L3: Introduction to Memory II

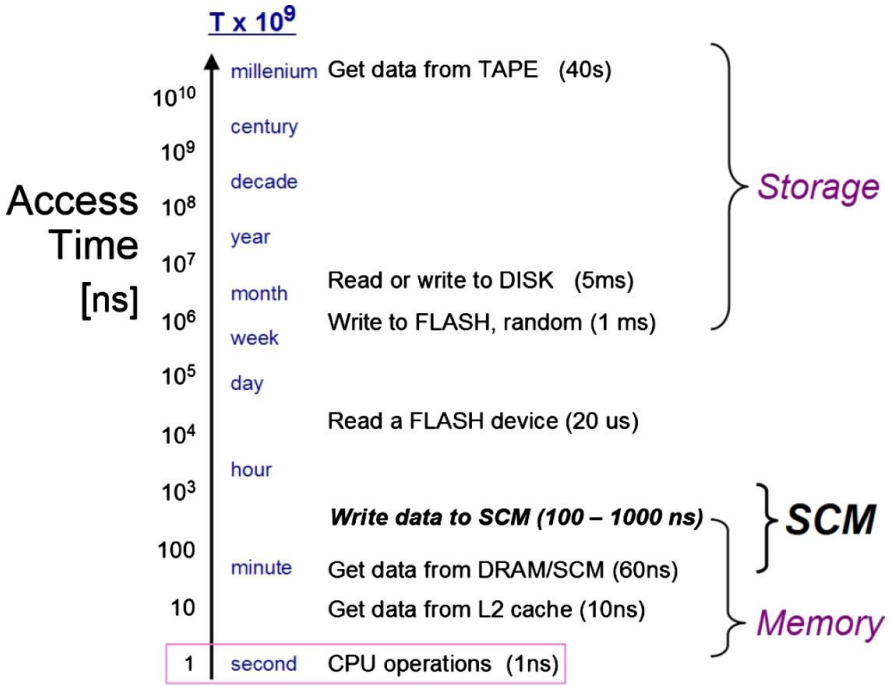
DRAM, SRAM, Flash, and Optical Memories

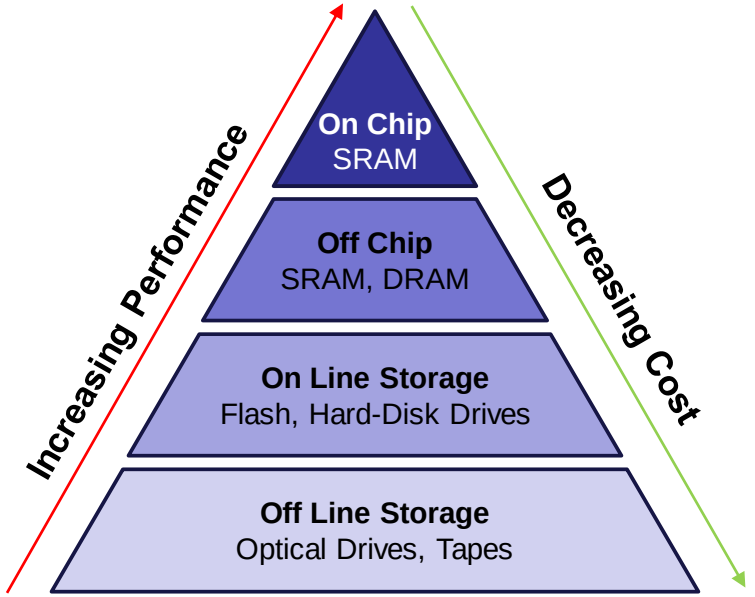
Individual Term Paper Presentation Schedule

- Finalize your topic with the instructor first before you can sign up for a presentation slot (first-come-first-serve)
- 29 slots:
  - 4 slots on Feb 12th
  - 7 slots on Feb 19th
  - 7 slots on Feb 26th
  - 4 slots on Mar 18th
  - 4 slots on Mar 25th
  - 3 slots in Apr 1st
- Finalize your topic by next Wednesday, **Feb 5th**
- Link: https://goo.gl/BkgftK
- Written report (4-page IEEE-style paper) due on **Feb 26th**

Recap

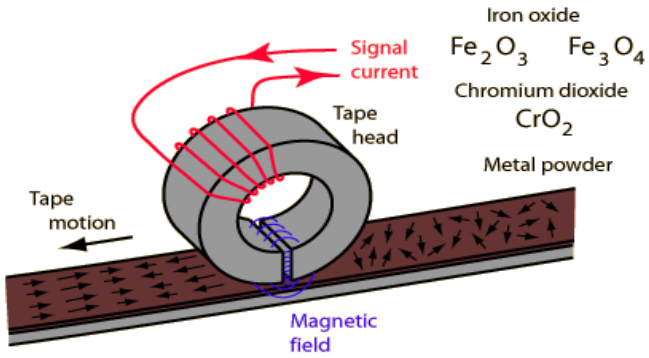
- History of memory
- Memory vs. Storage
- Memory hierarchy
- Important traits
  - Cost
  - Speed
  - Density/Capacity
  - Energy
  - Scalability
  - Endurance
  - Volatility
  - Radiation



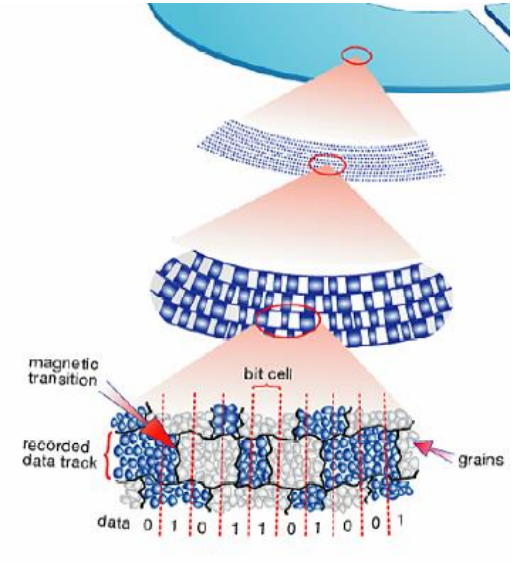


Recap: MRAM

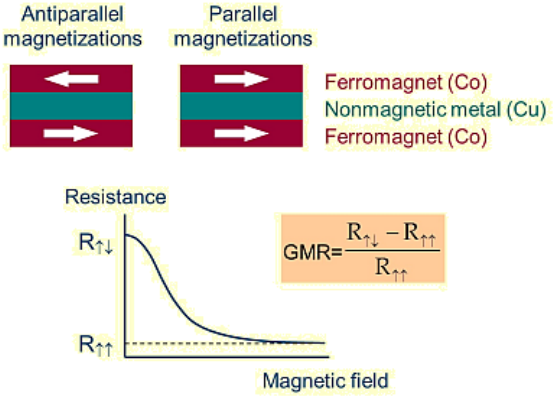
Magnetic Tape



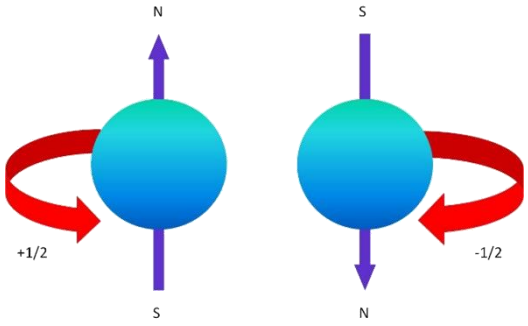
HDD



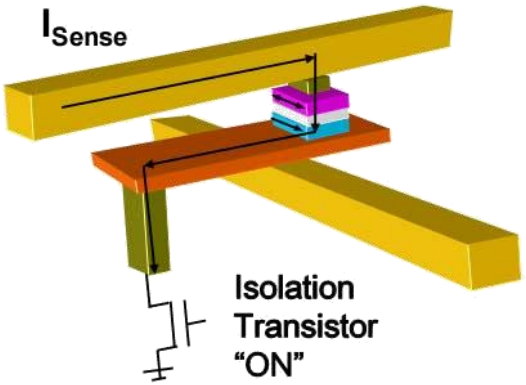
GMR Effect



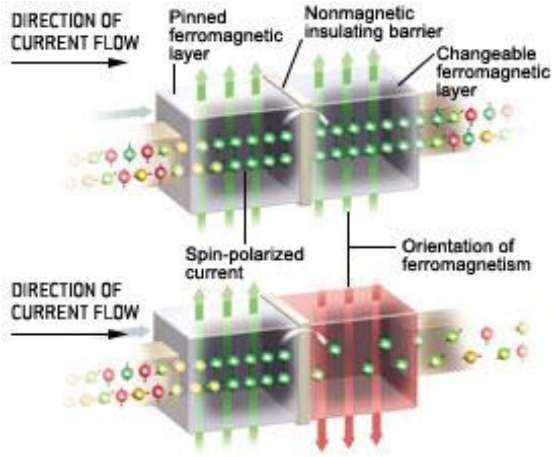
Spintronics



MRAM



STT RAM

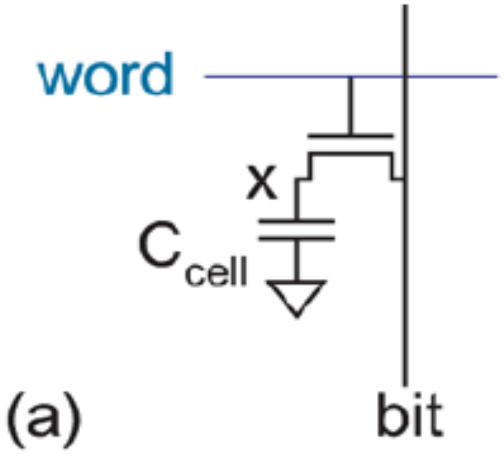


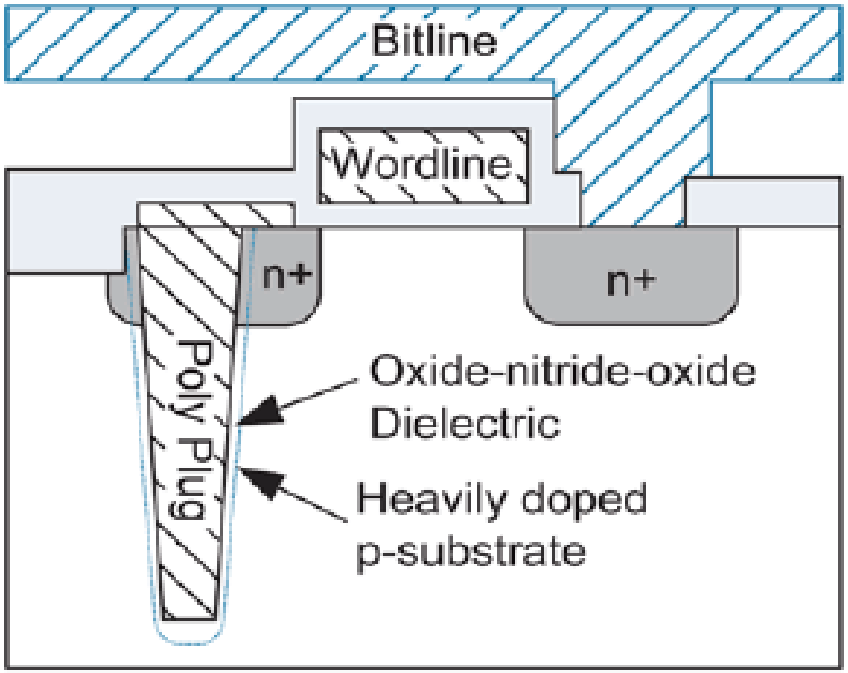
Dynamic Random Access Memory (DRAM)

- DRAM based on capacitive charge storage
- High speed
- Nearly infinite endurance >10^{16}
- Volatile and needs constant refresh
- Main memory for computers
- Moderate capacity

