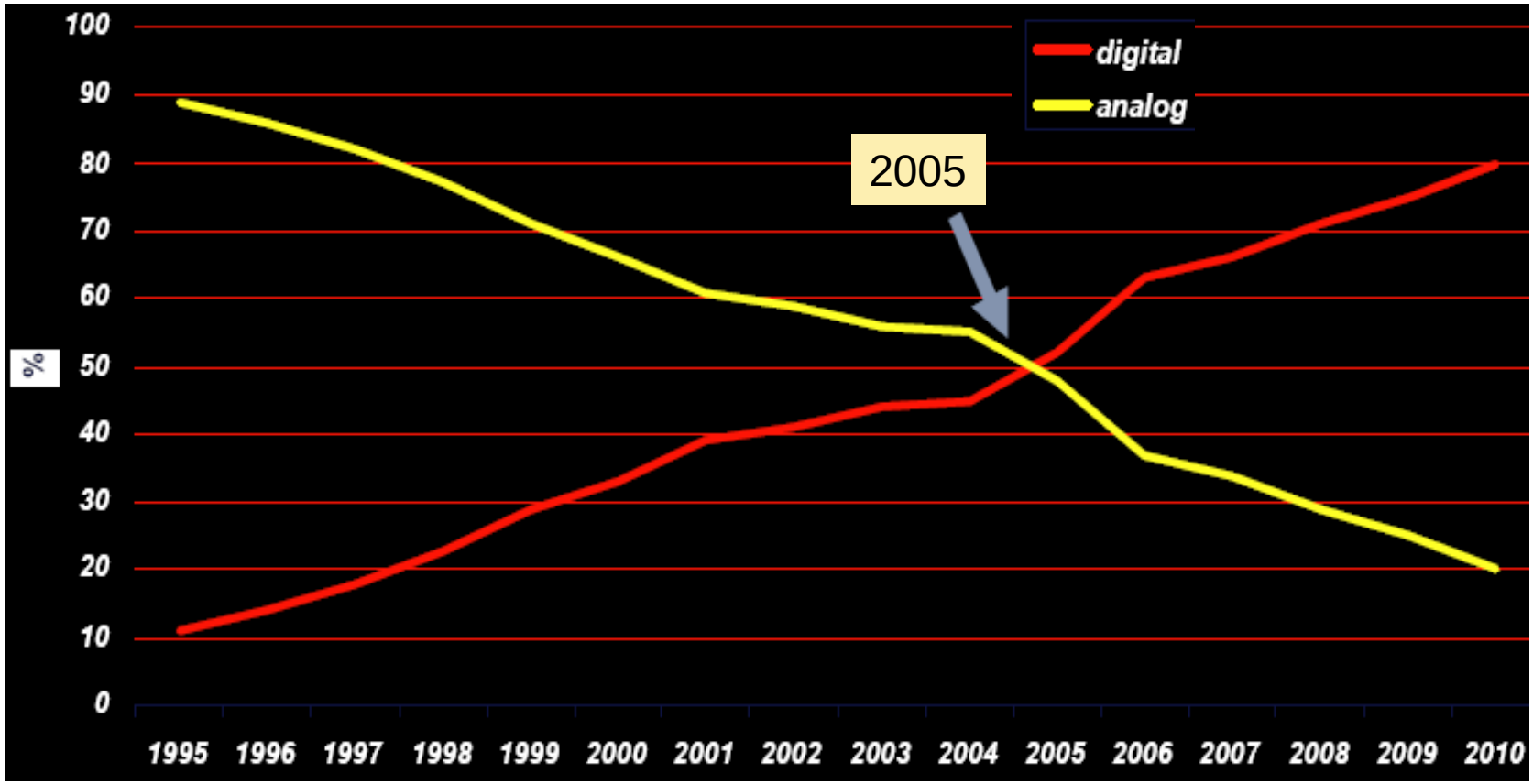


** decoding*



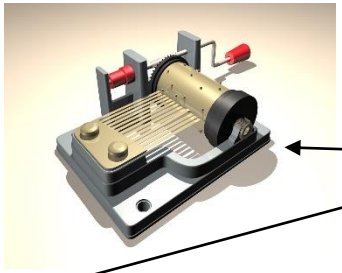
Note: DIGITAL >does not has to be BINARY :
MULTI-LEVEL CODE

Analog vs. digital devices



DIGITAL coding & storage

Metal rolls with pins
- Music box 19th century

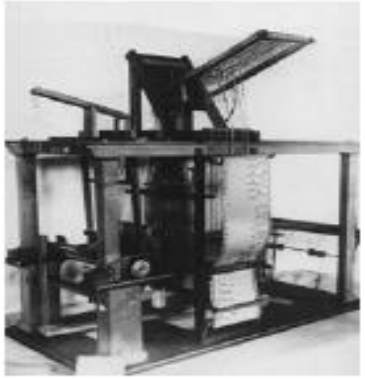


use for **Control**

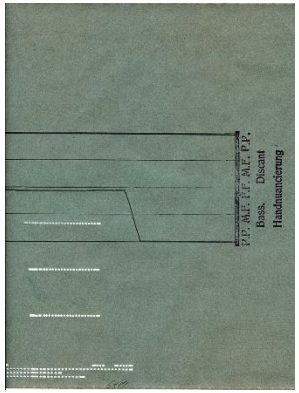
1st use for **Data**

Punched carton

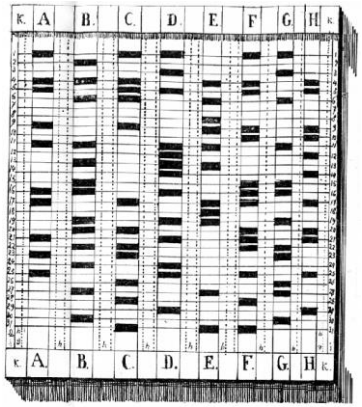
Jacquards:
Weaving loom (1801)



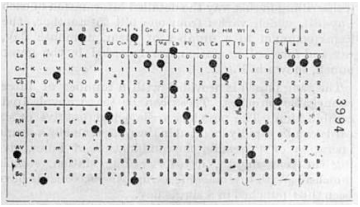
Piano Roll
(1896)



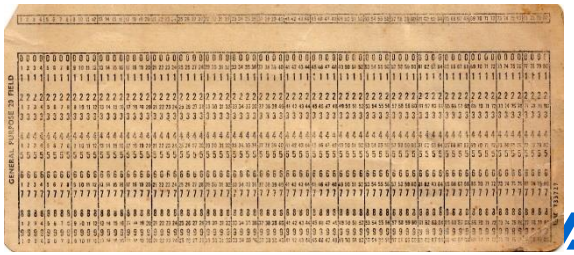
Korsakov Punched Card
(1832)



Hollerith‘s Census Machine
(1890)



IBM Punched Card
(1928)



DIGITAL coding & storage

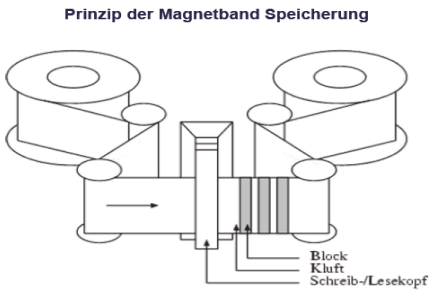
Magnetic

- Magnetic Tape
- Magnetic Core Memory
- Harddisk

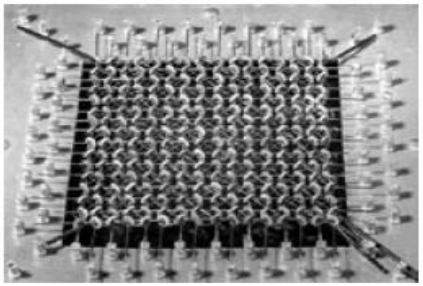
Optical

- CD / DVD

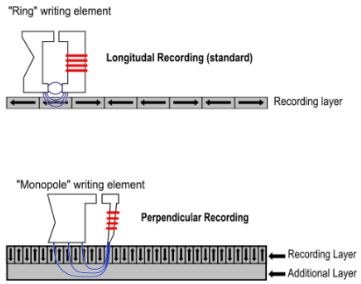
Pfleumer:
Magnetic tape (1927)



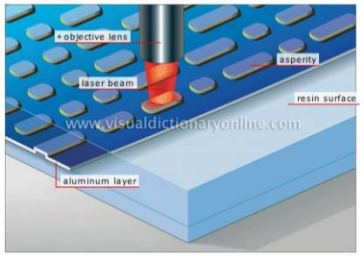
Wang/Forrester:
Magnetic core
(≈1950)



HDD
IBM (1956)

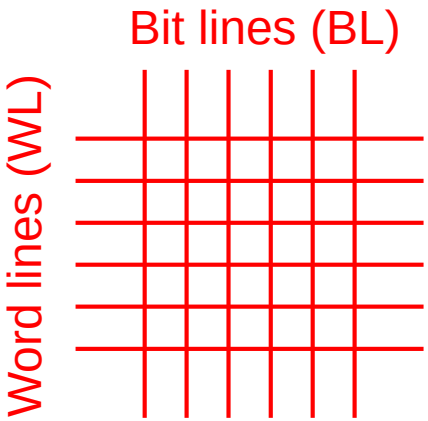


Compact Disk
Ottens (Philips 1974)



Random Access Memories vs. Mass Storage Devics (MSD)

	Mass Storage Devices	Random Access Devices
System	• access unit (e.g. R/W head) • storage medium	• matrix of conductor lines • storage elements at nodes
Addressing of Data	mutual positioning of the R/W unit & storage medium	application of columm & row address signals
Data exchange	• mechanical • optical • magnetic fields • electric fields	electronic access via matrix (voltage mode or current mode)
Future		



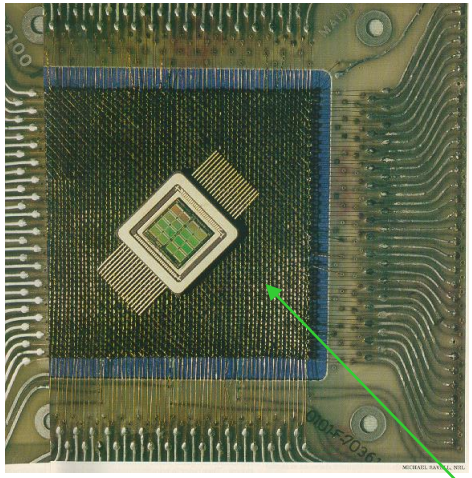
Random Access Memories vs. Mass Storage Devics (MSD)

Random Access Devices

- **Matrix arrangement**
- (fast) access to each single cell

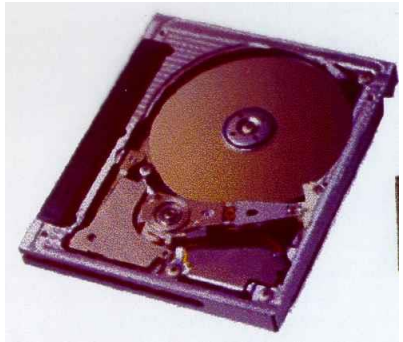
MSD:

- Sequential access
- **no** (fast) access to each single cell



1 Kbit magnetic-core memory

16 Kbit thin film memory (Honeywell)



Hard Disc :
Mixed (sectors)

Tape :
Fully serial

What is best ? → depending on application & cost

- tape: lowest cost/bit: OK for BACKUP of large amount of data
- CPU memory need fast and random access : fast SRAM arrays (higher cost/bit)

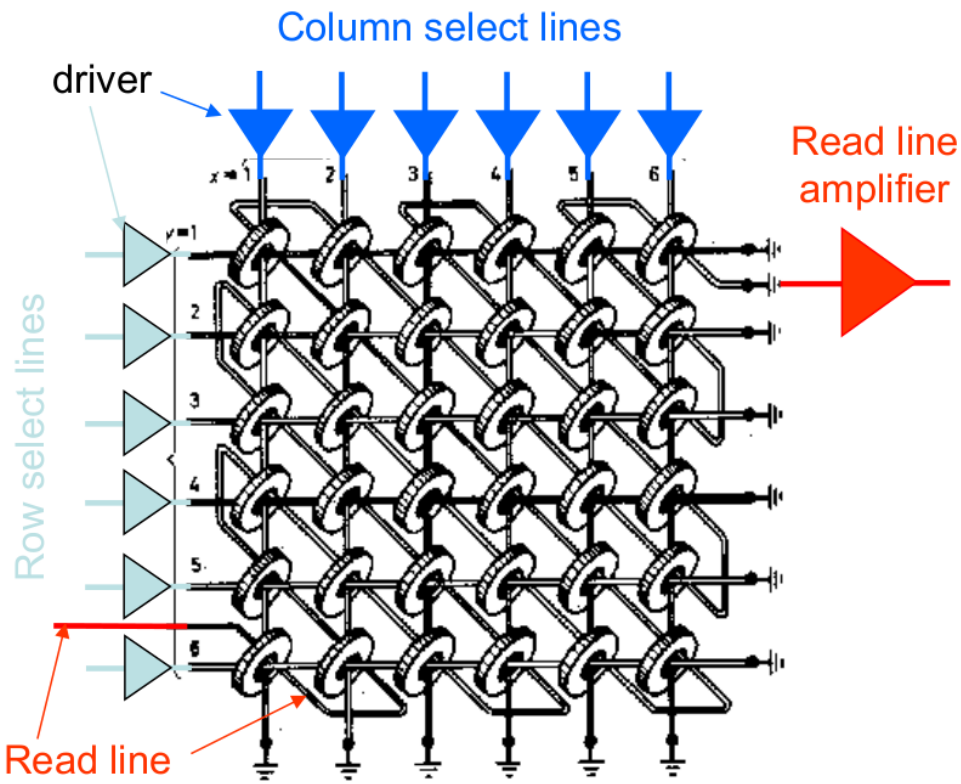
HDD: ≈ ms ≈ 1 Gbit

Note: difference between ACCESS TIME (latency) and BANDWITH (BIT rate)

SUPPLEMENTAL: Magnetic-core memory

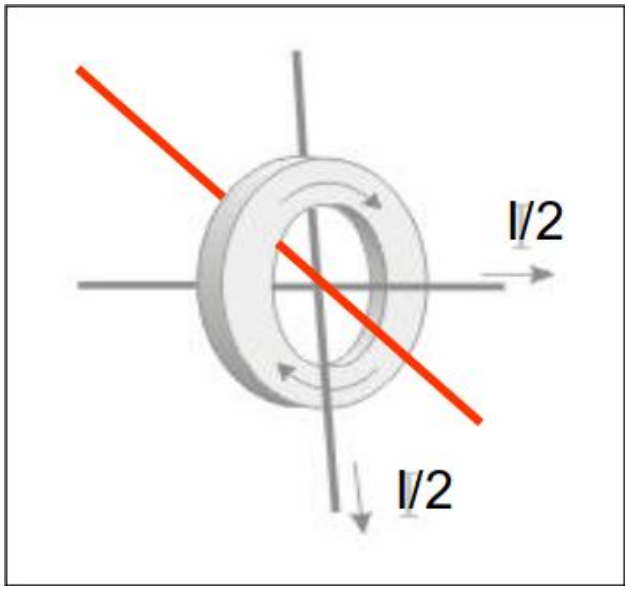
Illustration of

- matrix arrangement
- random Access



Read/Write:

- **Induction**
- Destructive read-out (DRO)
- Non-volatile (NVM)

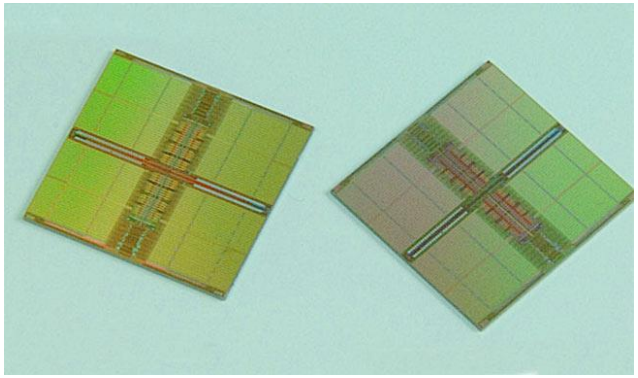


5.2 Integrated memories

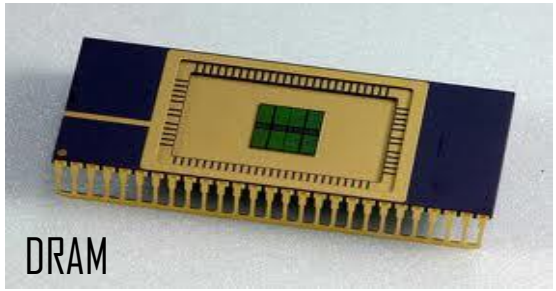
Integrated Memories

A Memory System is not only Storage elements
- Need to access, read (and/or write) particular memory element (Bit)

Previous examples:
- Heterogeneous system : memory medium, read/write heads, scan mechanics
- Bulky, slow, high energy consumption



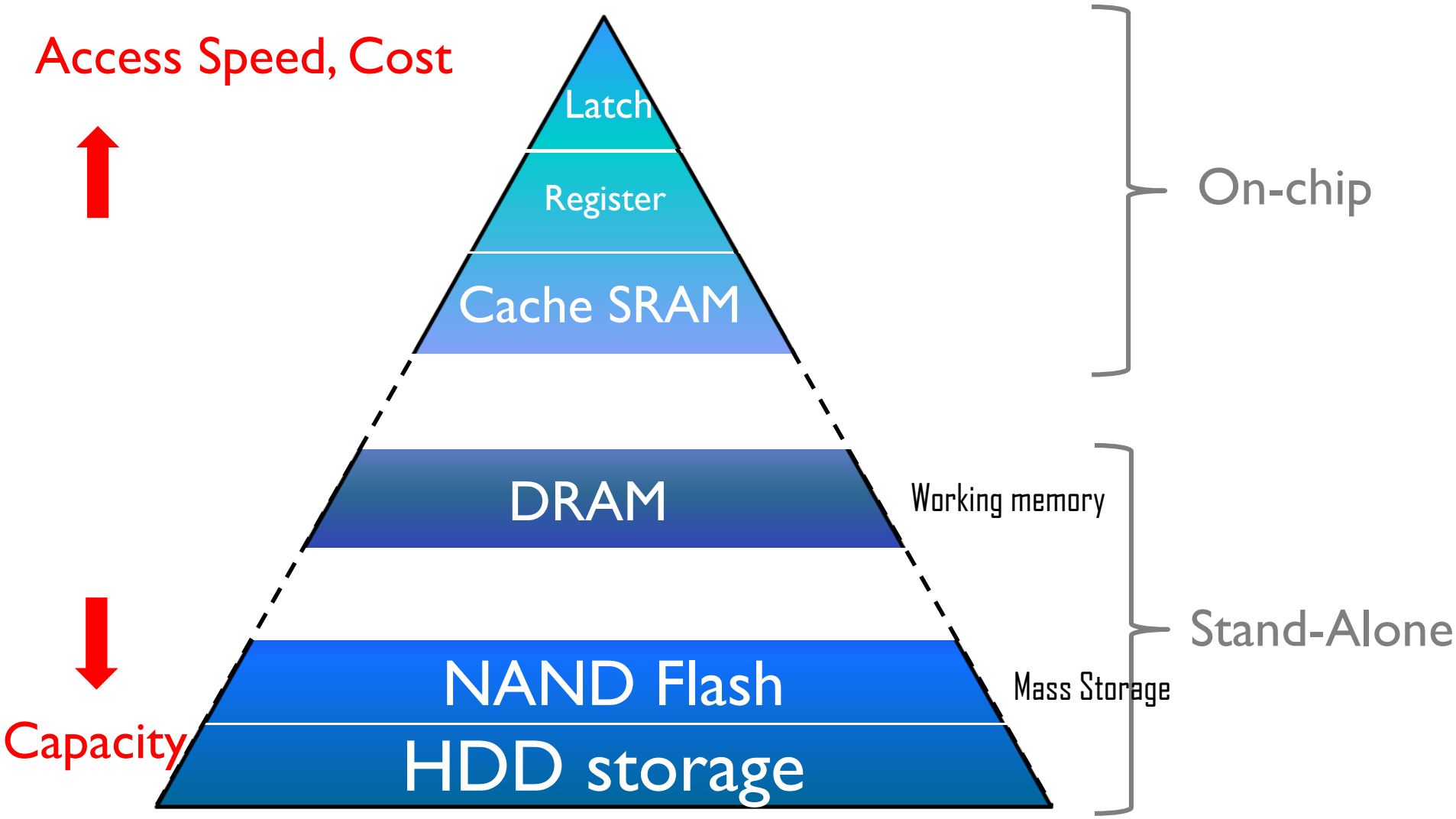
Improvement : **Integrated Memories**
- Complete memory system integrated in a Si based microchip
- No mechanics → robust/small form factor/low energy (portable!)
- Active vs Passive Matrix (performance)
- Low Capacitances : fast
- Low cost (due to faboulous SCALING according to Moore´s law!)