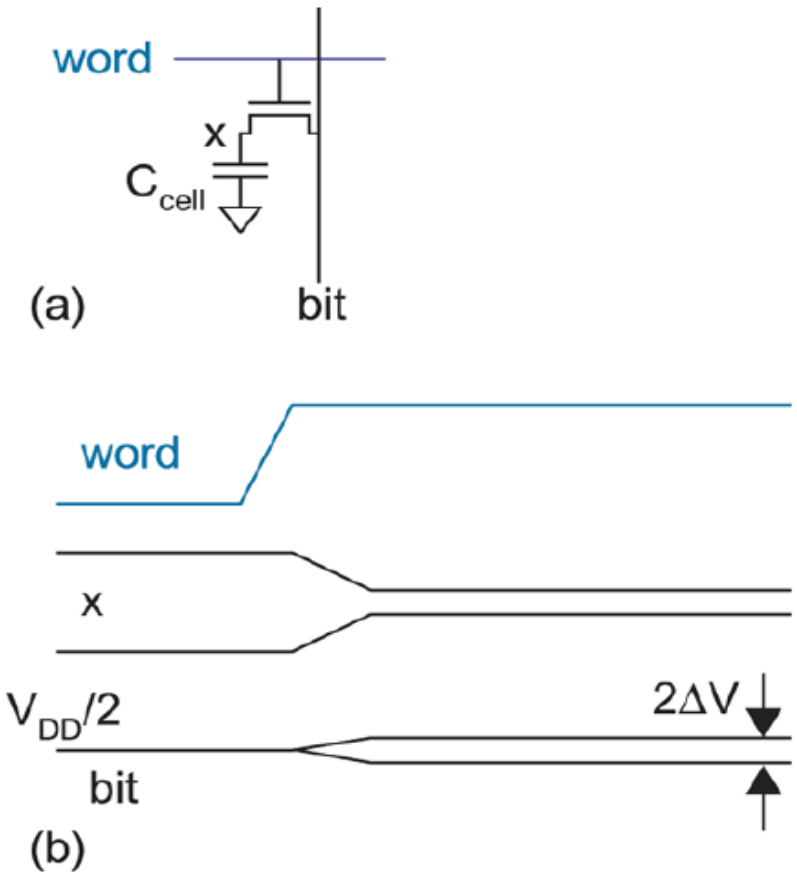
DRAM Cell

- Information stored as charge in the capacitance
- 1T1C – 1 transistor + 1 capacitor



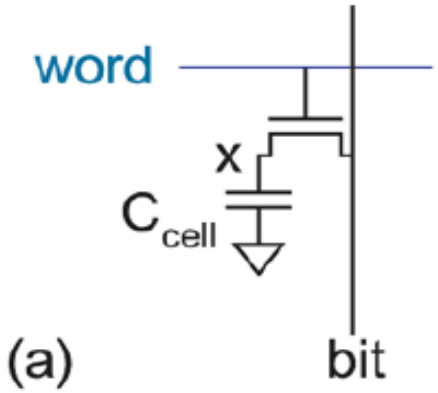


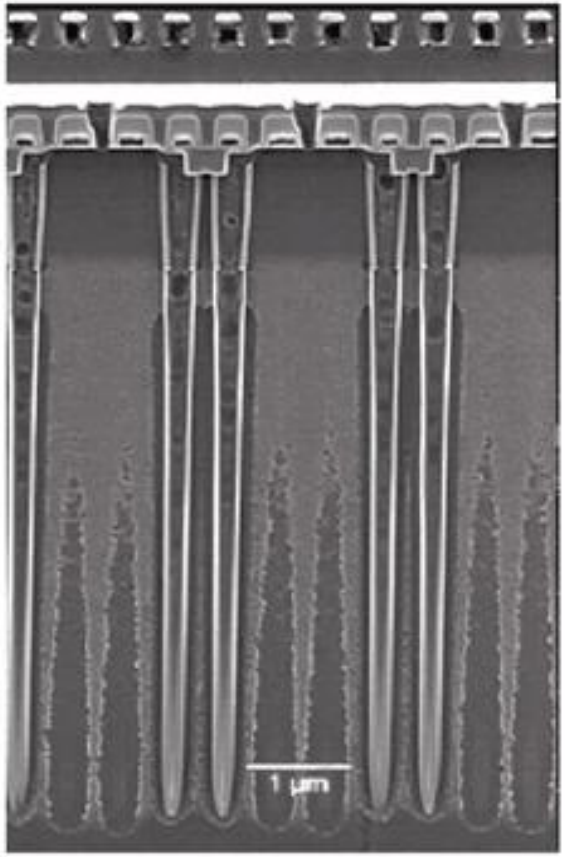
DRAM: Read



1. Bitline precharged to $V_{DD}/2$
2. Wordline rises, cap. shares it charge with bitline, causing a voltage ΔV
3. Read disturbs the cell content at x, so the cell must be rewritten after each read

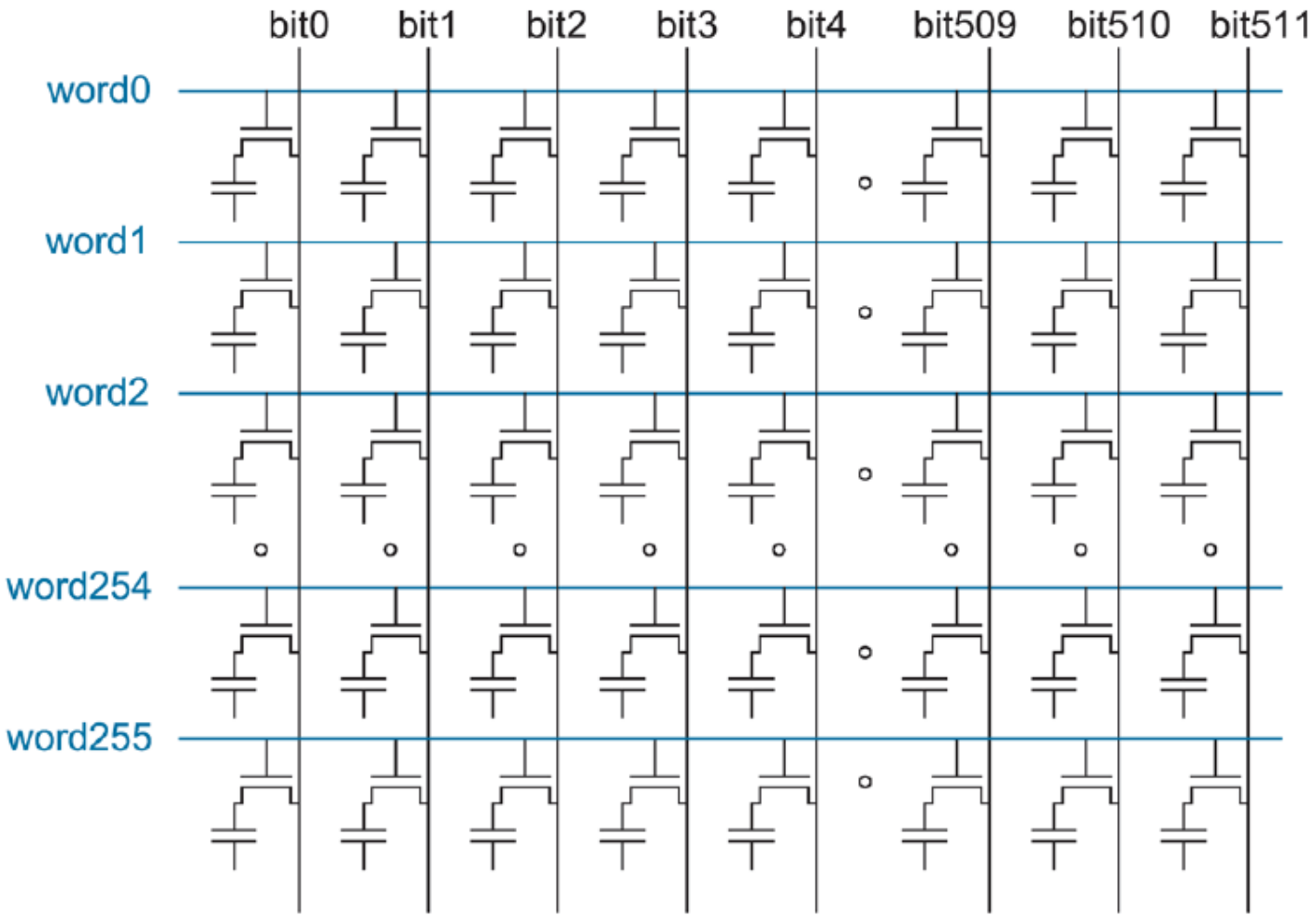
DRAM: Write





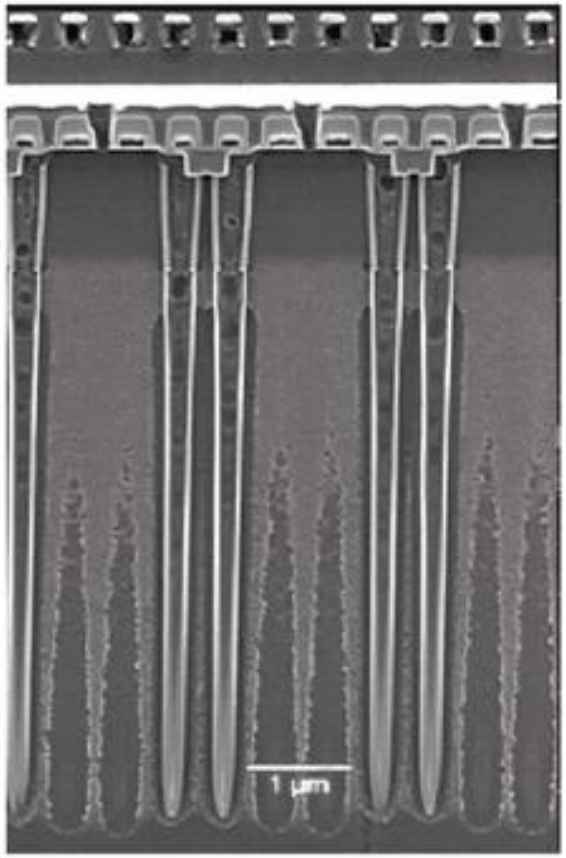
- Write – bitline is driven high or low
- Fill and empty the capacitor
- Needs constant refreshing ~ 64 ms
- Refresh = a dummy read

DRAM Array


bit0
bit1
bit2
bit3
bit4
bit509
bit510
bit511
word0
word1
word2
word254
word255

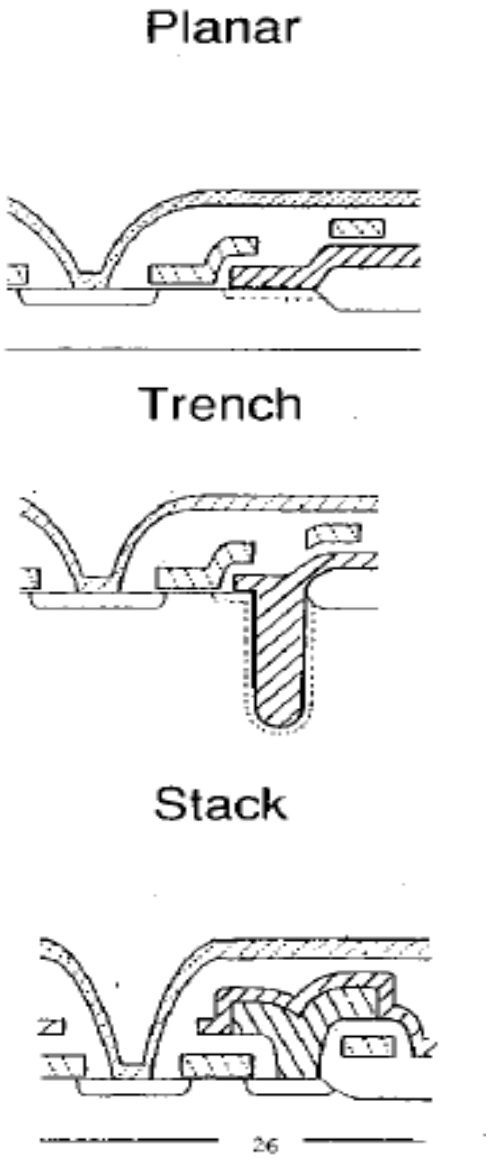
DRAM Size

- Capacitor is an order of magnitude larger than the transistor
- Cap needs to store enough charge to drive the bitline
- Limiting the density

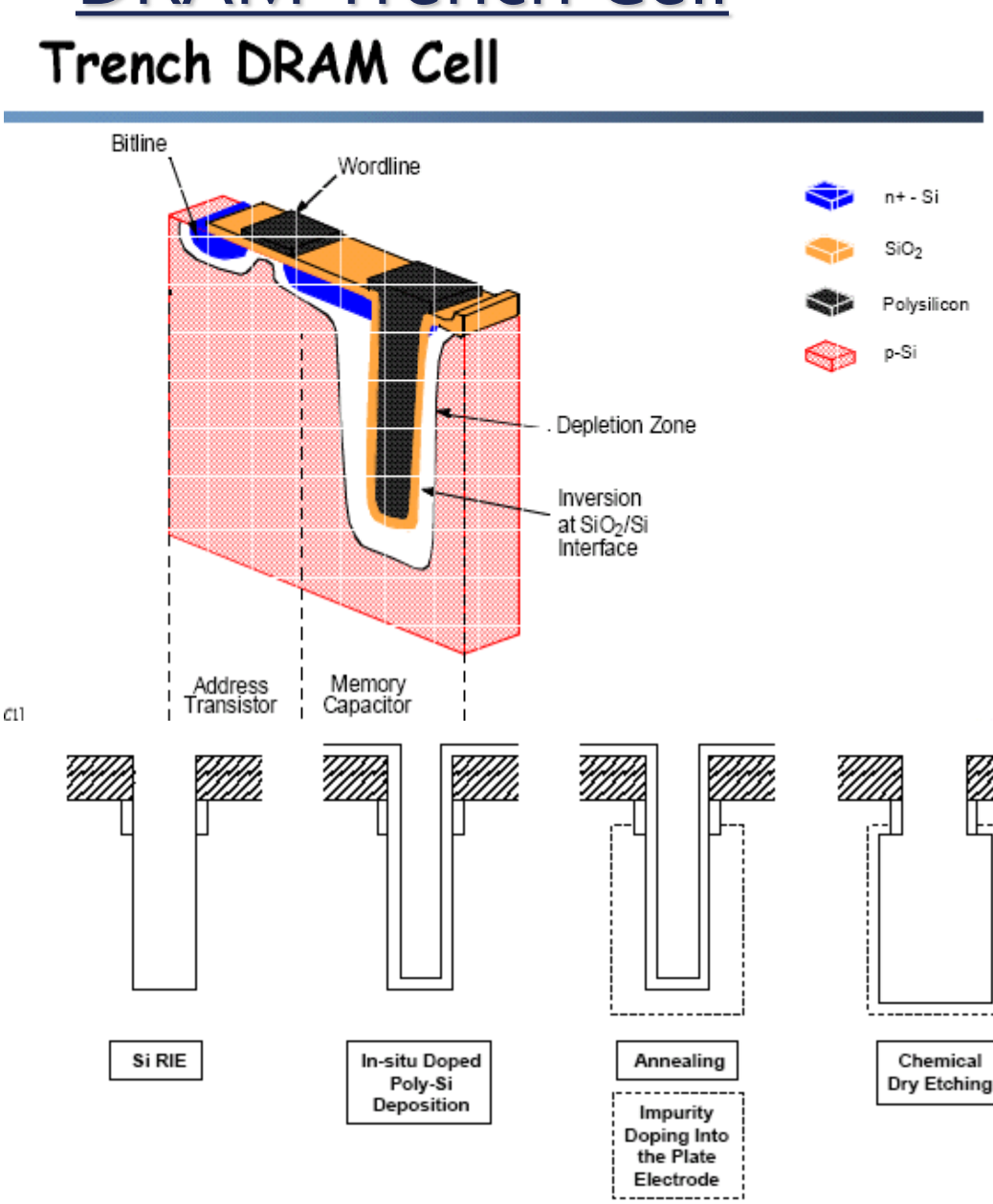


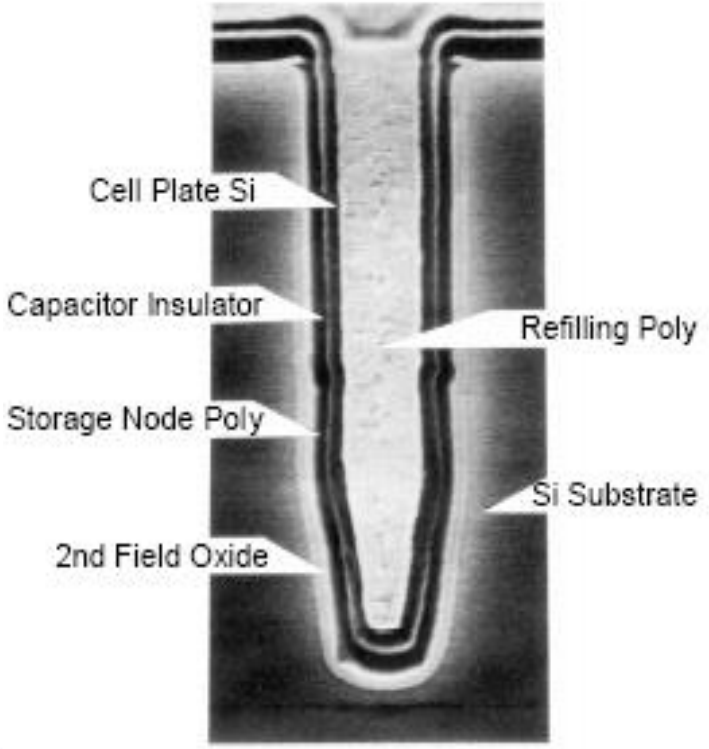
Evolution of DRAM Cell Structure

- Planar Capacitor
  - Up to 1Mb
  - C decreases linearly with feature size
- Trench Capacitor
  - 4–256 Mb
  - Lining of hole in substrate
- Stacked Cell
  - > 1Gb
  - On top of substrate
  - Use high ε dielectric



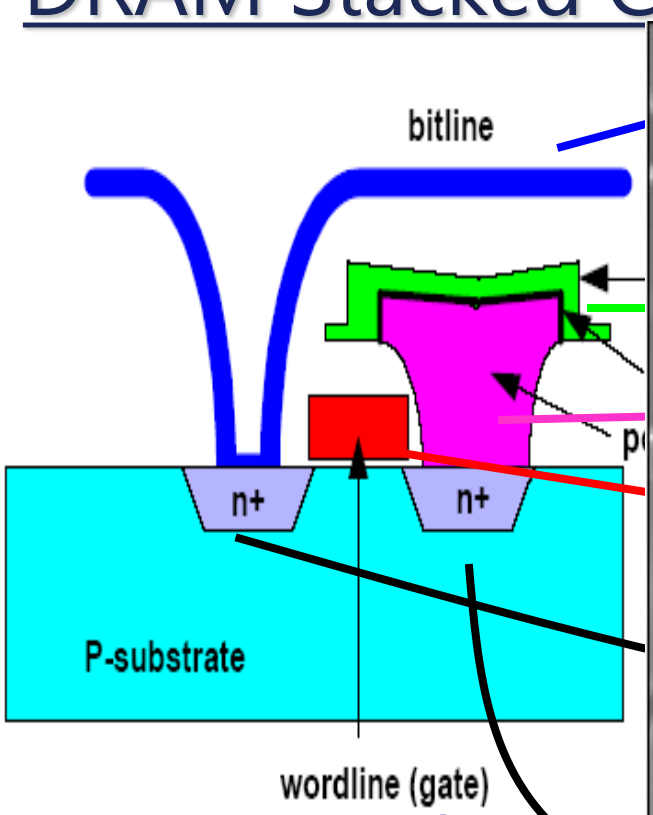
DRAM Trench Cell

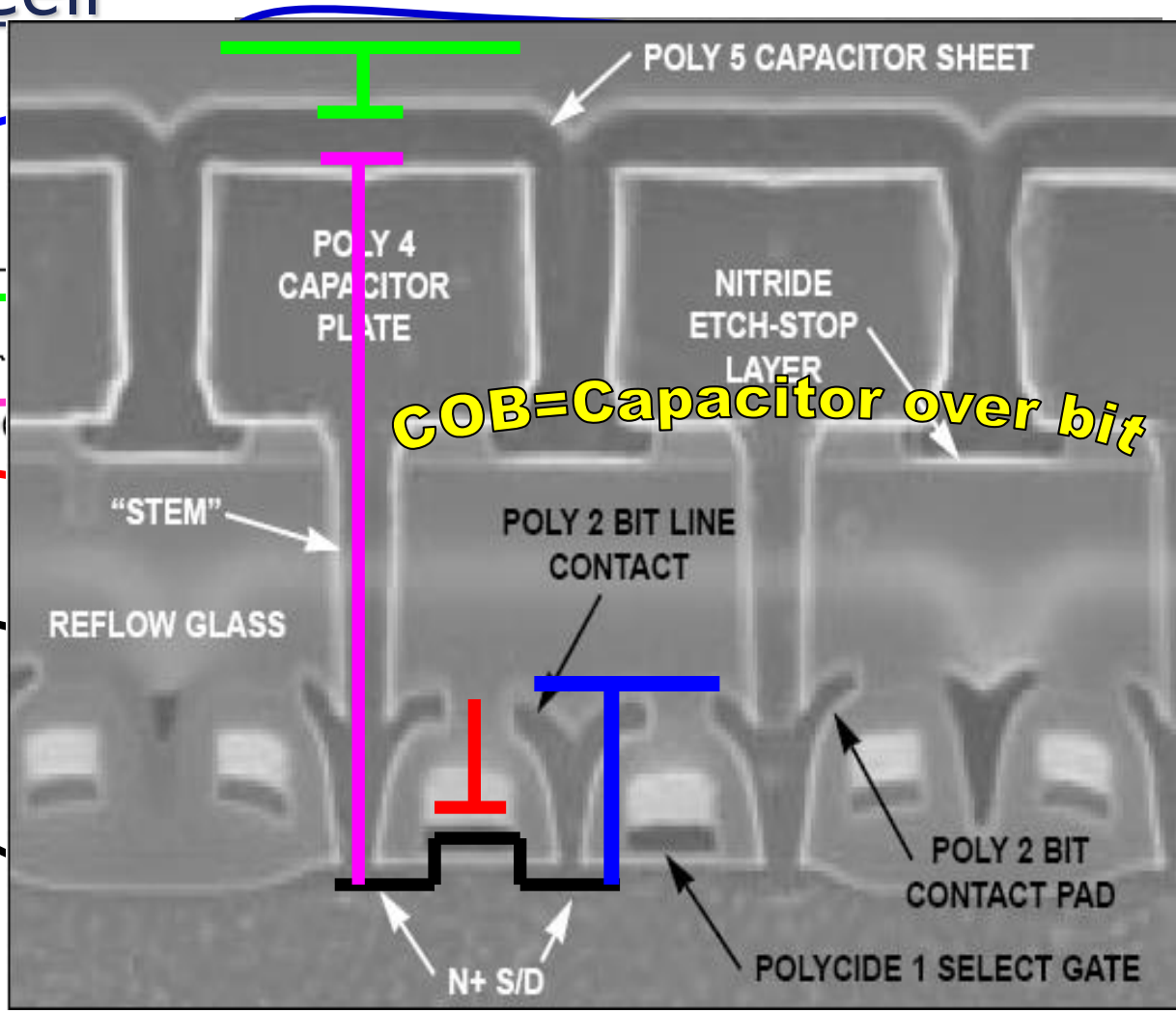




- Process
  - Etch deep hole in substrate
    - Becomes reference plate
  - Grow oxide on walls
    - Dielectric
  - Fill with polysilicon plug
    - Tied to storage node

DRAM Stacked Cell

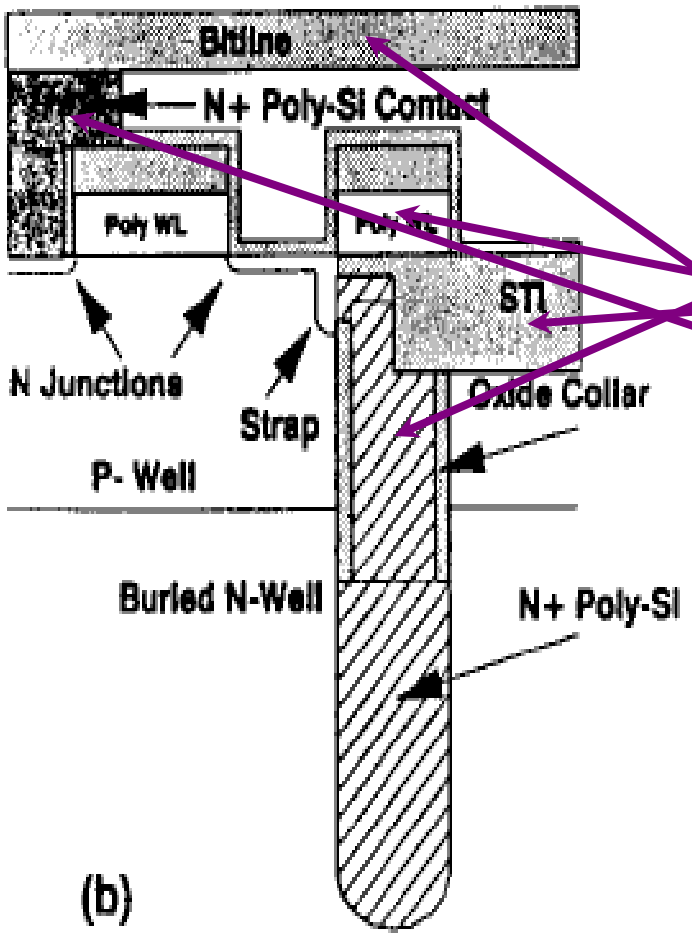




Samsung 64Mbit DRAM Cross Section

- Storage Capacitor
  - Rubidium electrodes
  - High dielectric insulator
    - 50X higher than SiO_2
    - 25 nm thick
  - Cell capacitance 25 femtofarads