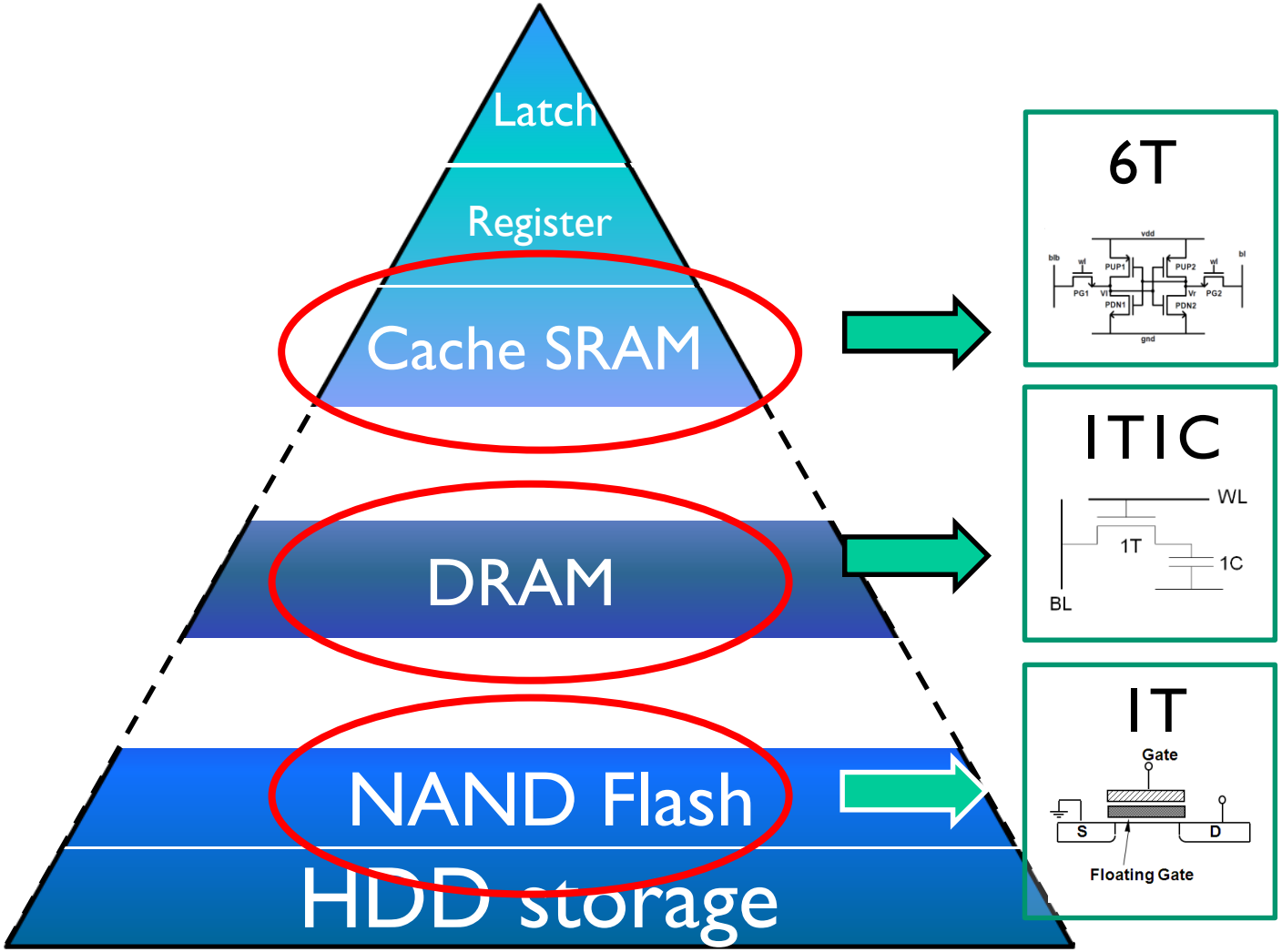
USB memory

SSD

The Memory Hierarchy

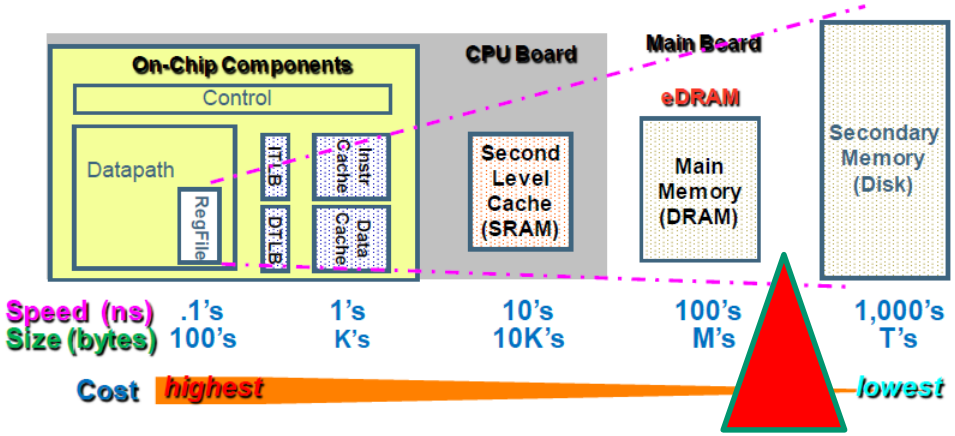


The Memory Hierarchy : Three main Si (transistor/charge) based Memory types

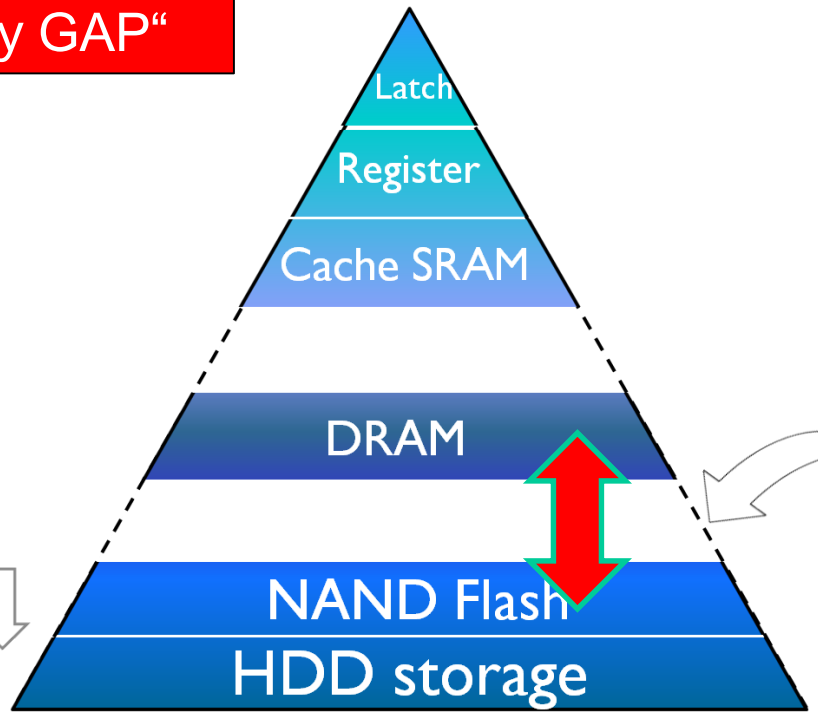


← 3dim

The Memory Hierarchy : changes/opportunities



Large „latency GAP“



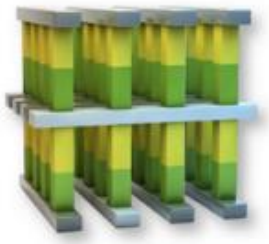
(1) Flash based SSD replacing HDD

(2) opportunity for new class of (NV) memories : **Storage Class Memory** (SCM) ~ performance better than FLASH

X-point technology

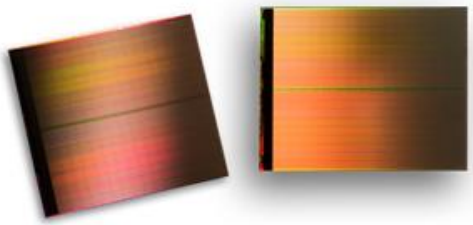
3D XPoint Technology

Intel and Micron 2015.7.28



Breakthrough Nonvolatile Memory Technology

The explosion of connected devices and digital services is generating massive amounts of new data. For this data to be useful, it must be stored and analyzed very quickly. 3D XPoint™ technology is an entirely new class of nonvolatile memory that can help turn immense amounts of data into valuable information in real time. With up to 1,000 times lower latency and exponentially greater endurance than NAND, 3D XPoint technology can deliver game-changing performance for big data applications. Its ability to enable high-speed, high-capacity data storage close to the processor creates new possibilities for system architects and promises to enable entirely new applications.



x1000 faster than NAND Flash
Greater endurance than NAND Flash
Lower cost per bit than DRAM
More dense than DRAM

Emerging technology in production phase

„Phase change memory“

Intel® Optane™

→ 128 Gbit-Chips

Need for new, non-volatile memory technologies

Scaling problems

- All 3 main existing memories (SRAM, DRAM, Flash) at limits of scaling

Advantage of non-volatile memories

– Performance : power / speed

→ Portable applications

→ SCM

Requirements

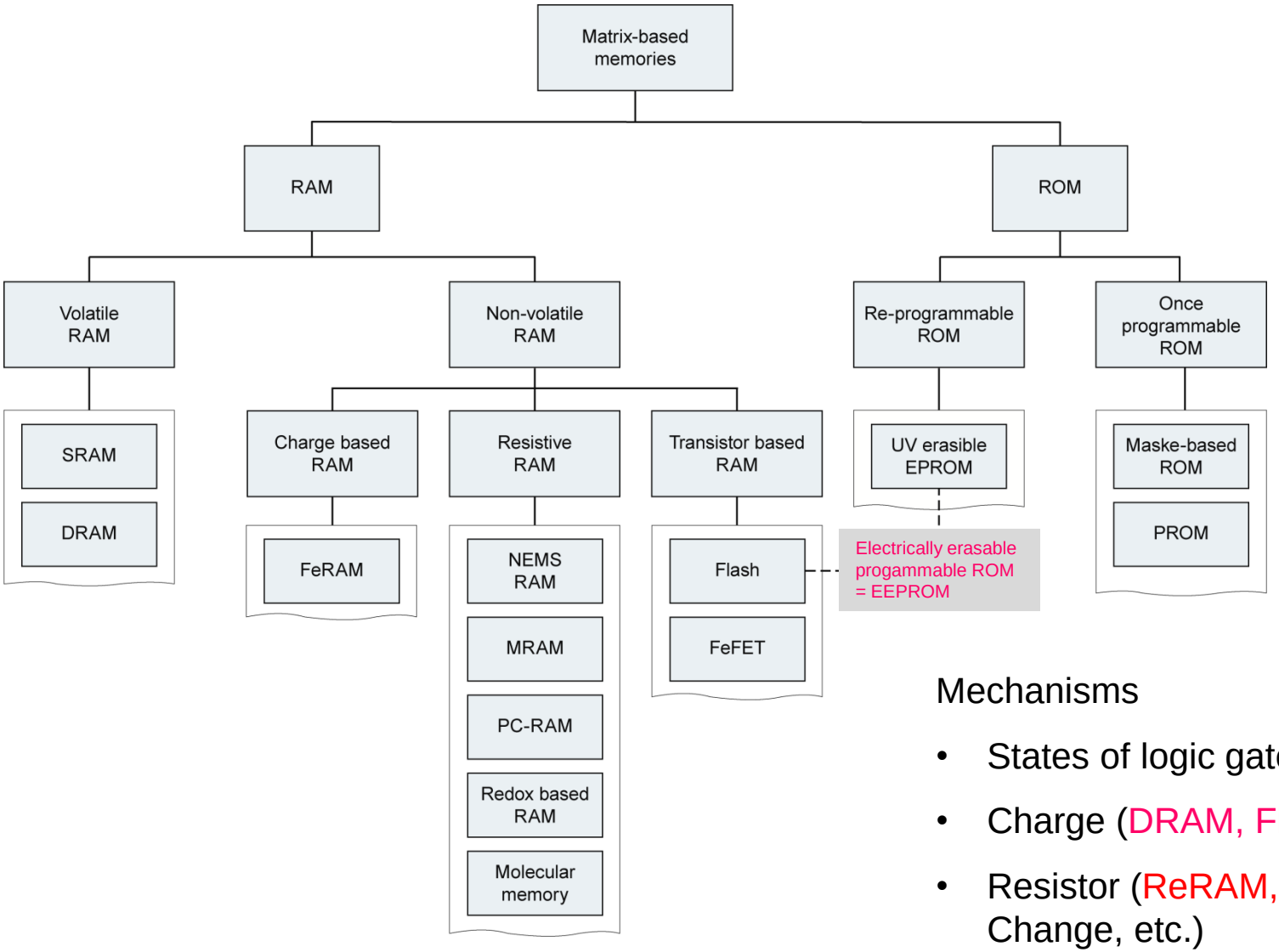
– Small cell size & scalable – high integration density
– Low energy consumption
– Fast R/W
– Large (infinite..) number of R/W cycles
– Compatibility with Si-CMOS technology, little additional processing costs

– Possible technologies

– Resistive switching (Redox, Phase Change)
– Ferroelectric RAM, FeFET
– Ferromagnetic MRAM (tunnel junction), STT
– Carbon Nanotubes CNT
– Molecular concepts

Charge based or
Resistance based

IC = Matrix based memories



Mechanisms

- States of logic gates - Flip-Flop (SRAM)
- Charge (DRAM, Flash, FeRAM, FET, FeFET)
- Resistor (ReRAM, STT-MRAM, Phase Change, etc.)