DRAM Buried Strap Trench Cell



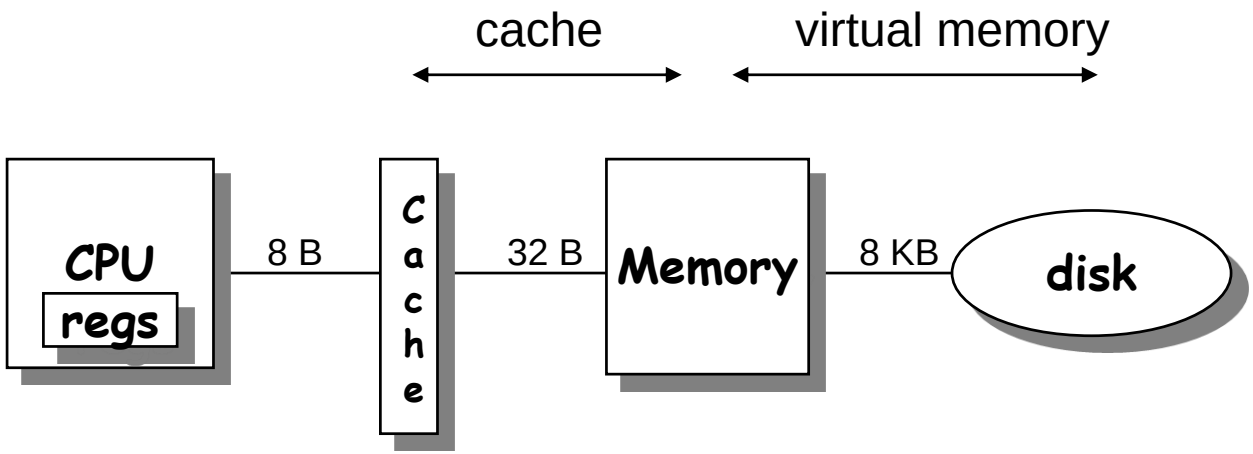
Process Module	Feature	Material Planarized
Deep Trench Capacitor	35 fF/cell NO Node Diel. 6nm TEQ.	PolySi
Shallow Trech Isolation	0.55 μm pitch	Oxide
Gate	n+ poly/WSi 0.55 μm pitch	BPSG
Contacts	poly-Si to array W to supports	poly-Si, Tungsten
Bitline	W Damascene 0.55 μm pitch	Tungsten
Via 1	W filled	Tungsten
Metal 1	Aluminum 0.55 μm pitch	None
Metal 2	Aluminum 2.2 μm pitch	None

DRAM Scaling

	1992	1995	1998	2001	2004	2007
Feature size:	0.5	0.35	0.25	0.18	0.12	0.10
– *Industry is slightly ahead of projection*						
DRAM capacity:	16M	64M	256M	1G	4G	16G
– *Doubles every 1.5 years*						
– *Prediction on track*						
Chip area (cm^2):	2.5	4.0	6.0	8.0	10.0	12.5
– *Chips staying small*						

- Scaling challenge
- Must keep C_{node} high as we shrink cell size

Computing Memory Hierarchy

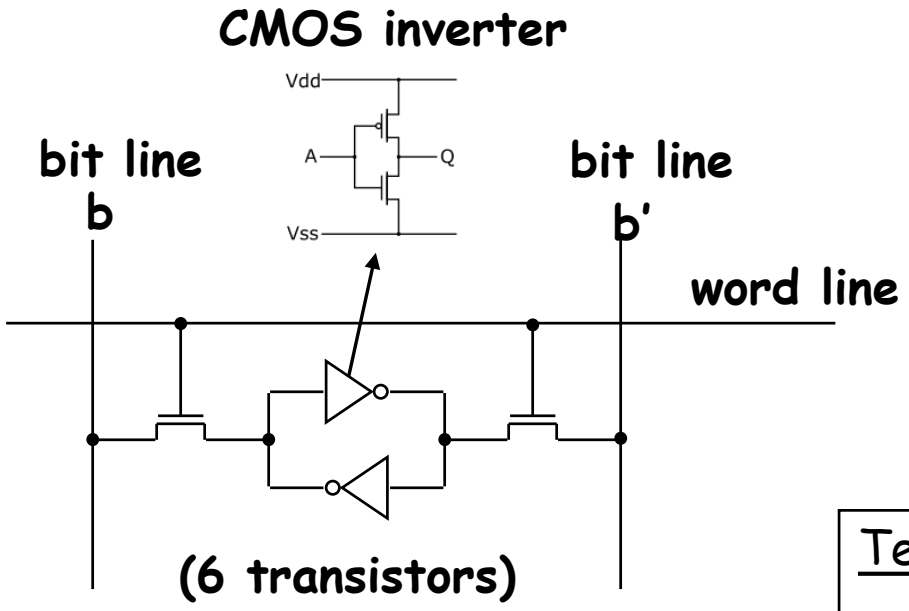


	Register	Cache	Memory	Disk Memory
size:	200 B	32KB - 4MB	128 MB	20 GB
speed:	2 ns	4 ns	60 ns	8 ms
$/Mbyte:		$100/MB	$1.50/MB	$0.05/MB
block size:	8 B	32 B	8 KB	
		?	**DRAM**	**HDD/Flash**

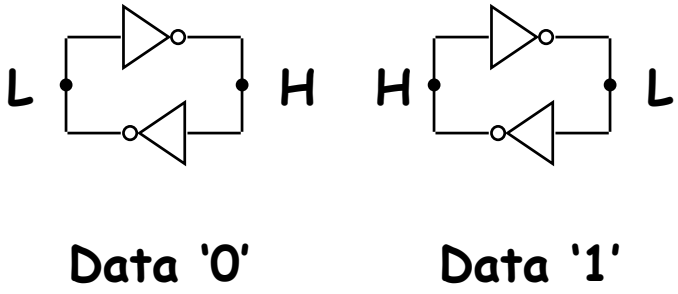
Static RAM (SRAM)

- High speed ~<5 ns
- Stable
  - Highly immune to noise/disturbance
  - NO refresh needed
- Large
  - 6 transistor per bit
- Expensive
  - ~$100/MB

SRAM Working Principle



Stable Configurations



Terminology:

bit line: carries data

word line: used for addressing

Write:

1. set bit lines to new data value
   - **b'** is set to the opposite of **b**
2. raise word line to "high"

⇒ sets cell to new state (may involve flipping relative to old state)

Read:

1. set bit lines high
2. set word line high
3. see which bit line goes low

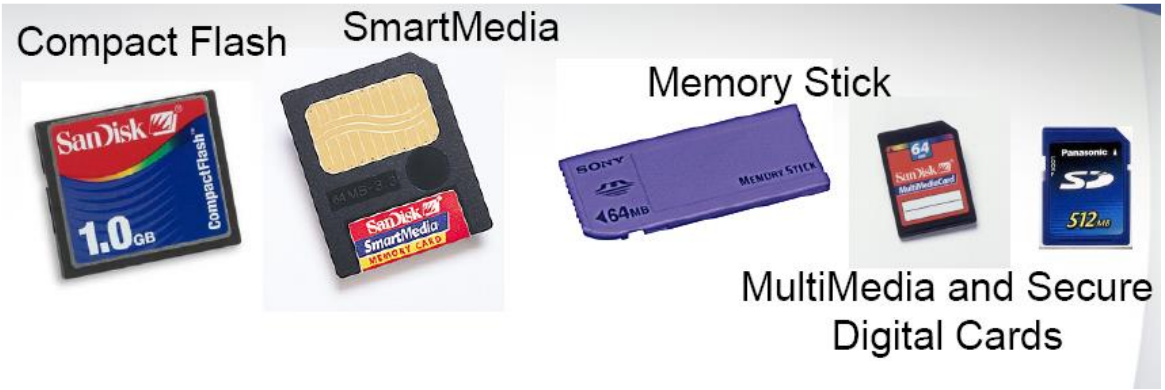
DRAM vs. SRAM

- DRAM (Dynamic RAM)
- Used mostly in main mem.
- Capacitor + 1 transistor/bit
- Need refresh every ~50 ms
  - 5% of total time
- Read is destructive (need for write-back)
- Requires sensing amplifier
- Access time ~20 ns
- Density (25-50):1 to SRAM

- SRAM (Static RAM)
- Used mostly in caches (I, D, TLB, BTB)
- 1 flip-flop (4-6 transistors) per bit
- Draws power even during standby
- Read is not destructive
- Bitline is driven to high/low
- Access time ~ 2 ns
- Speed (8-16):1 to DRAM

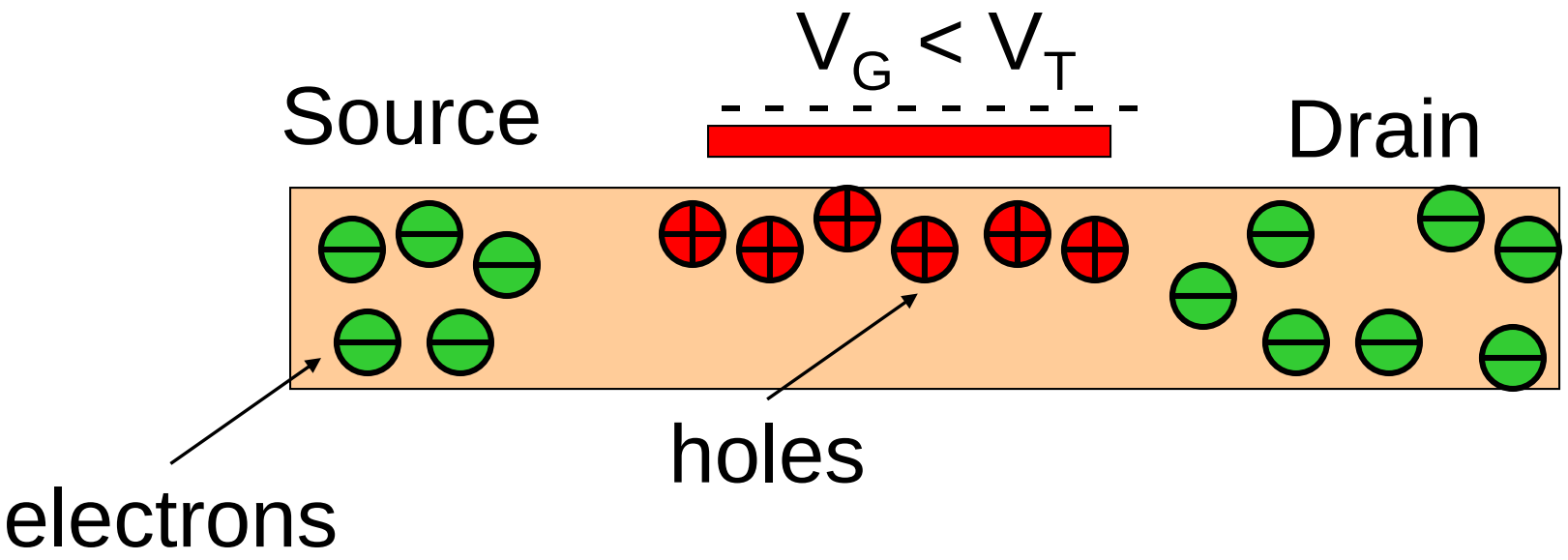
Flash Memory

- Moderate cost per bit
- Good density
- Slow speed: ~25 µs (read time) to ms (write time)
- Random access
- Non-volatile
- Low power consumption
- Poor endurance
- Erase before write

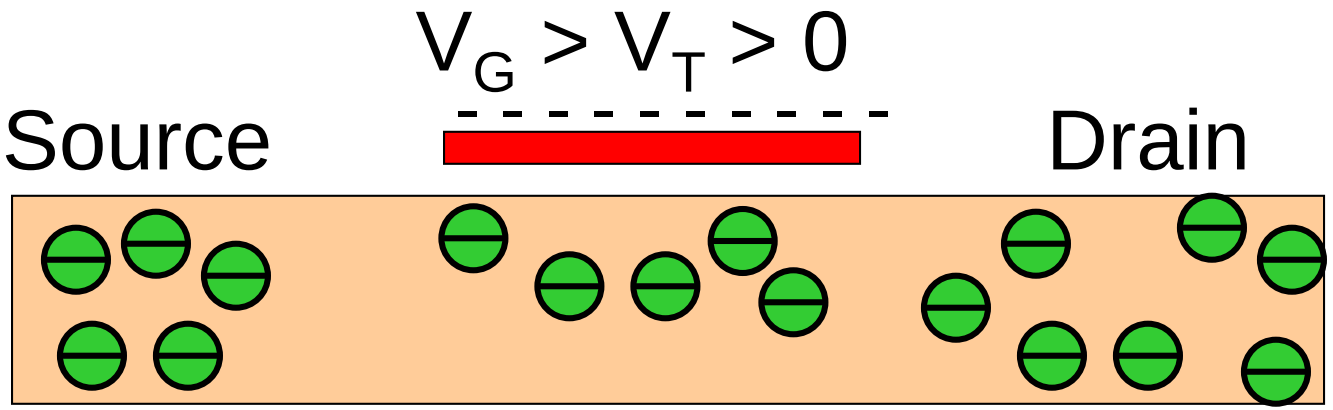


MOSFET

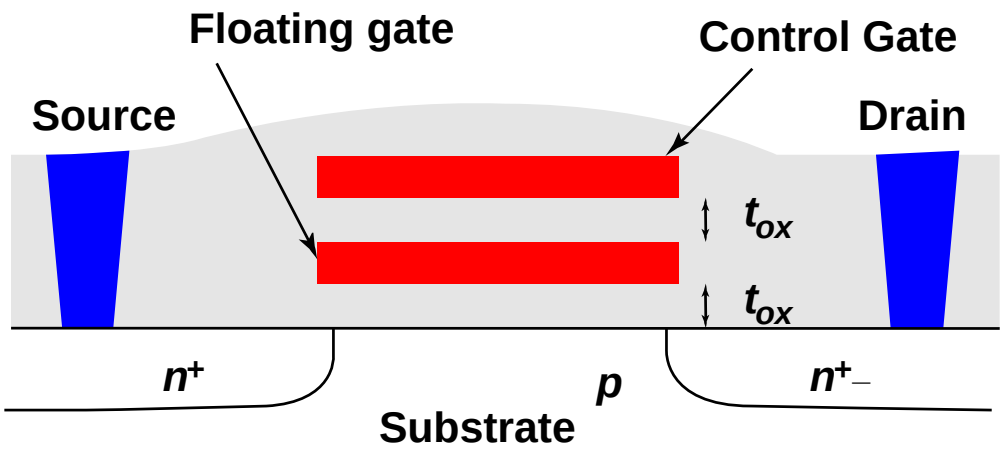
OFF



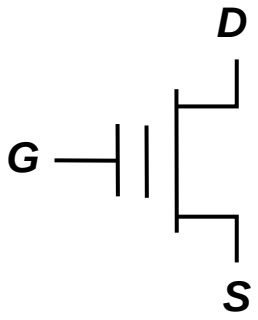
ON

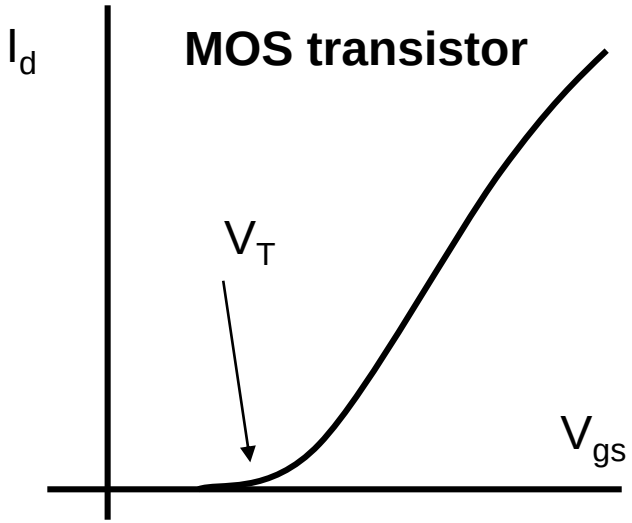


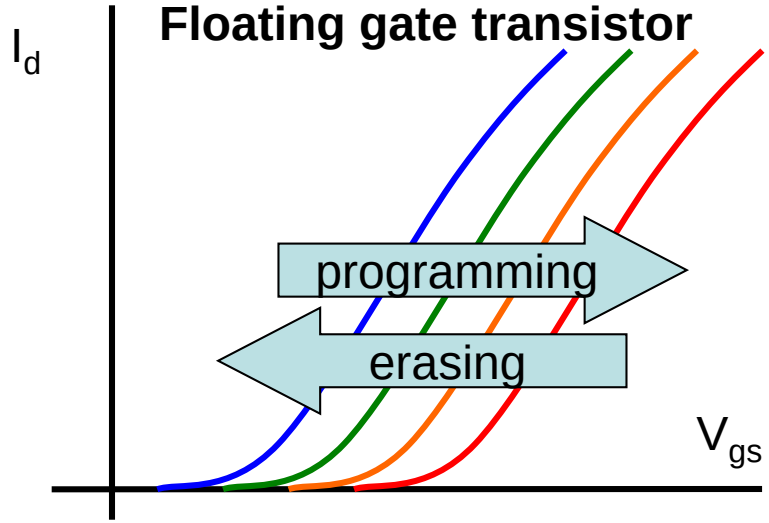
Floating Gate Transistor



Circuit symbol

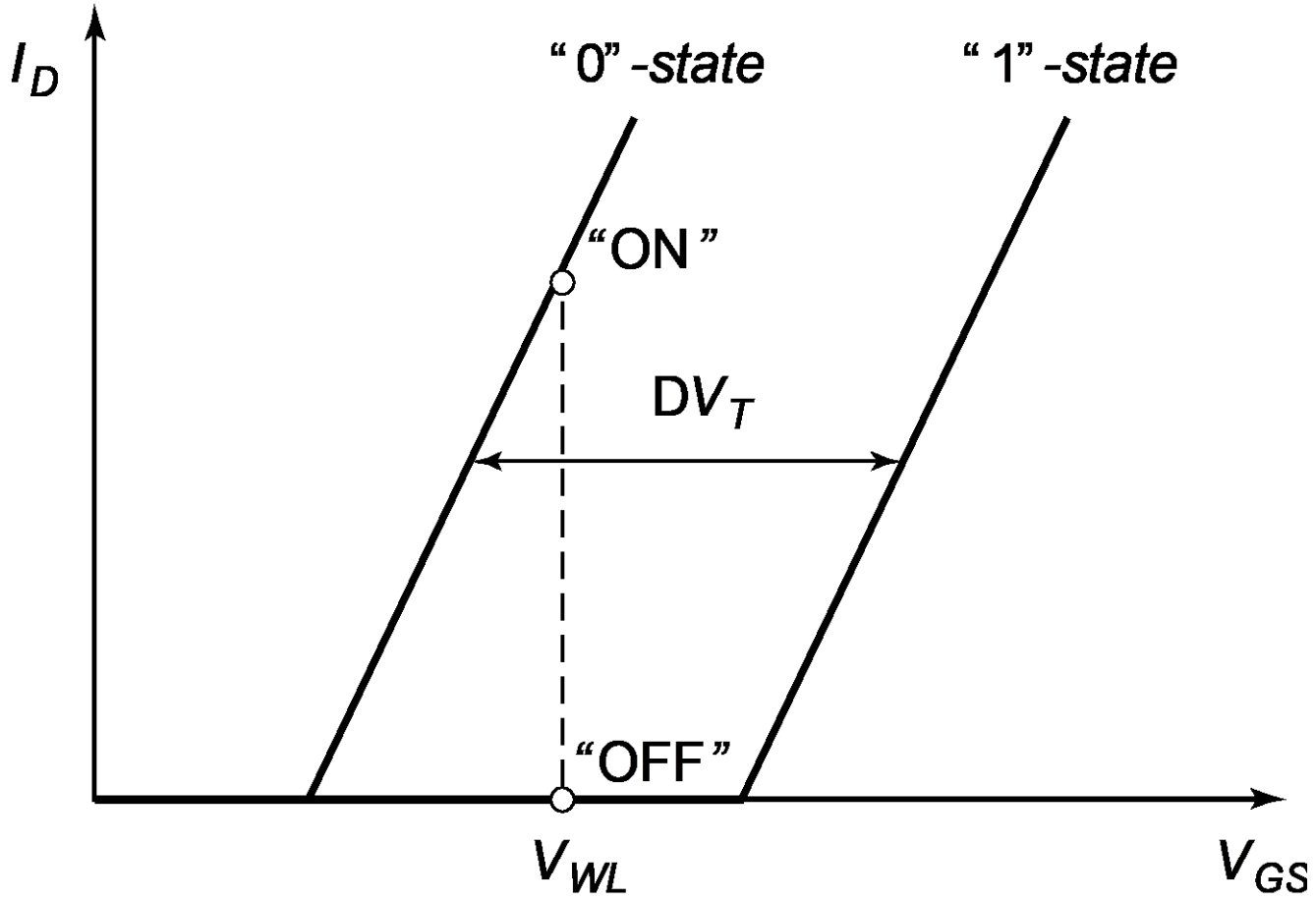






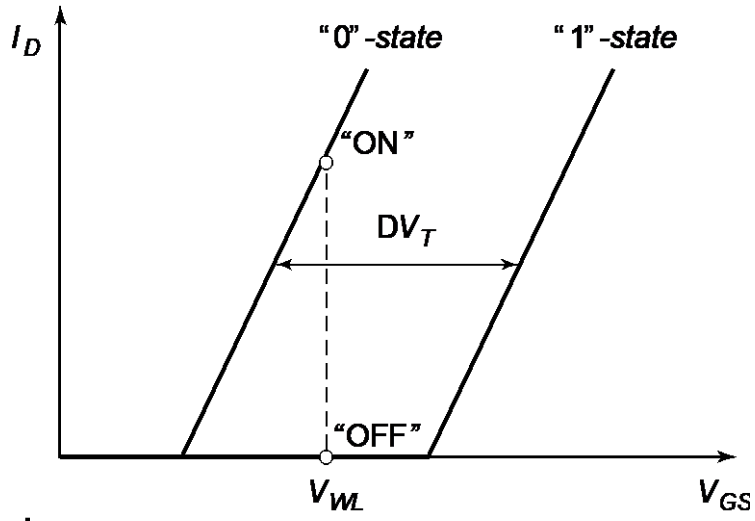
- MOSFET: fixed threshold voltage V_T
- Floating gate: V_T tuned by program/erase

Programmable Threshold



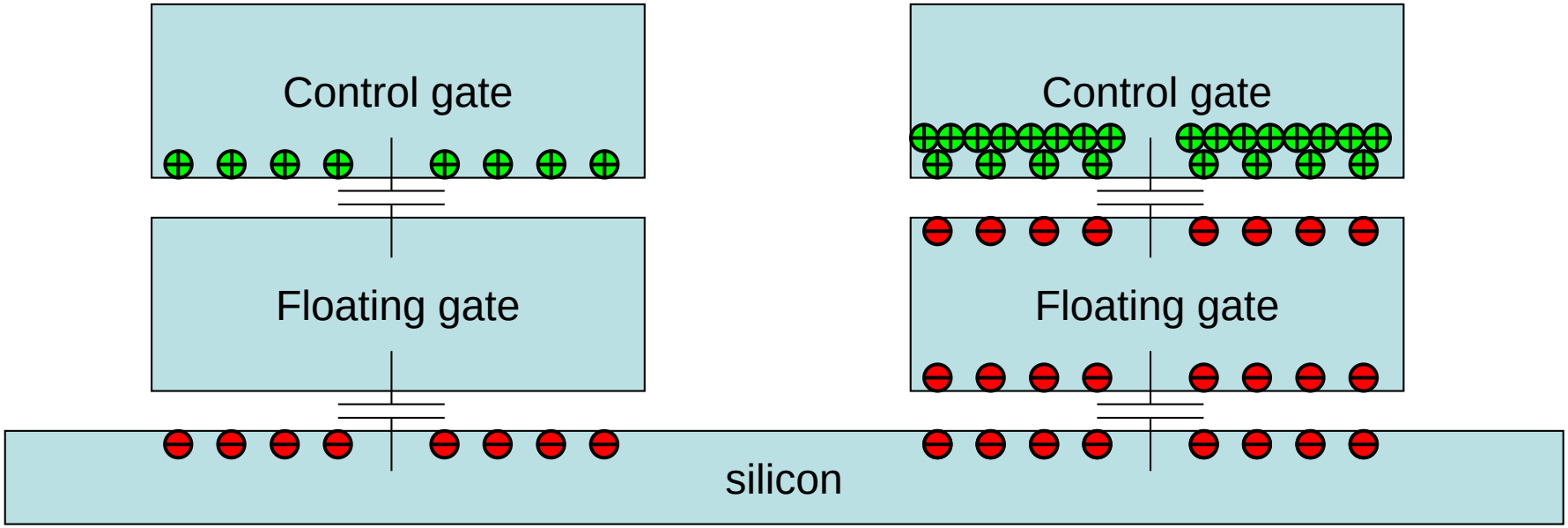
- Read
  - Apply V_{WL} on control gate, measure I_D
  - "0" → $I_D >> 0$, "1" → $I_D = 0$