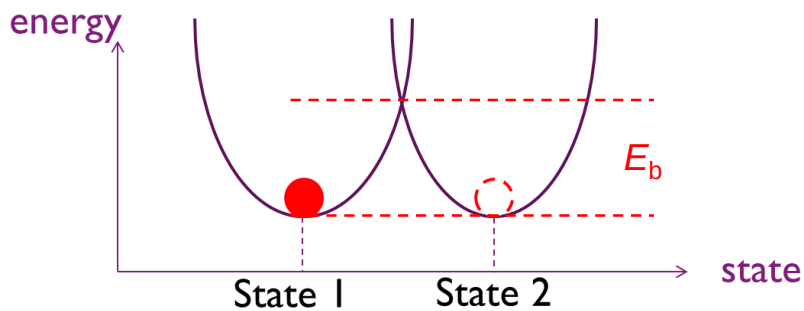
5.3 Requirements for (non-volatile) memories

Requirement for a non-volatile memory

1. Needs "**something**" that can be in 2 "states"

Stable states:

- each state = energy minimum (ideally)
- minimal energy barrier to move the "memory content" out of each of the states (unwanted stimuli, thermal fluctuations!) e.g., barrier for electron tunneling / barrier of ion drift/diffusion

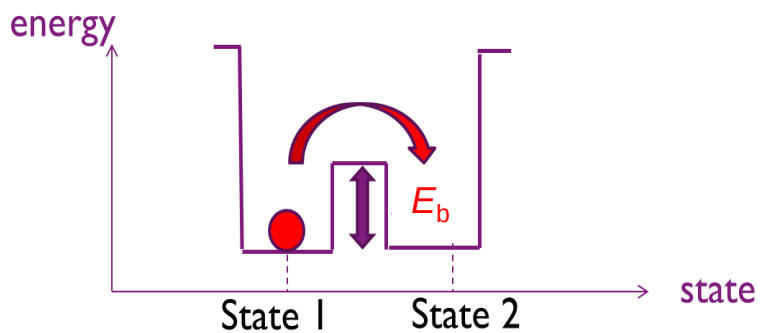


Estimation of Eb (V. Zhirnov, IEEE EDS Webninar 03/14/2013):

$$f_{tr} = f_0 \exp\left(-\frac{E_b}{k_B T}\right) \quad \Rightarrow \quad t_{tr} = \frac{1}{f_{tr}} = \frac{1}{f_0} \exp\left(\frac{E_b}{k_B T}\right)$$

thermal attempt frequency f_0, max. f for high T

$$\Rightarrow E_b = k_B T \ln\left(t_{tr} f_0\right)$$



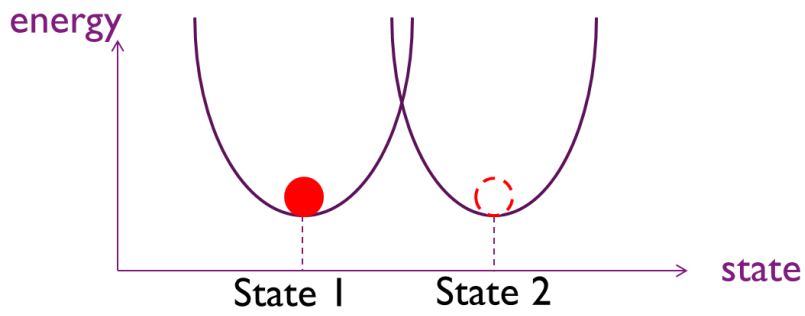
- Storage device requirement: $t_{tr} = 10$ y at $T = 400$ K
- Physical properties for

electrons	atoms/ions	spins
$f_0 \approx 10^{13}$ Hz	$f_0 \approx 10^{12}$ Hz	$f_0 = 10^9 - 10^{10}$ Hz
$E_b \approx 1.7$ eV	$E_b \approx 1.6$ eV	$E_b = 1.4 - 1.5$ eV

$$\Rightarrow E_b = 1.4 - 1.7 \text{ eV}$$

Requirement for a non-volatile memory

1. Needs "**something**" that can be in 2 "states"



Memory state = information "carrier" and "form"

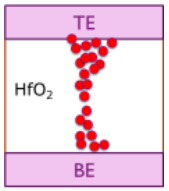
Information carriers:

- Electrons
- Atoms
  - Neutral
  - Charged: ions/dipoles
- Spins
- Molecules
- Mesoscopic structures
- …

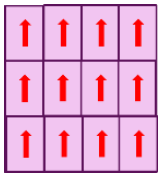
Information form (ordering)

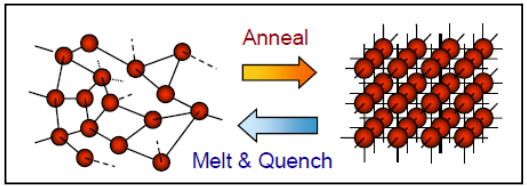
- Bucket type" (No - **0D ordering**)
  - Memory state is number of particles in an energy well
    - e.g. number of electrons in a floating gate or capacitor
    - only number counts, not how they are configured

- "filament type" (or **1D ordering**)
  - Memory state is number of particles in/forming a conducting filament
    - e.g, conductive filament made by Cu ions in CBRAM
    - State is not only depending on number but also on configuration
  - (**2D variant** : "interfacial" switching RRAM)



- "Phase type" : Memory state depends on **3D ordering**
  E.g., FeRAM, MRAM, PCRAM...



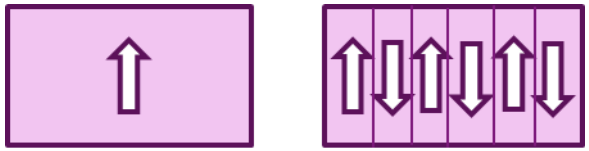


Spins 3D order in Magnetization
Dipoles 3D order in Polarization

Requirement for a non-volatile memory

2. Must be **stable** states:

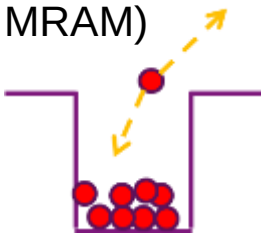
- often one of the states is only a metastable state
  - one has a lower minimum than the other
  - mix may have even less stability
  - e.g, PCRAM:
    - » Amorphous phase is meta-stable
    - » Partially crystallized is less stable than fully amorphized
- not always the right way to depict system stability :
  - in some systems, minimum is a volume mix of both states
  - e.g., multidomain state energetically favorable
    (scale dependent: monodomain for V<Vcrít)



Requirement for a non-volatile memory

3. Need a **WRITE** (Program-Erase) scheme

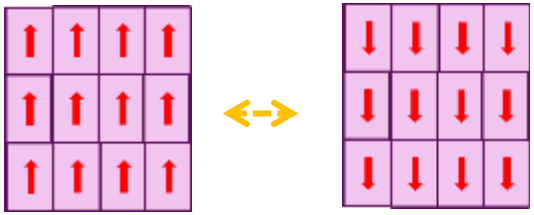
- able to change the state of our memory, in an electrical way
- Writing process: overcome the energy barrier:
  Need to excite the states by DE (E_b resp.) – sufficient fast!
- Alternative:
  Tunnel through energy barrier (e.g. NAND FLASH)
  Decrease of energy barrier
  e.g. by increase thermal energy of the system (thermal assisted write MRAM)



2 main different writing modes:

Individual : particle by particle >> current of particles
Typical for bucket type of memories
But also for 1D and 2D

Collective : “coherent” change of ordering of assembly of particles
Typical of “phase” (3D ordering) type of memories
Assembly of particles typically associated with “domains”



Unipolar / Bipolar write:

Field driven barrier transition : need bipolar write
Thermal energy driven : unipolar write possible, state transition determined by energy min/kinetics

Requirement for a non-volatile memory

4. Need a **READ** scheme

How to sense the state of the memory ?

Direct **counting** of the number of particles
- Counting = current measurement
- Needs charged particles !
- E.g. measure discharge current of DRAM capacitor

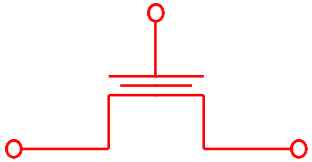
Indirect: by **field effect** generated by the particles
- Needs charged particles
- E.g., current in floating gate transistor
- E.g., dipole compensating charge in ferroelectric capacitor

Indirect: by **resistance change** induced by particles/order/M)
- Can be either charged or uncharged particles or spins
- E.g, Resisitive RAM / Phase Change RAM / MRAM

Memory element structure

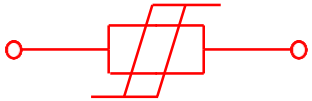
3-terminal device :

Offers decoupling of program and sense (read) operation
e.g. floating gate transistor



2-terminal device :

Programming and reading using the same terminals
Critical for decoupling them! → disturb/retention



Read / Sense at lower voltages than those that induce programming

Need highly nonlinear programming process

E.g. RRAM : need high voltages to mobilize ions, at low V only electrons mobile...➔ **voltage-time dilemma**

Alternative: same mechanism & destructive readout (DRO)

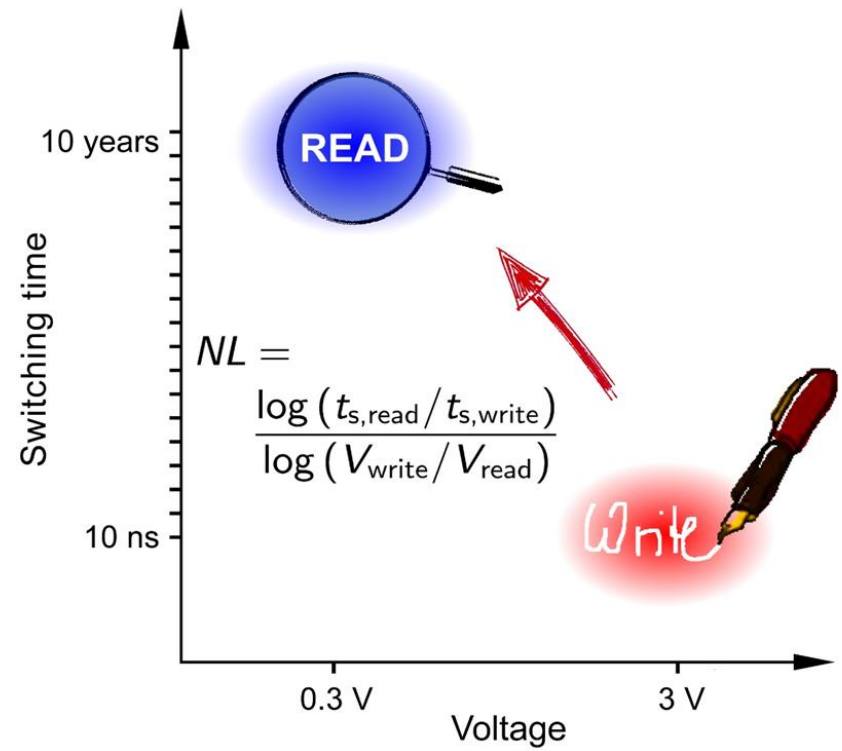
E.g. ferroelectric capacitor memory

Requirement for 2-terminal non-volatile memory

Strong non-linearity of switching kinetics with voltage
= Voltage – time dilemma

- Fast programming
  At ~**3V**, need programming in **~ 10nsec**

- Good retention
  At ~**0.3V**, need stable read for **~ 10 years**

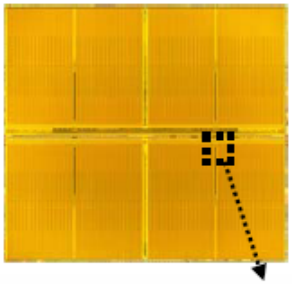


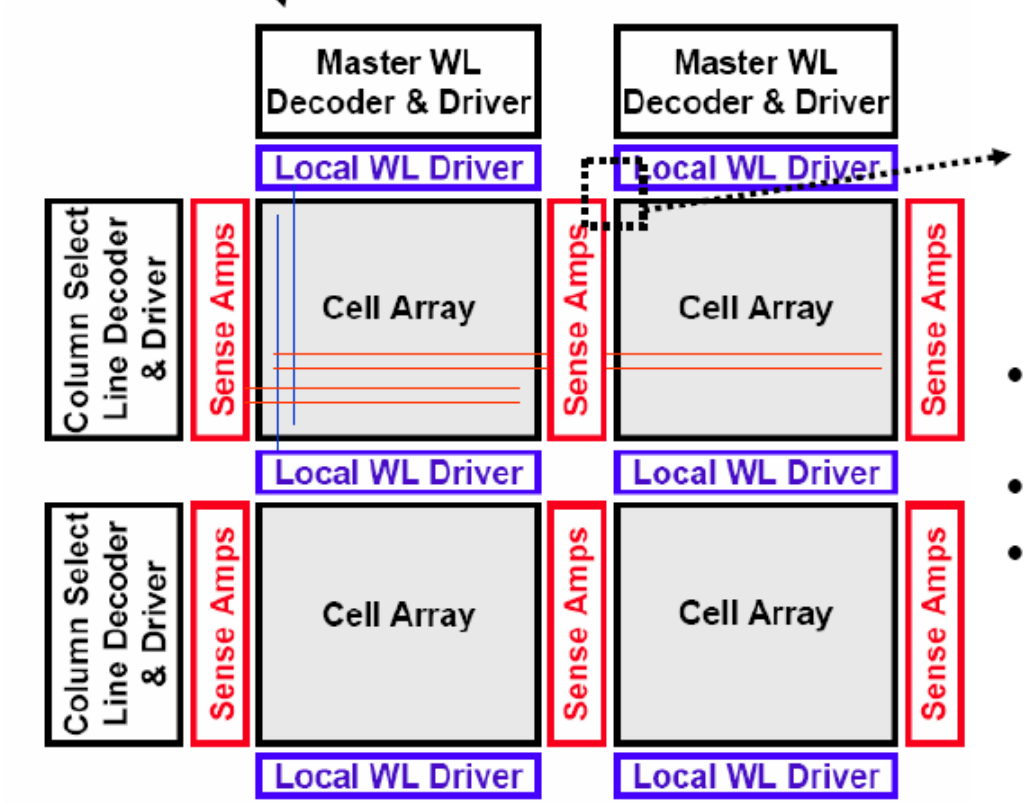
Need operation mechanism with extreme non-linear behavior:

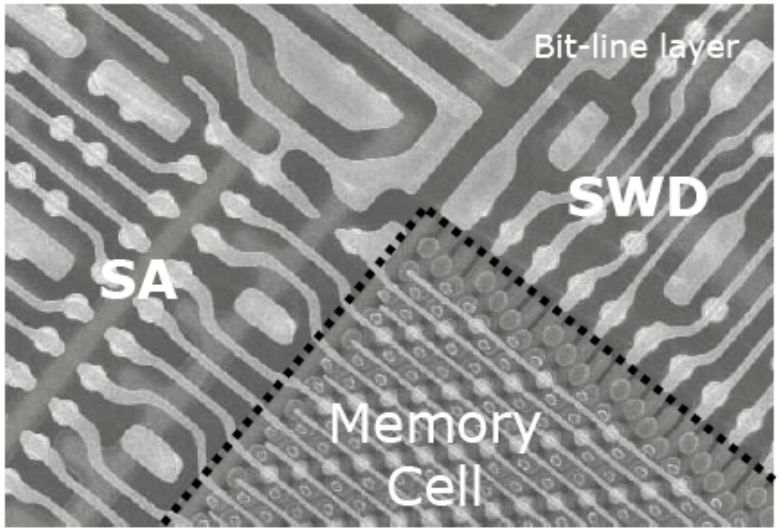
>15 orders of magnitude in time over 1 decade in voltage

5.4 Matrix organization of memories

Memory Structure



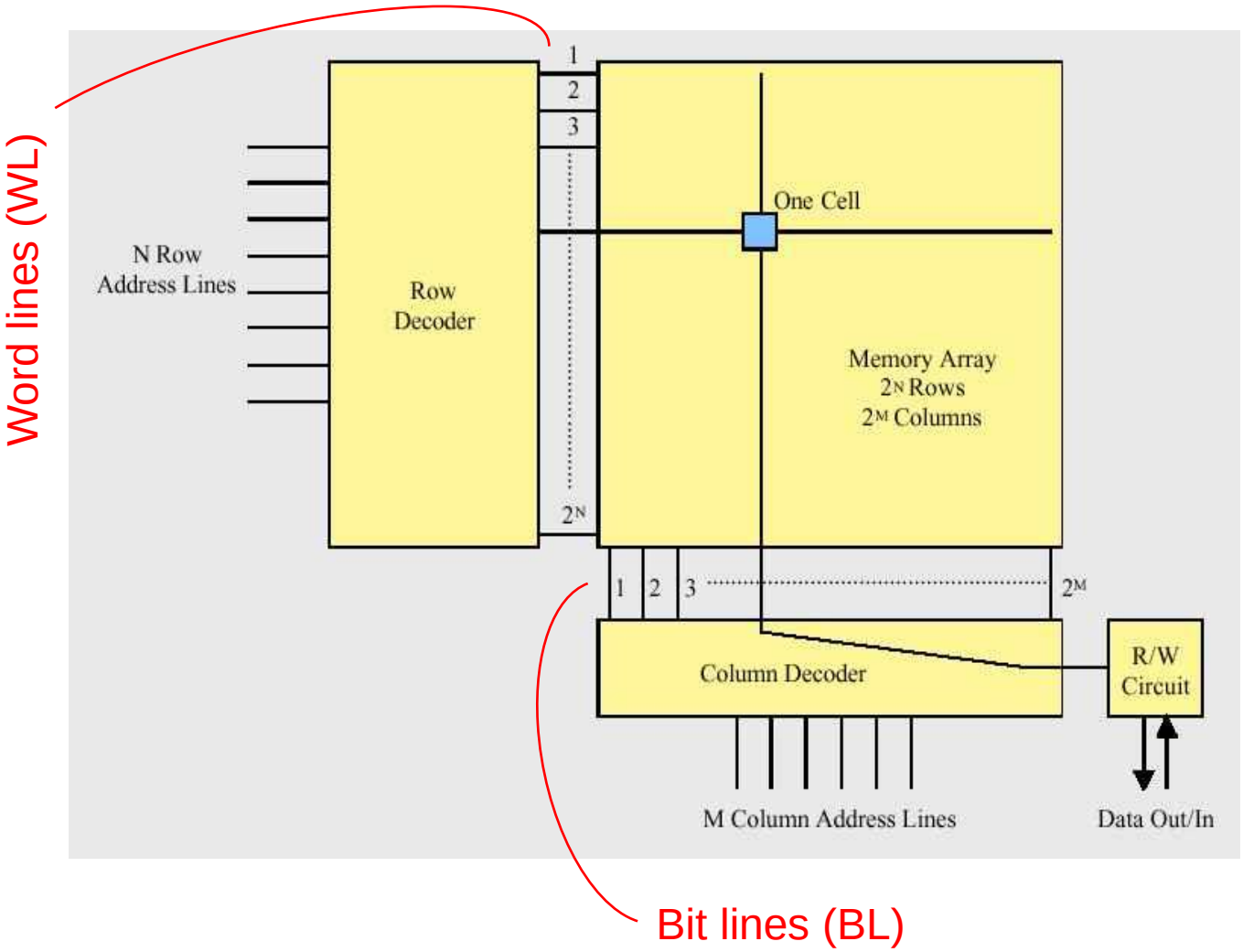




- Arrays are divided and partially activated.
  » Low power / Low C_{par}
- Cell arrays are surrounded by SA & SWD.
- Periodic patterns in both SA and SWD regions

* Sense amplifier
Sub (or local) WL driver

Matrix organization



Address bus

- N rows
- M columns

$\Rightarrow 2^{N+M}$ cells addressable

e.g. N=M=2

$\Rightarrow$ 16 cells

e.g. N=M=10

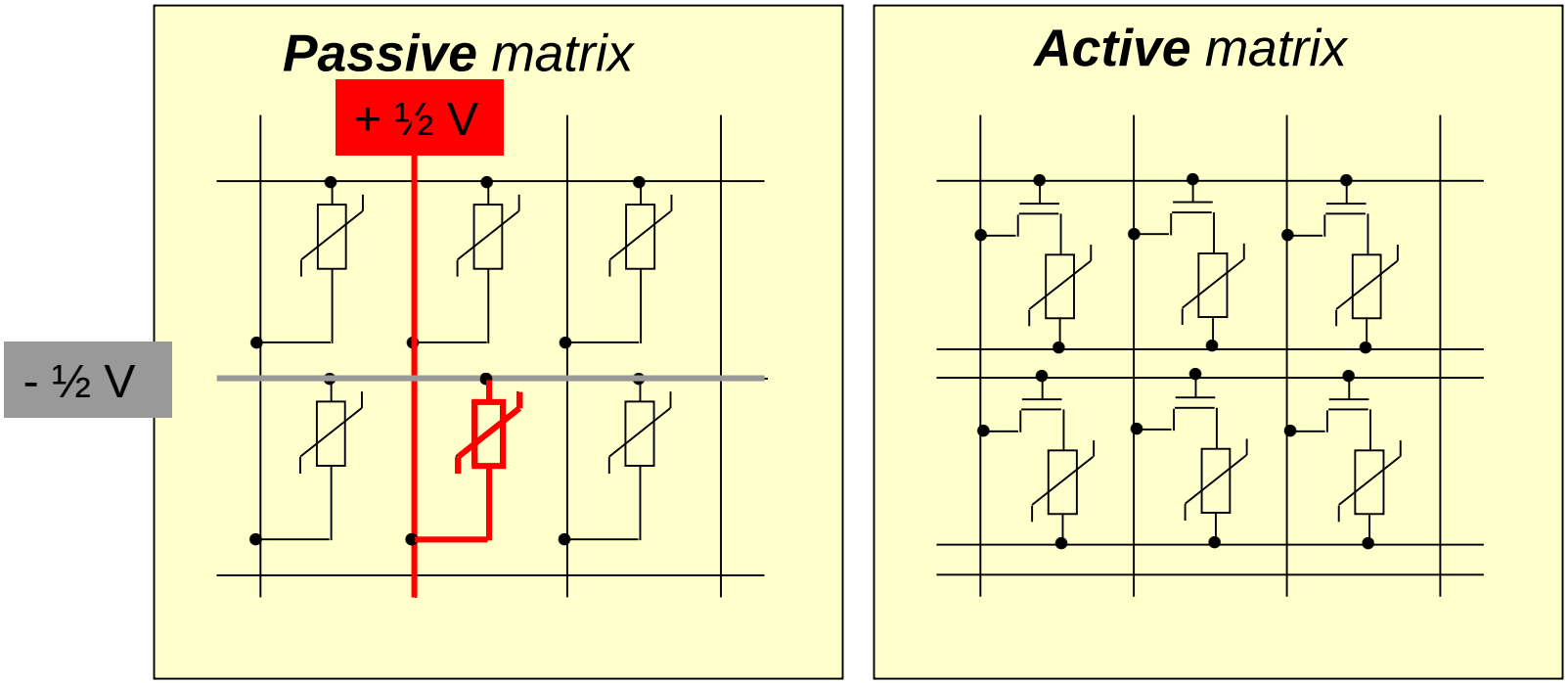
$\Rightarrow$ 1.048.576 cells = 1 Mb

Passive vs. active memory matrix

RAM:

- fast, random access to each cell

1/2 Voltage scheme

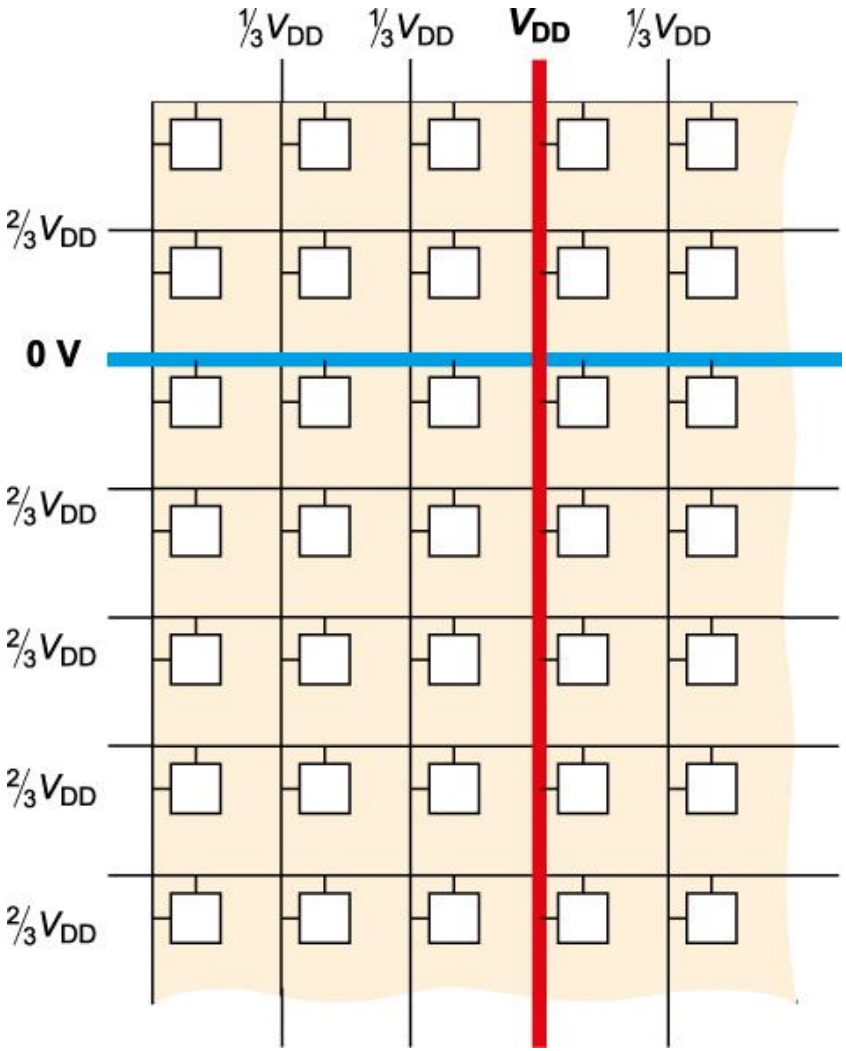


arbitrary binary storage element

- Total voltage V = ½ V – (- ½ V) only across the selected element
- All other elements: V = ½ V or V = 0 V