Flash Memory Working Principles

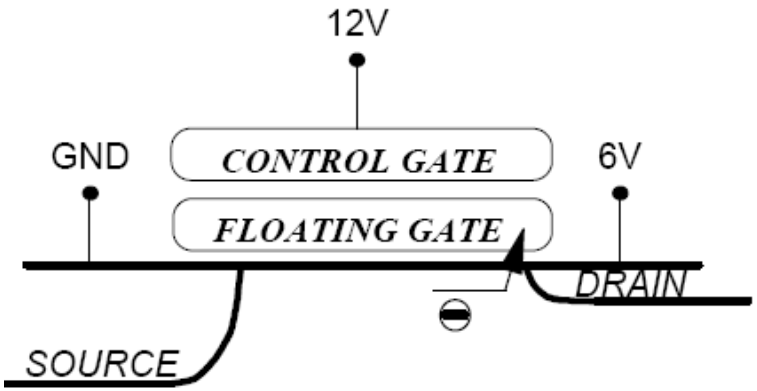


erased

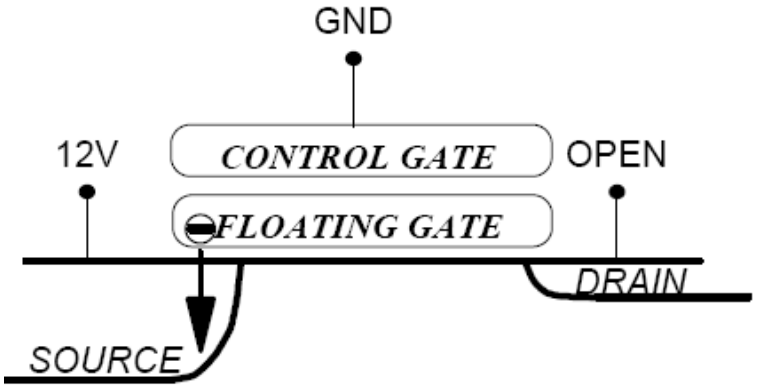
programmed



Program and Erase Method



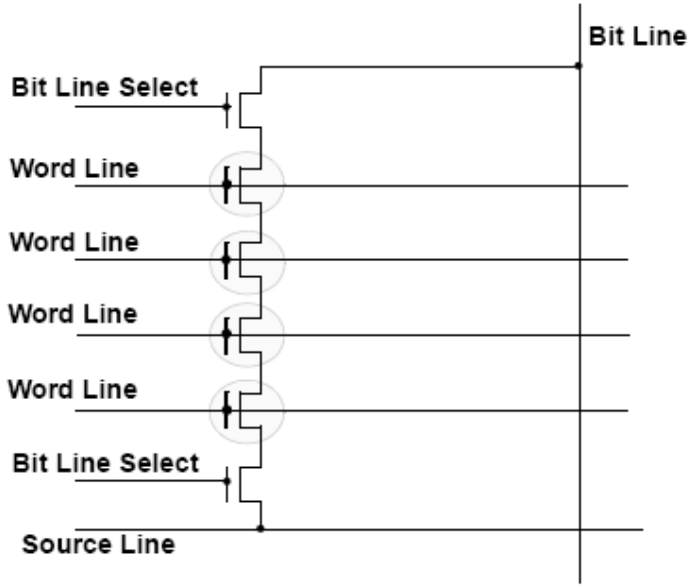
- Programming
- channel hot electron (CHE)
- Injection to floating gate at drain side



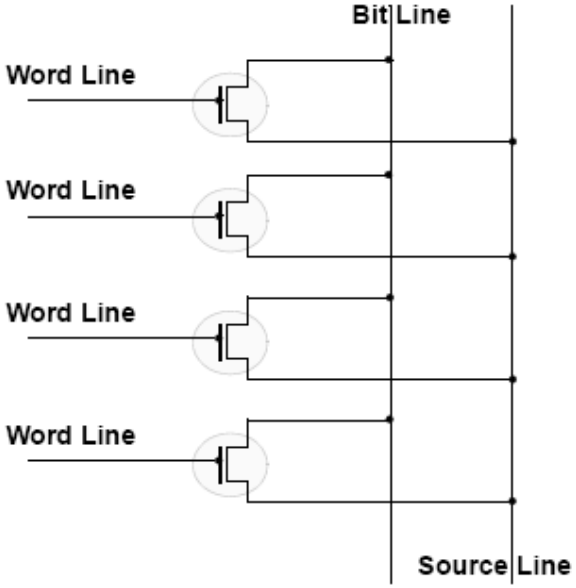
- Erasing
- Fowler-Nordheim tunneling
- Through oxide at source side

NAND vs NOR Flash

NAND



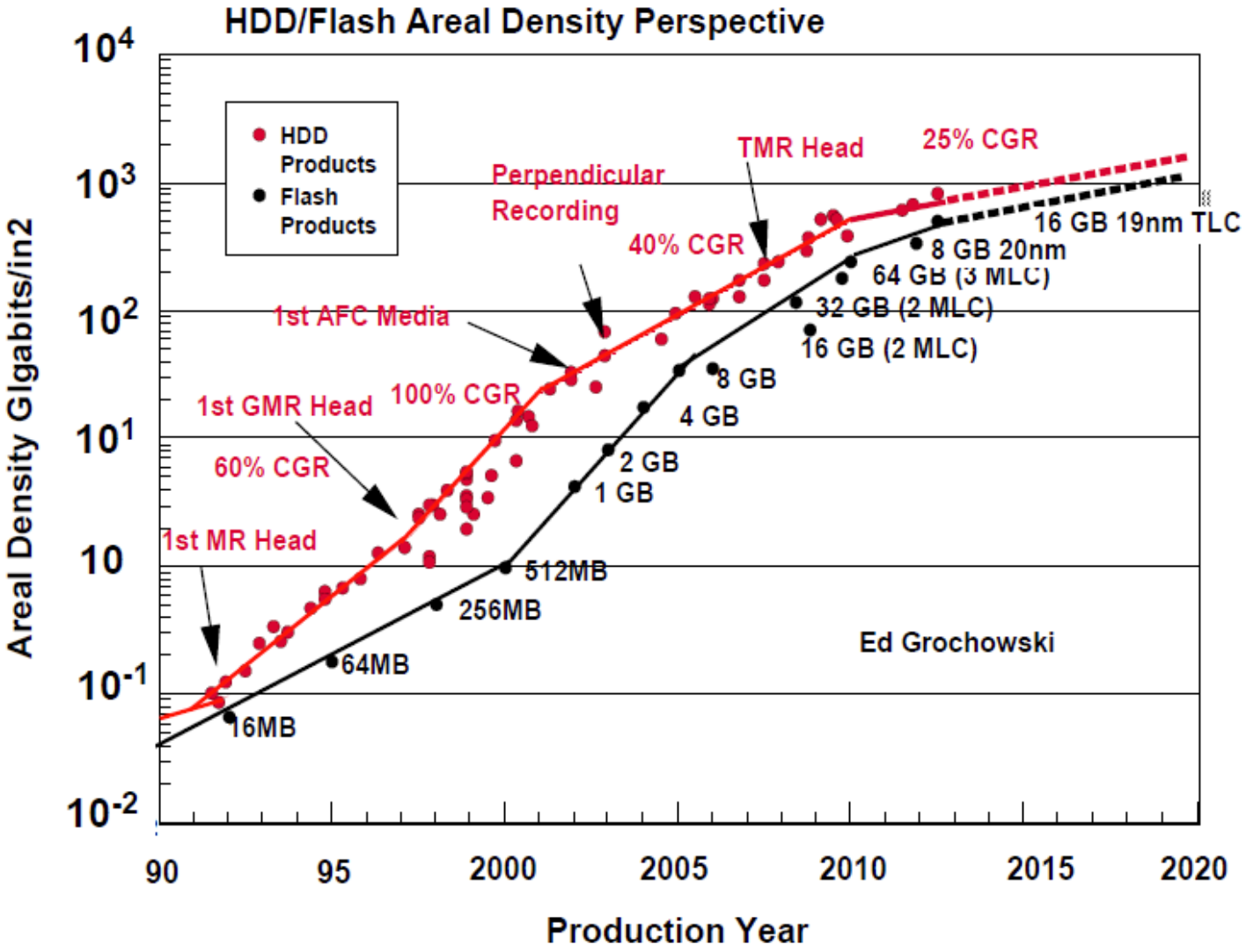
NOR



- Small cell, sharing contact
- Easy to scale down
- Sequential access 1 μs
- Fast write/erase 1 μs
- Lower cost

- Large cell
- Difficult to scale
- Fast random access 100 ns
- Slow write/erase 10 μs
- More expensive

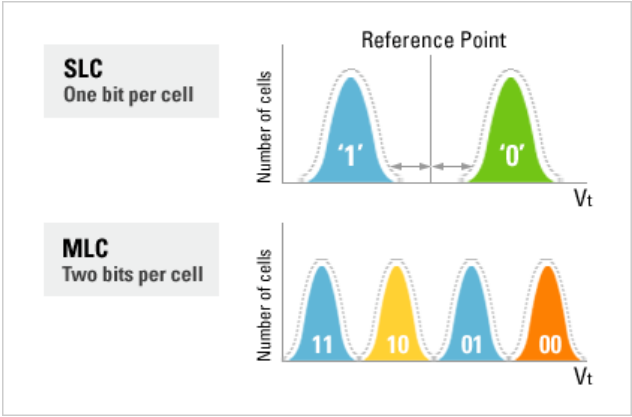
Flash Scaling



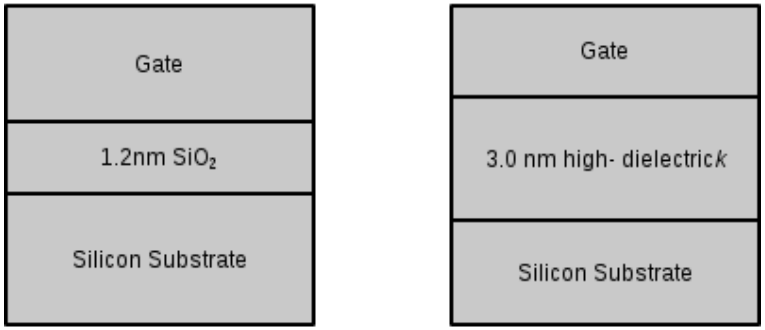
E. Grochowski et al., “Future technology challenges for NAND flash and HDD products”, Flash Memory Summit 2012

Flash Scaling Tricks

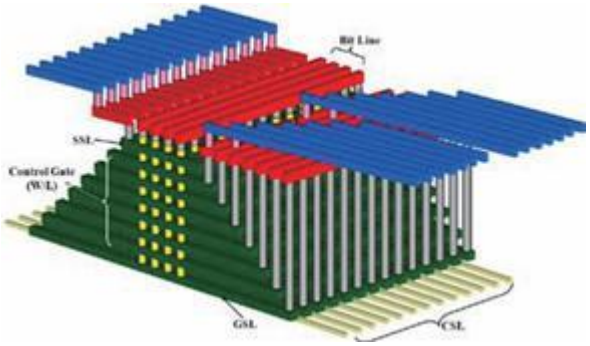
- Multi-level cell



- High-k layer



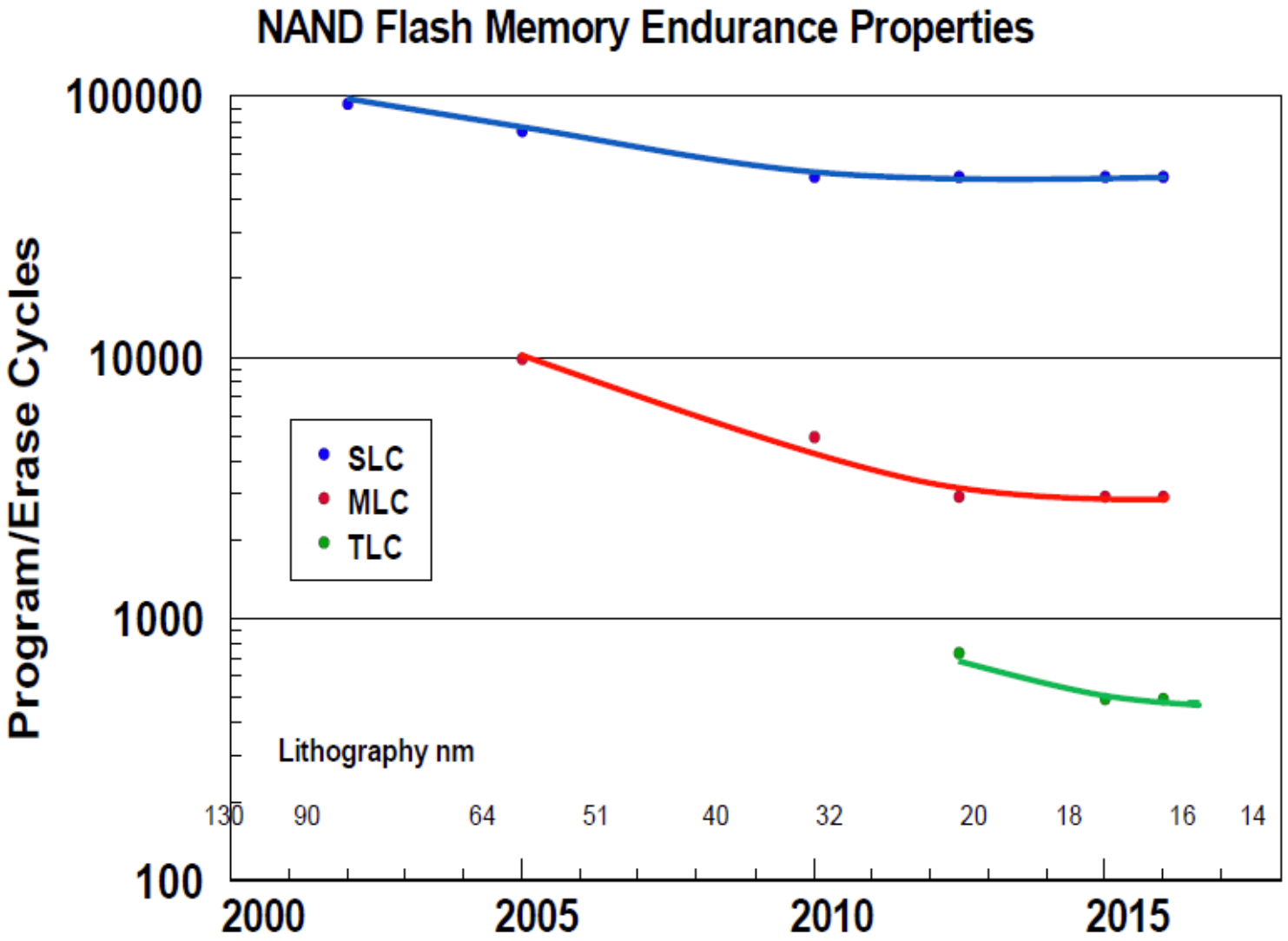
- 3D flash



Flash Memory Endurance and Retention Issues

- Endurance
  - device failure from hot carrier injection
  - oxide breakdown under stress (high voltage)
  - oxide charge trapping
  - Shift in operating window
  - even worse for multi-level cell (MLC)
  - ~1000× now
- Retention
  - Requiring no charge leak for ~10^8 s,
  - i.e. cannot lose >2 electrons per day
  - Oxide defects, mobile ions, contamination
  - High temperature/radiation pose thread
  - More difficult with scaling
  - Less tolerant for multi-level cell (MLC)