Other scheme‘s also possible, as $^1/_3$ V

Passive matrix: 1/3 Voltage scheme

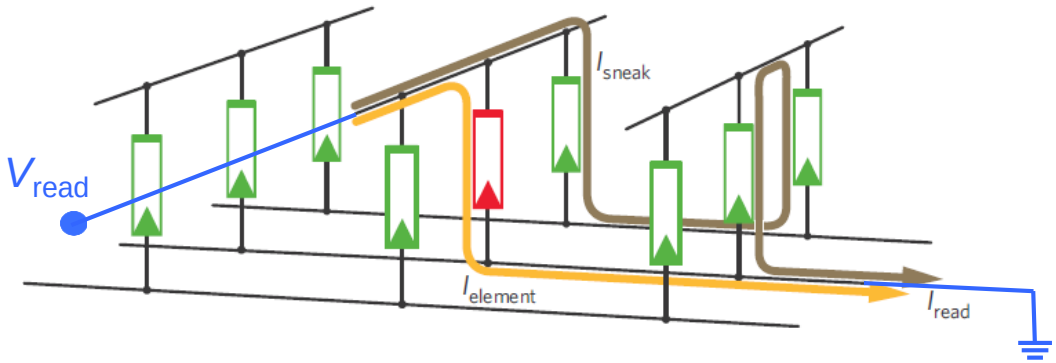


Issues of passive matrix (e.g. resistive memories)

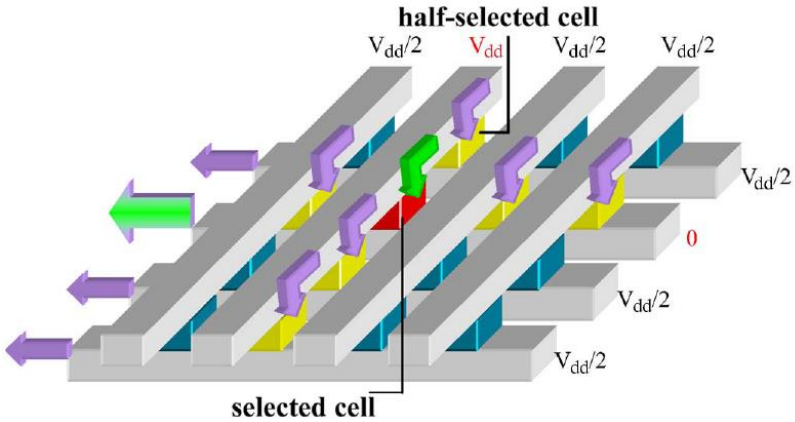
- Read errors due to sneak current paths
- Program disturbs on half-select cells (1/2 or 1/3 V scheme)
- Power dissipation due to current through half-selected cells

HRS: High resistive state („OFF“)

LRS: Low resistive state („ON“)



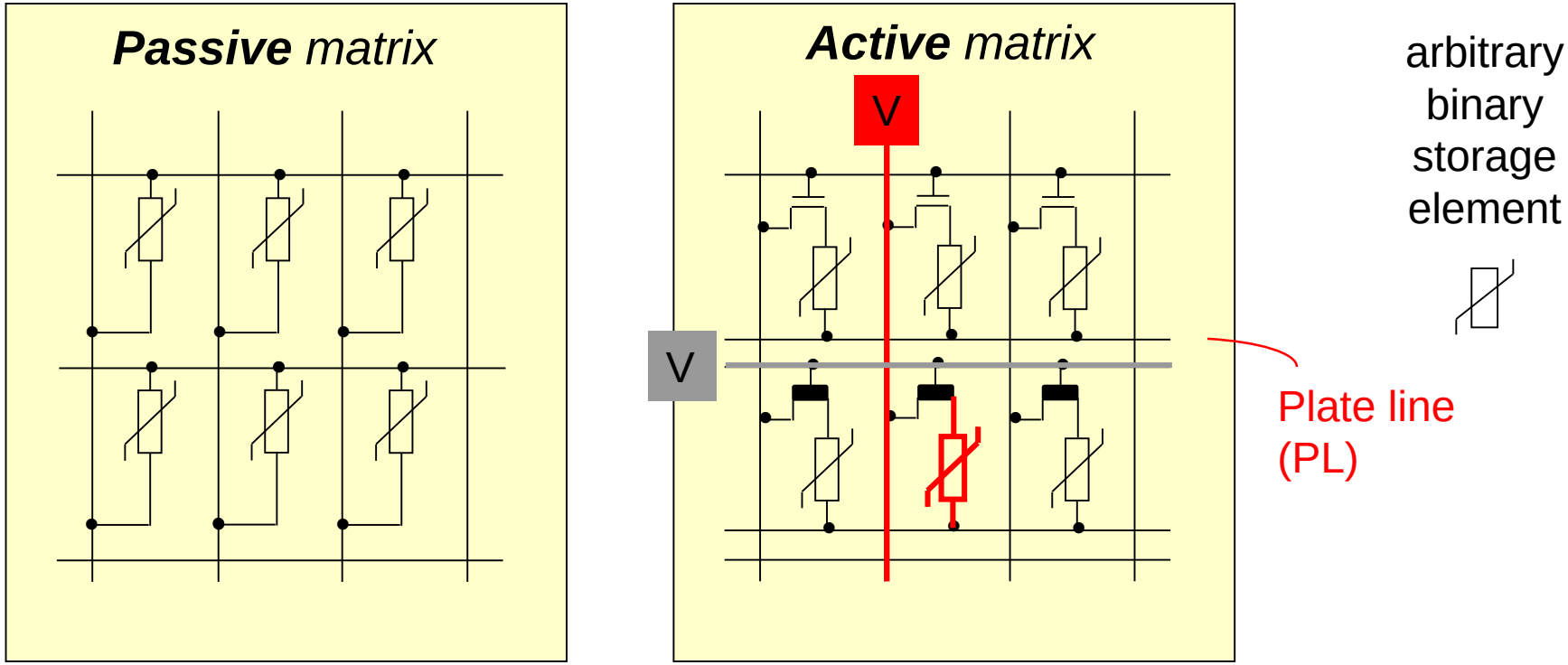
E.Linn en al, NATURE MATERIALS VOL 9 p. 403 MAY 2010



J.Liang et al., IEEE TRANSACTIONS ON ELECTRON DEVICES, VOL. 57, NO. 10, p.2532 OCTOBER 2010

Passive vs. active memory matrix

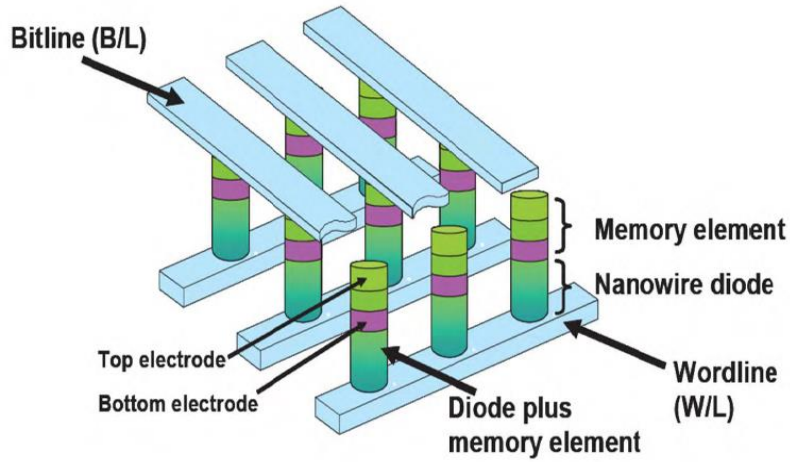
RAM:



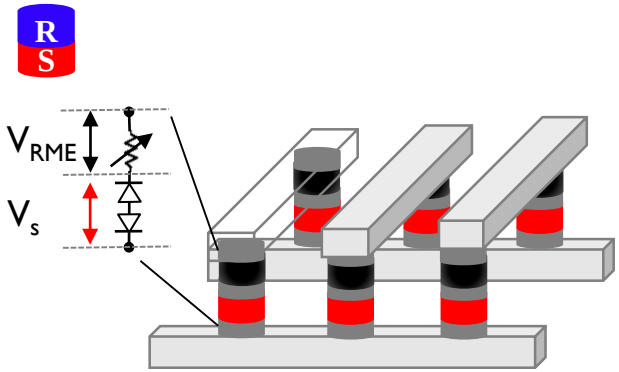
- Voltage only across the selected element by
  Additional **SELECTOR** element (here: transistor)
  - Larger cell
  - Additional connection required (transistor GATE)

Other possibilities: passive selectors (e.g., DIODE)

- Small 2 terminal element (diode) that can be stacked on top of the memory element
  - Small cell size still possible
  - Limited selectivity (non-linearity worse than transistor)



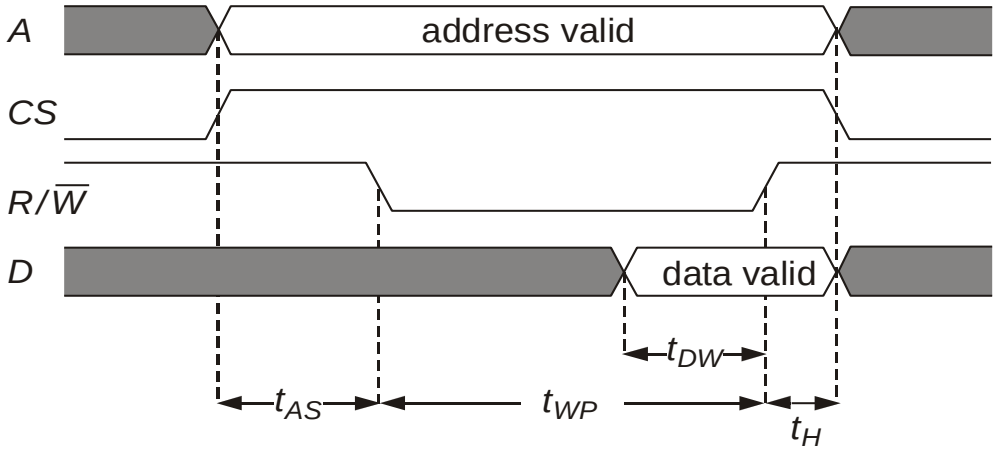
P.Wong (Stanford)



Along the sneak path:
diode(s) in blocking state

5.5 Timing Schemes of read and write

Timing scheme



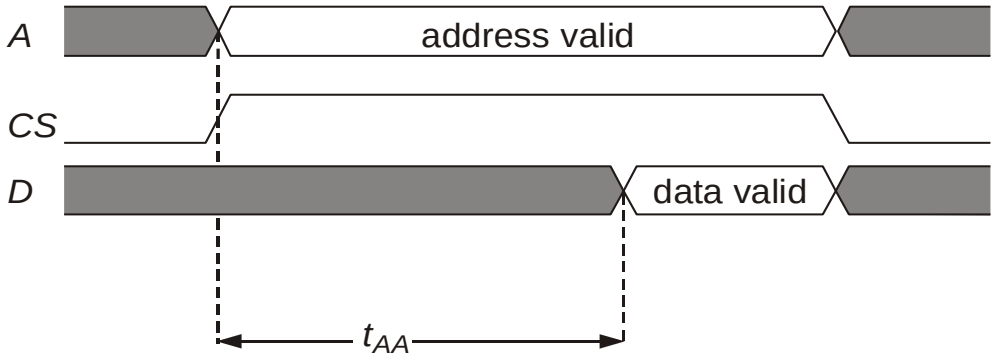
Write

t_{AS}: Address Setup Time

t_{WP}: Write Pulse Width

t_{DW}: Data Valid to End of Write Time

t_{H}: Hold Time

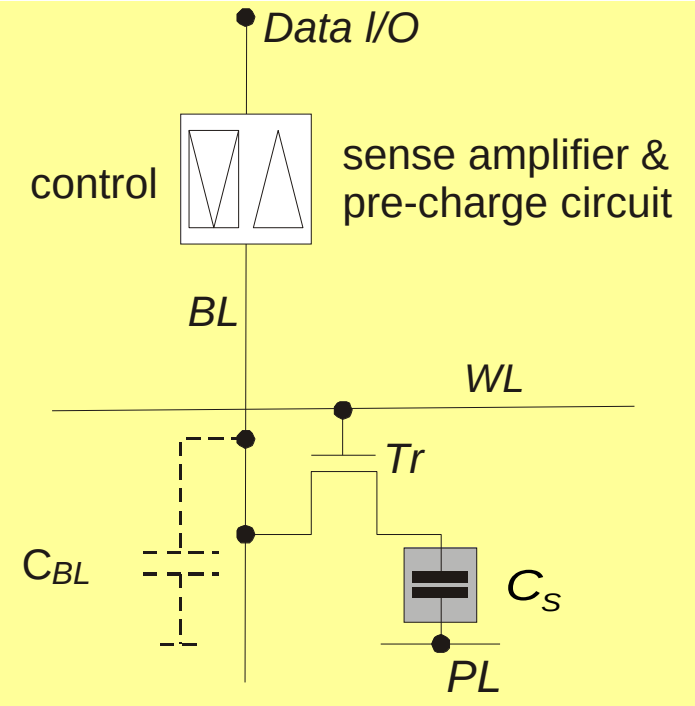


Read

t_{AA}: Address Access Time

Charge-based RAM: *Read*

1T-1C cell (Node)

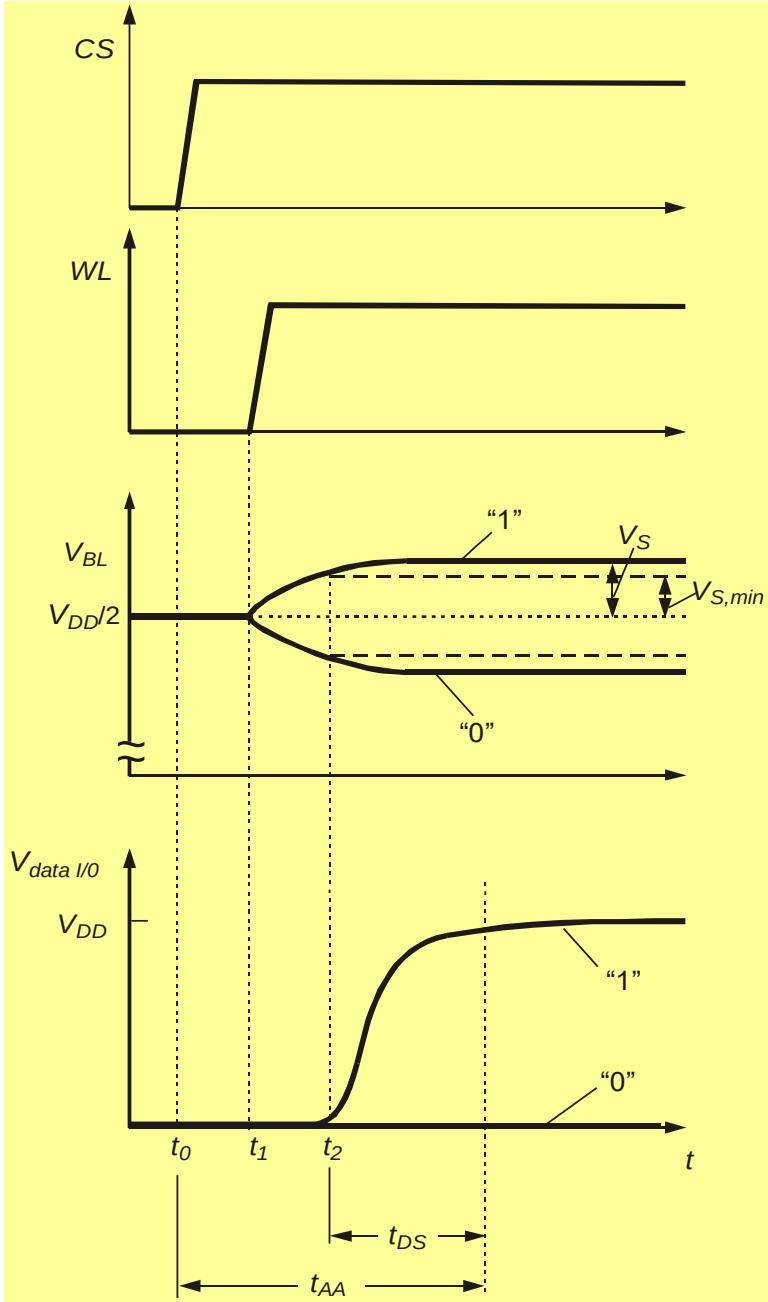


DRAM

$$V_{BL} = \frac{V_{DD}}{2}\left[1 \pm \frac{C_S}{C_{BL} + C_S}\right]$$

$$= \frac{V_{DD}}{2} \pm V_S$$

with $V_S = \frac{C_S}{C_{BL} + C_S}\frac{V_{DD}}{2}$

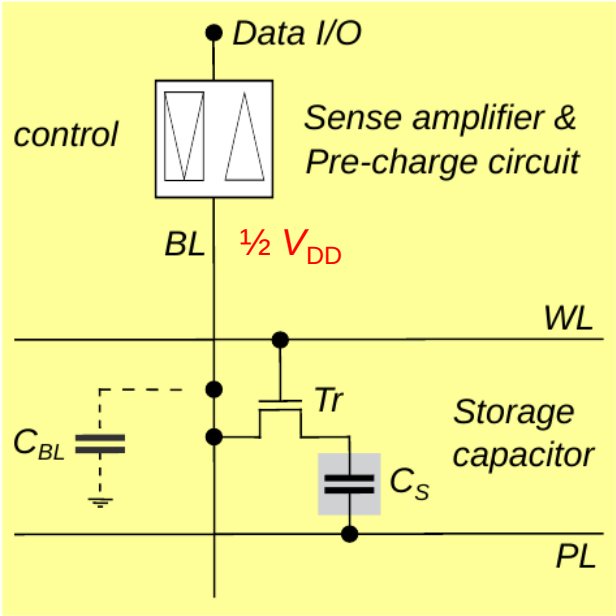


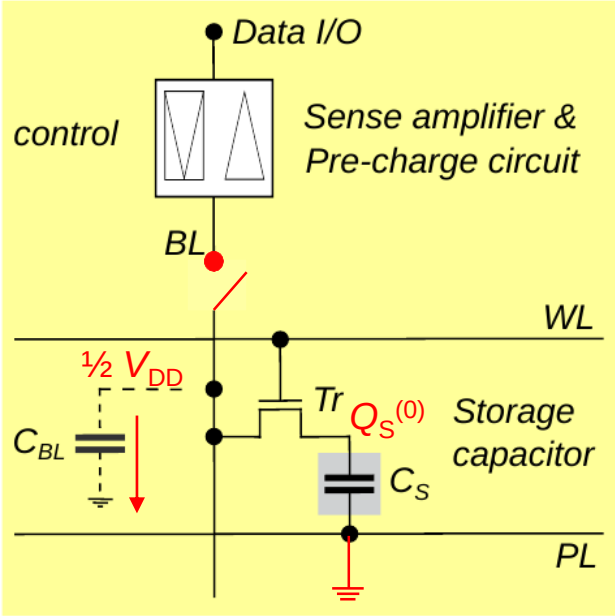
Tr select transistor

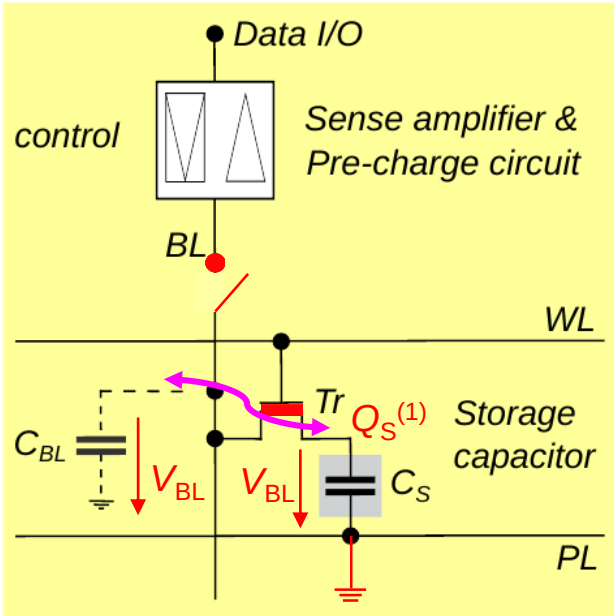
C_S storage capacitor

PL(DRAM) = GND

Charge-based memories – DRAM read







- precharging BL:

  $V_{BL}^{(0)} = ½\, V_{DD}$

- cut-off BL: separated charges

  BL: $Q_{BL}^{(0)} = ½\, C_{BL}\, V_{DD}$

  cap: $Q_S^{(0)}$

- Addressing Tr → parallel path to GND → redistribution of Q:

  $Q_{BL}^{(0)} + Q_S^{(0)} = (C_{BL} + C_S)\, V_{BL}$

- Evolution of V_{BL}:

$$V_{BL} = \left(1 - \frac{C_S}{C_{BL} + C_S}\right) \frac{V_{DD}}{2} + \frac{Q_S^{(0)}}{C_{BL} + C_S}$$