Flash Endurance



Optical Data Storage

- Cave painting

- Drawing

- Photography

Compact Disk Read Only Memory (CDROM)

- Invented by James Russell in 1970
- Mass production since 1985 by Philips and Sony

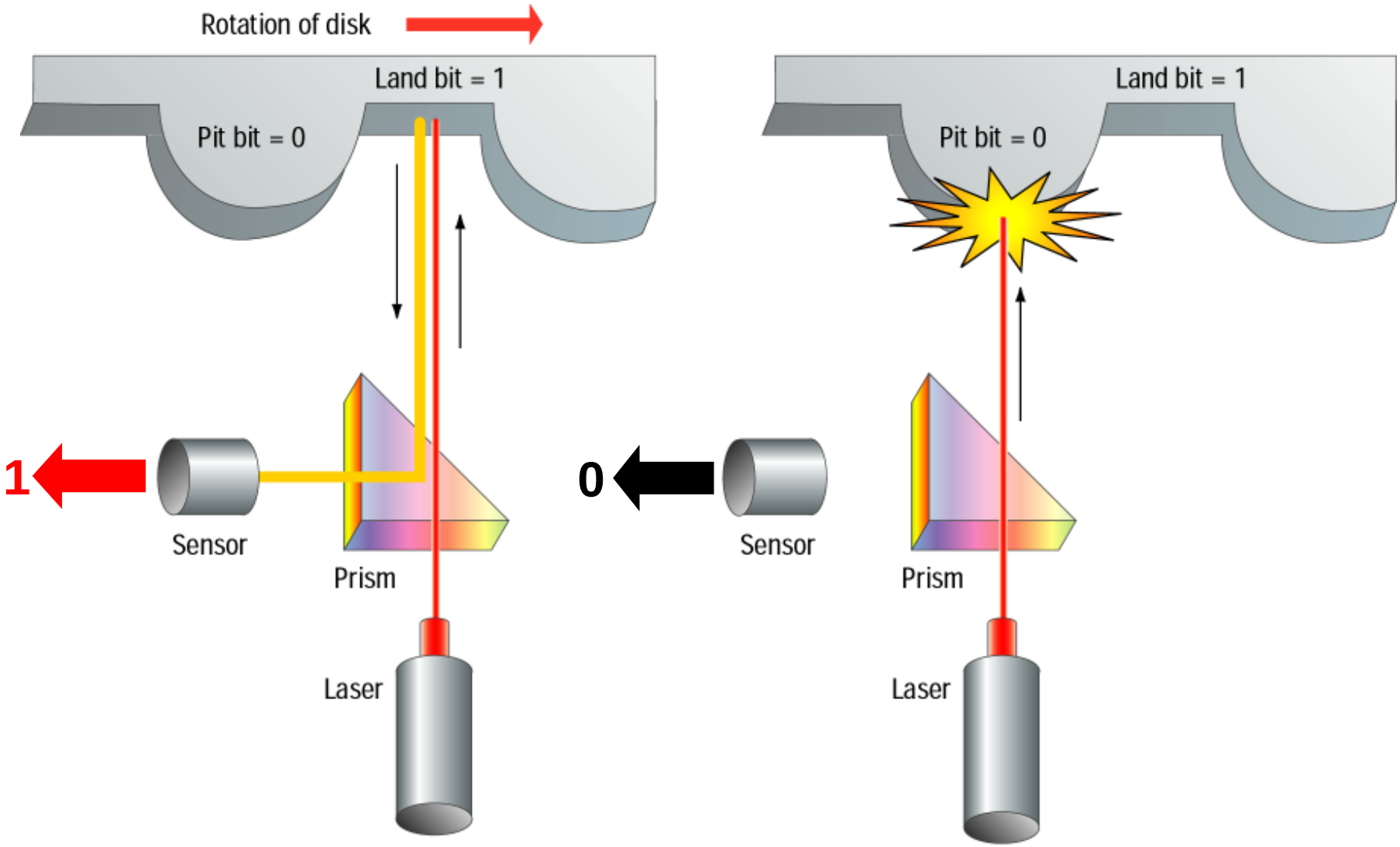
- Basis
  - Optical recording technology developed for audio CDs
    - 74 minutes playing time

- Bit Rate
  - 150 KB / second

- Capacity
  - 74 Minutes * 150 KB / second * 60 seconds / minute = 650 MB

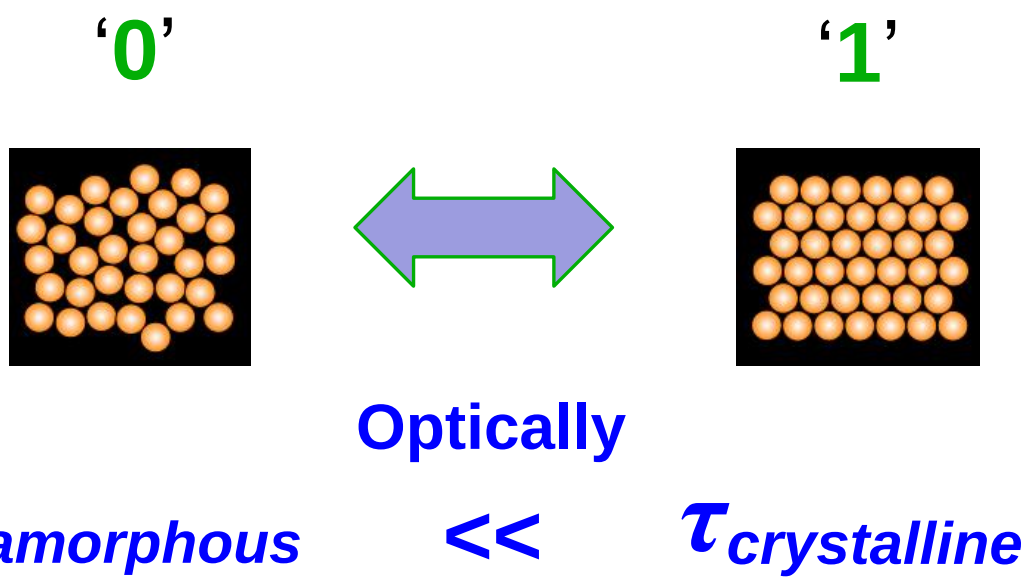
- Read only, cannot be overwritten

CDROM Working Principle

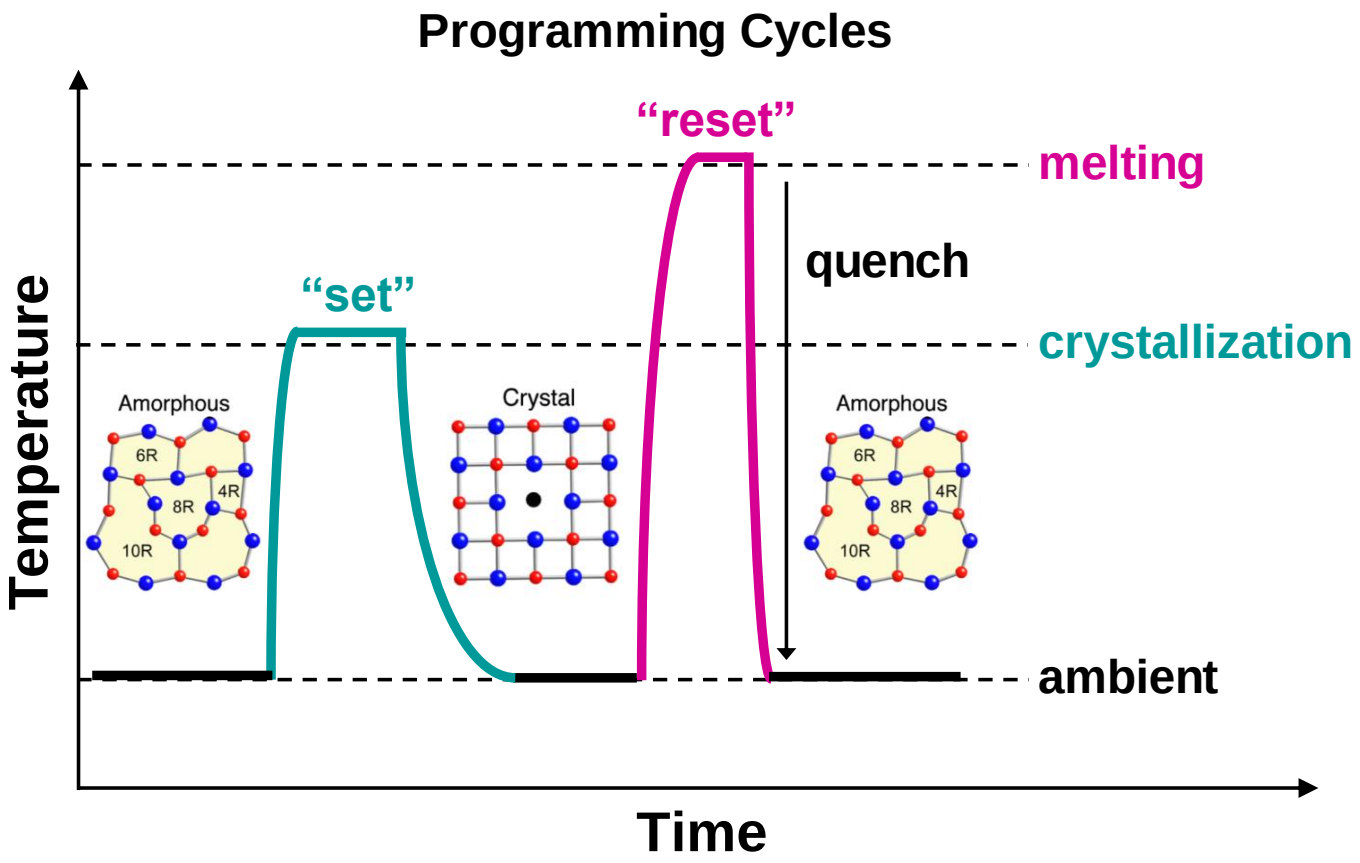


CD-Rewritable (CD-RW)

- Allows writing new data over recorded data
- Endurance: 100-1000 times
- Based on phase change materials



Programming

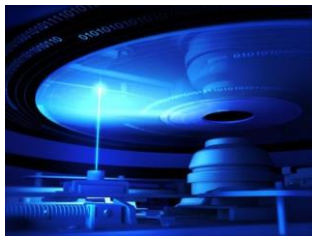


- "set": crystallization → data rate limiting (~10 ns)
- "reset": melt-quench → power limiting (~600 C)

DVD

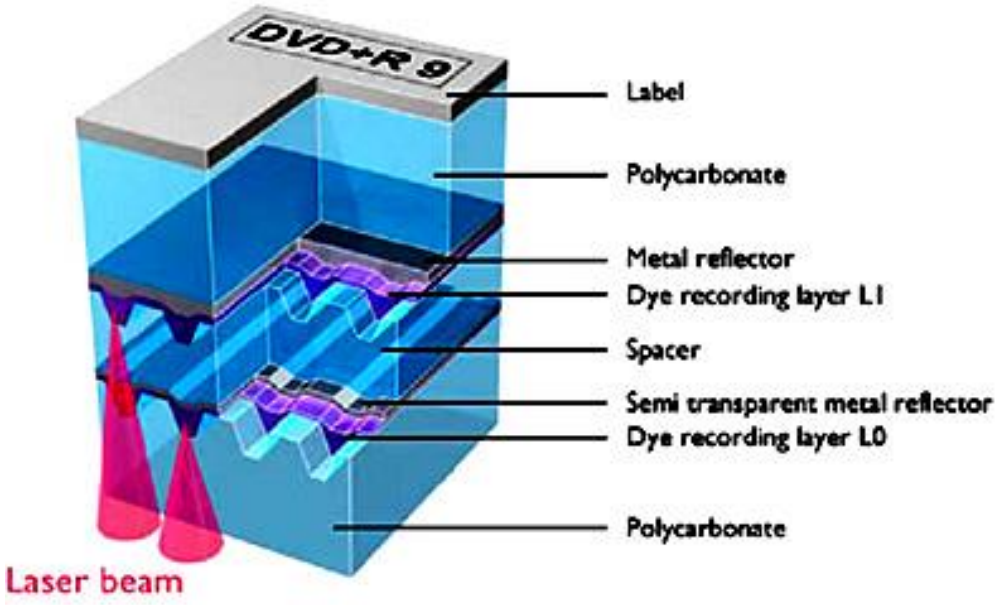
- Improved technology upon CD-RW
- Smaller wavelength → higher density
- Better mechanical control
- Improved error correction