- Status of C_S: **uncharged**: $Q_S^{(0)} = 0$

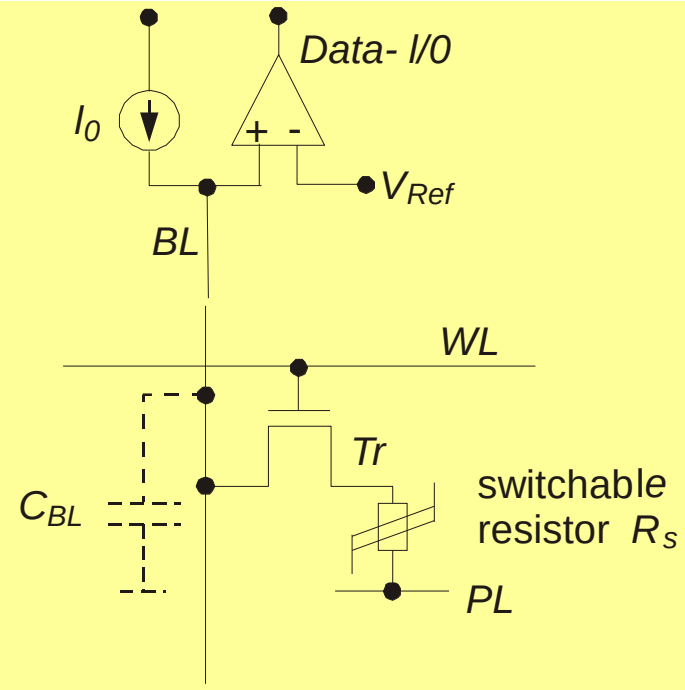
  charged: $Q_S^{(0)} = C_S\, V_{DD}$

$$V_{BL} = \frac{V_{DD}}{2} \pm V_S$$

$$\text{with } V_S = \frac{C_S}{C_{BL} + C_S} \frac{V_{DD}}{2}$$

Resistance-based RAM: Read

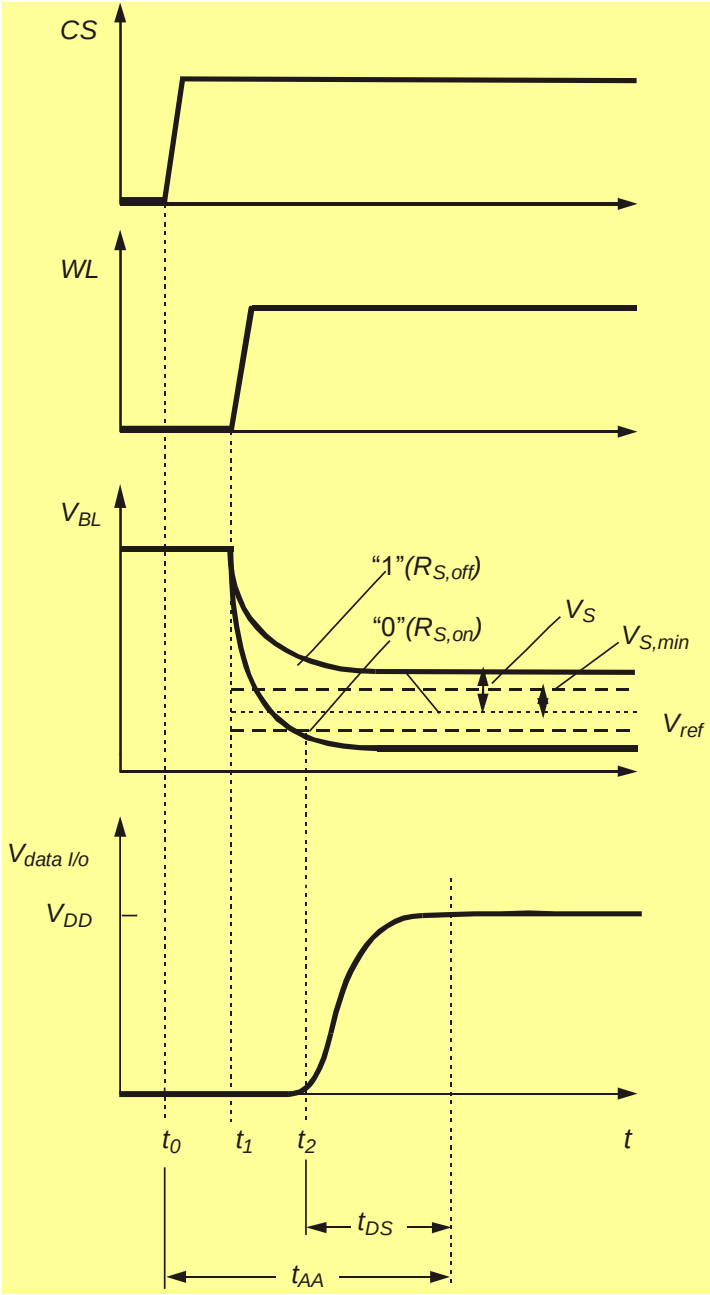
1T-1R cell (Node)



$$V_{BL}^{(on,off)} = \left(R_L + R_S^{(on,off)}\right) I_0$$
$$= V_{REF} \pm V_S$$

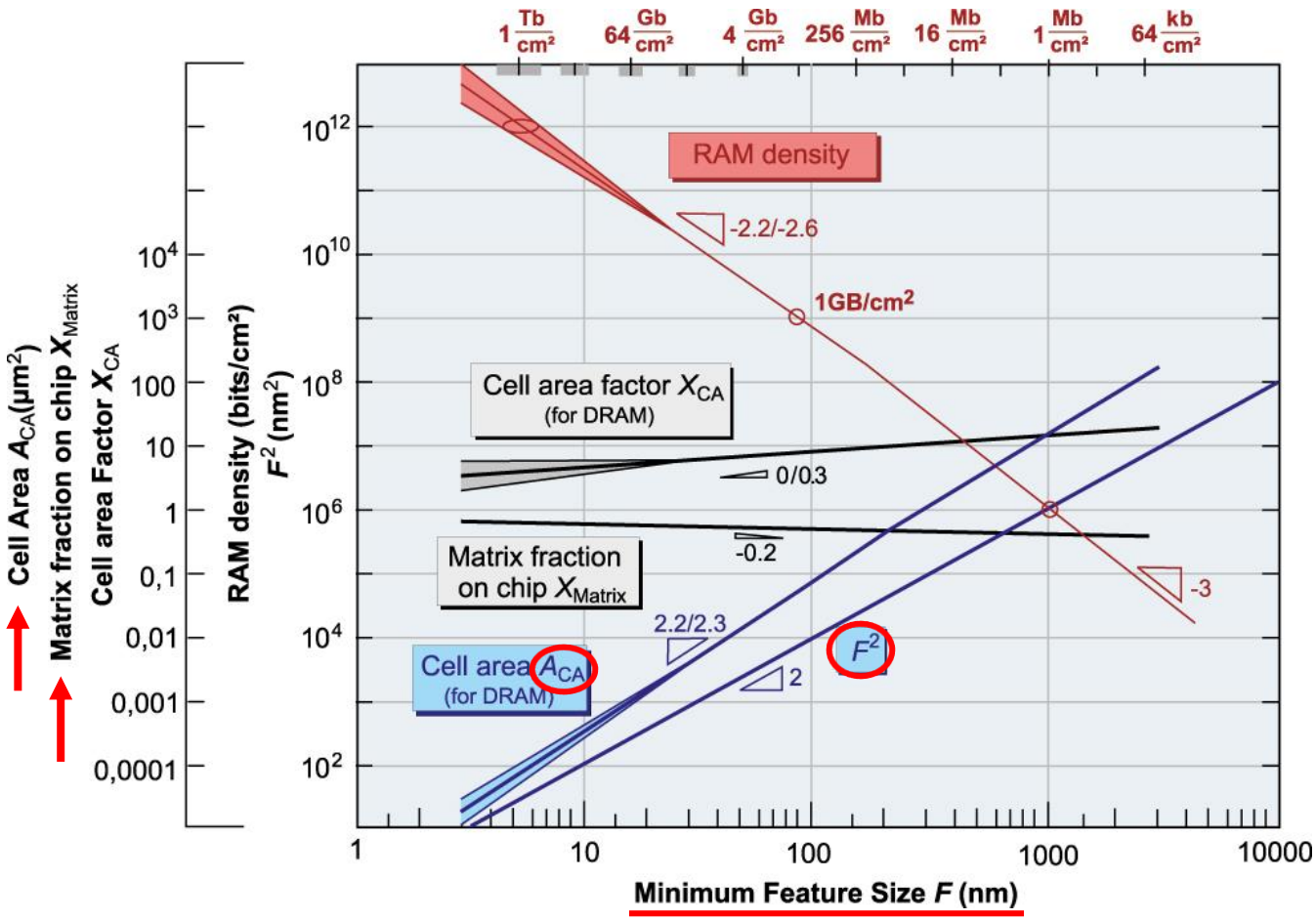
$$V_S = \frac{I_0}{2}\left(R_S^{(off)} - R_S^{(on)}\right)$$

ReRAM, MRAM



5.6 General scaling rules

Geometrical scaling aspects



$$X_{Matrix} = 0.6...0.7$$

Cell size in F^2

$$A_{CA} = X_{CA}\, F^2$$

DRAM: 6

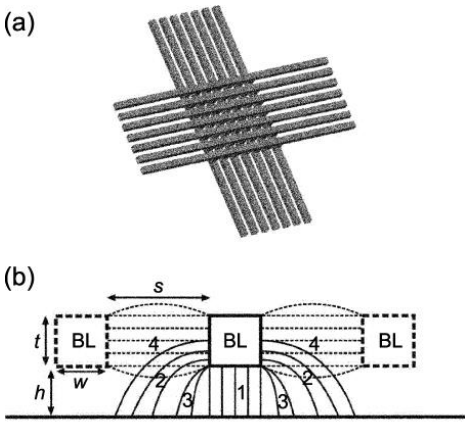
RAM density =

$$X_{Matrix}/A_{CA} = X_{Matrix}/(X_{CA}\, F^2)$$

! Limits for e.g. DRAM: Cap scaling,
Here only geom. Considerations.

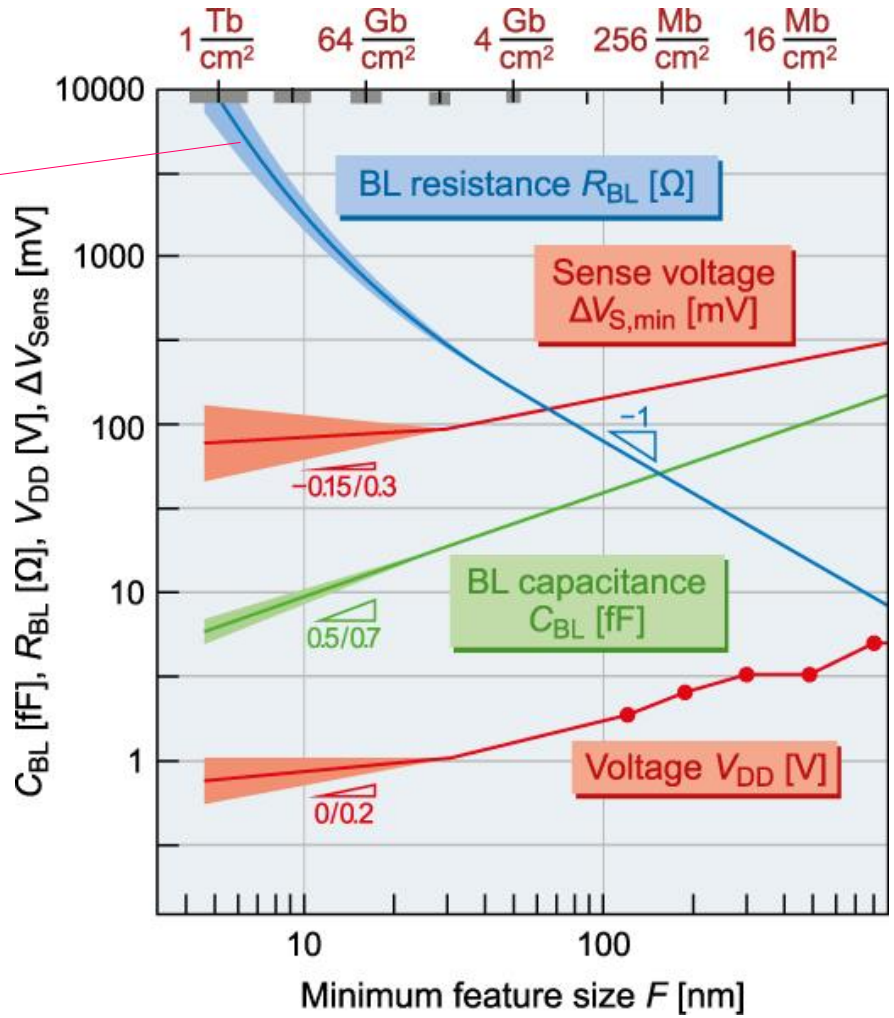
Electrical scaling aspects

Ideal (empty) memory crossbar array



Contributions to the bitline capacitance

Fuchs-Sondheimer Relation
(Scattering at the ME surface)



V_{DD} ↘
C_{BL} ↓
V_S ↘
R_{BL} ↑↑

$$Q_S = V_{S,min} C_{BL}$$