- Larger capacity
  - Standard – Up to 4.7 GB, 7 times more than CD-ROM
  - Double layers – 8.5 GB
  - Double-sided – 17 GB
  - Blu-ray (BD) disk – 25 GB
  - Dual layer BD – 50 GB



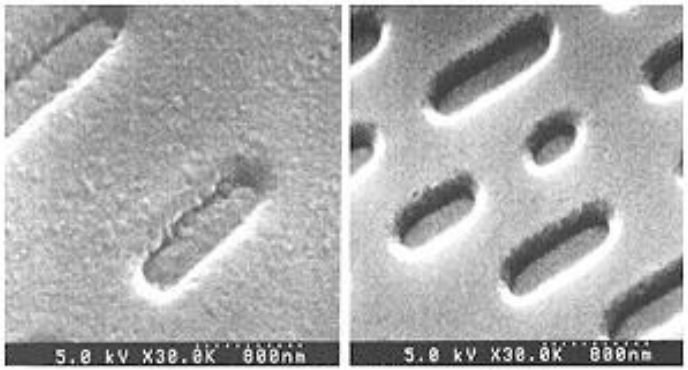
Dual Layer Technology

- Benefits
  - Increased durability
  - Increased capacity
- Detriments
  - Decreased S/N
  - Decreased data density



Numerical Aperture

- NA = n sin(θ/2)
- Spot size = λ/NA
- CD-RW λ ~ 780 nm IR
- DVD λ ~ 650 nm red
- Blu-ray λ ~ 405 nm blue



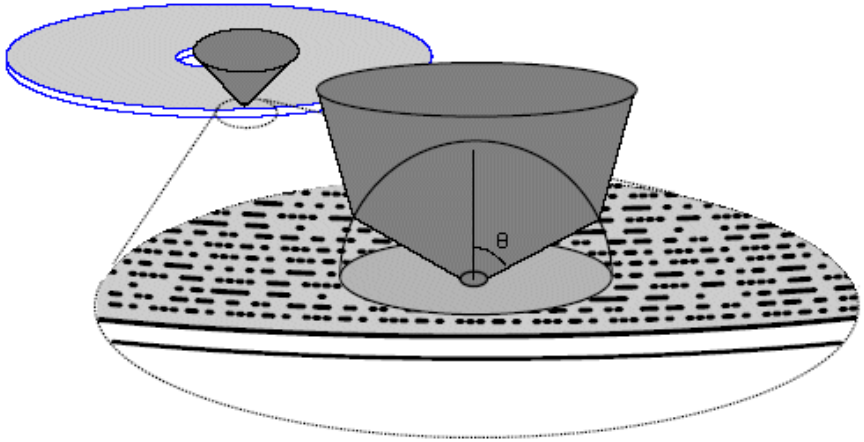
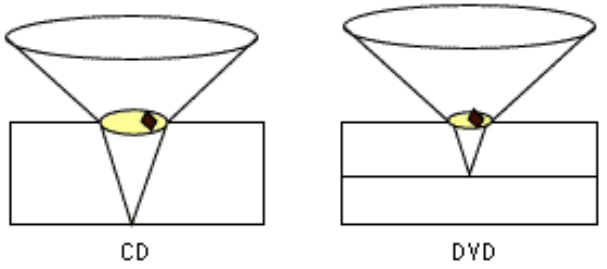
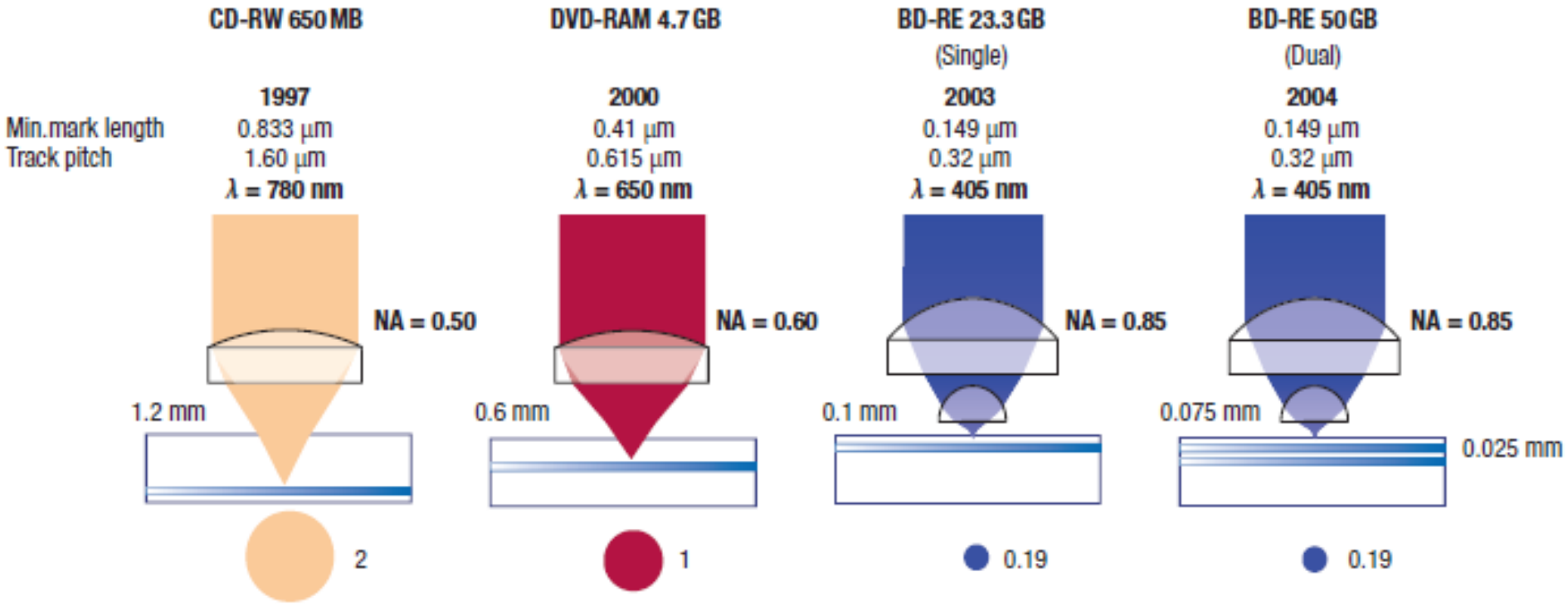


Figure 2: A solid–immersion lens (SIL) can increase the effective NA beyond 1.0, further increasing density but requiring evanescent coupling between the SIL and disk.



Comparison of Optical Storage



Depth of Focus (DoF)

- DoF = λ/NA^2
- Determines spacing of layers
- Decreasing depth of focus → more layer → higher density
- Affects S/N; places an upper limit on NA.

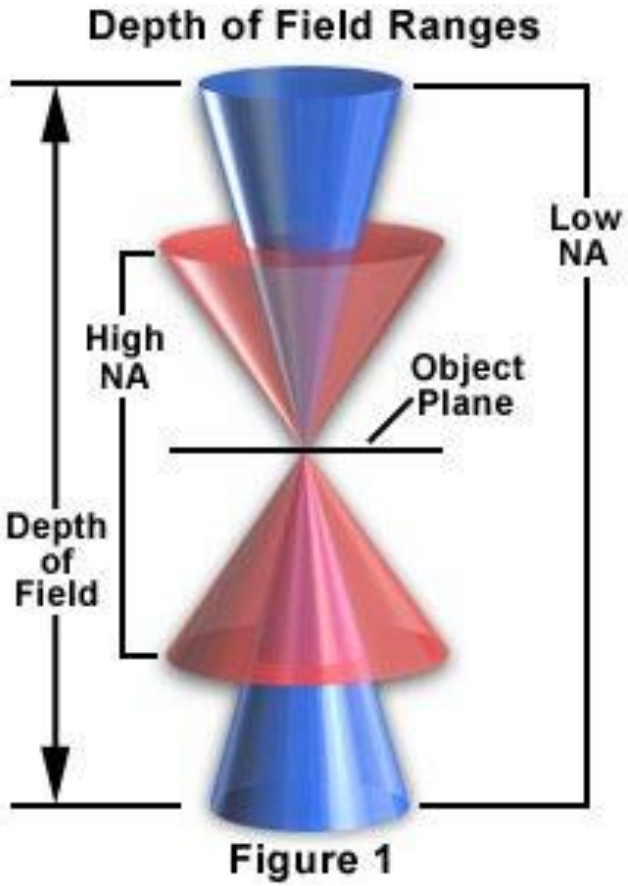
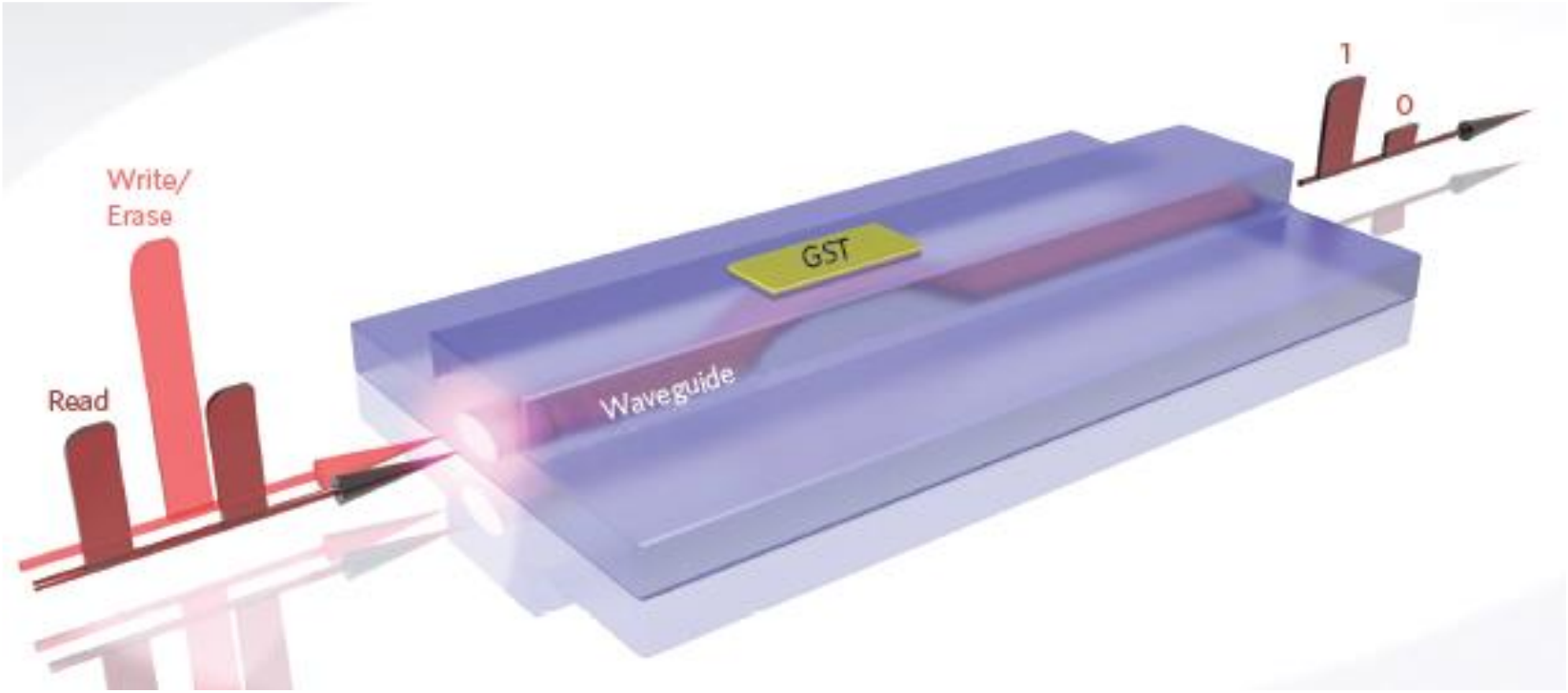


Figure 1

Photonic Memory



- Change in Absorption in photonic waveguide
- Optical memory and all photonic circuit