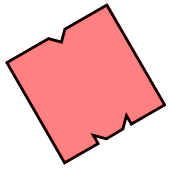
CS-590.26, Spring 2014

High Speed Memory Systems: Architecture and Performance Analysis

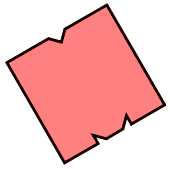
Alternative Solid State Memories: Flash, MRAM, FeRAM, PCRAM

Credit where credit is due:
Slides contain original artwork (© Jacob, Wang 2005)



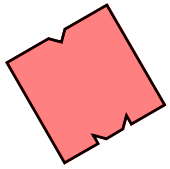
What is DRAM?

- **(Relatively) fast reads and (relatively) fast writes**
- **Unlimited number of writes**
- **Volatile - loses data storage without power**
- **Dynamic - loses data without periodic refresh**
- **Could be fabricated using similar materials and (relatively) similar silicon based process technologies as leading edge processors**

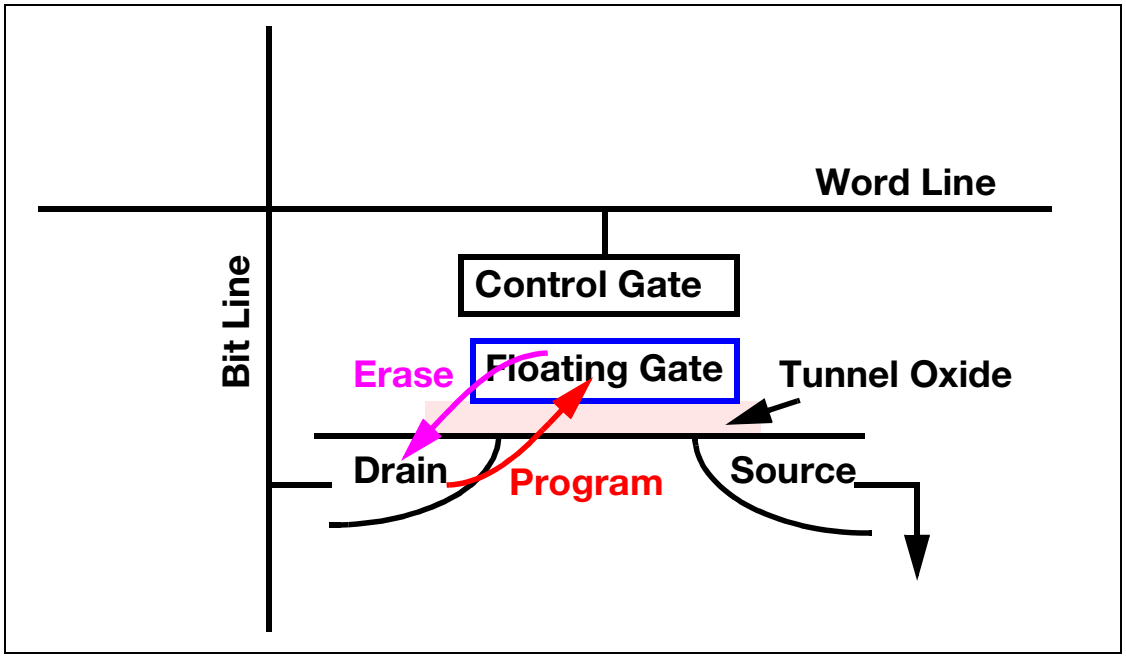


Alternatives:

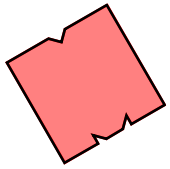
- **(Relatively) slower reads and (on some) really slow writes**
- **(Some) limited number of writes**
- **Non-Volatile - keeps data storage without power**
- **May require new materials and (relatively) different process technologies as leading edge processors**



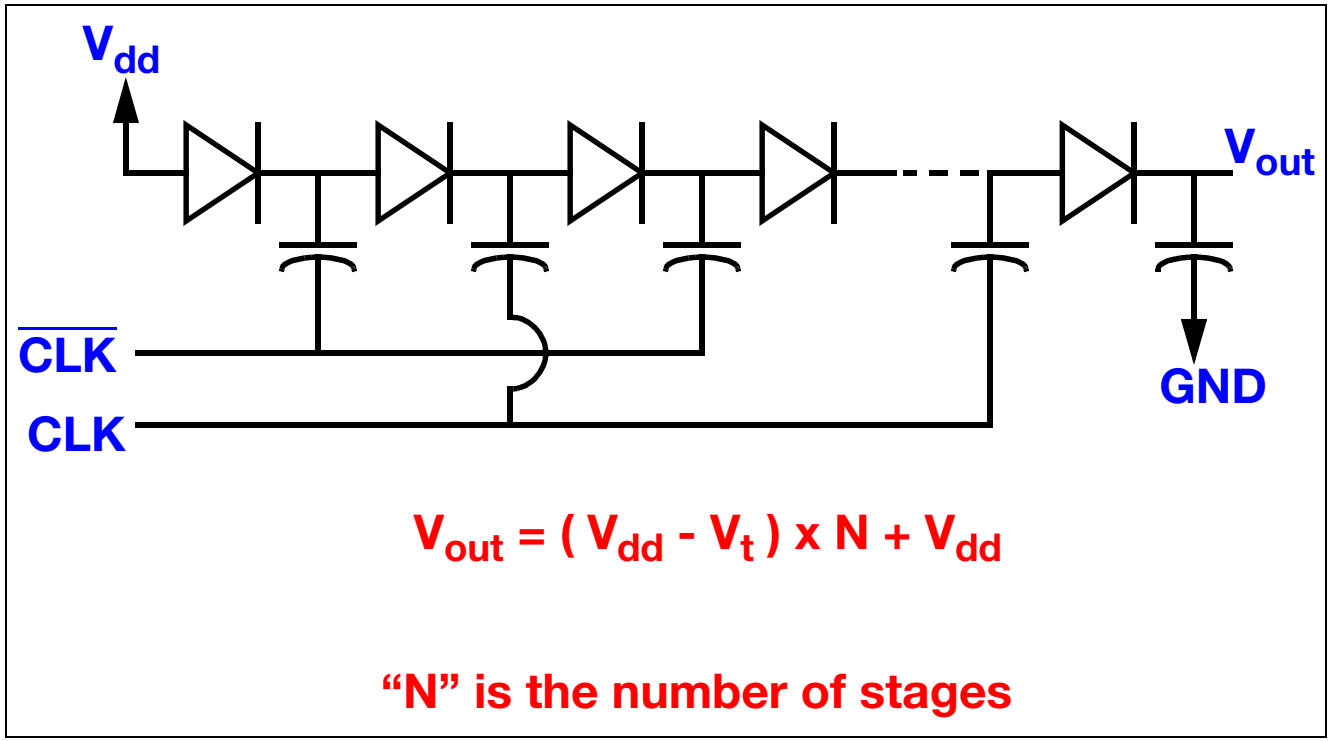
Flash: Basic Idea



- Electrical charges are forced to tunnel through oxides and get trapped in the floating gate.
- High voltage forces tunneling
- Trapped charges in floating gate then alters V_t
- Differences in V_t of transistor then sensed as 0/1
- Explicit program and erase cycles

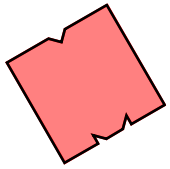


Charge Pump

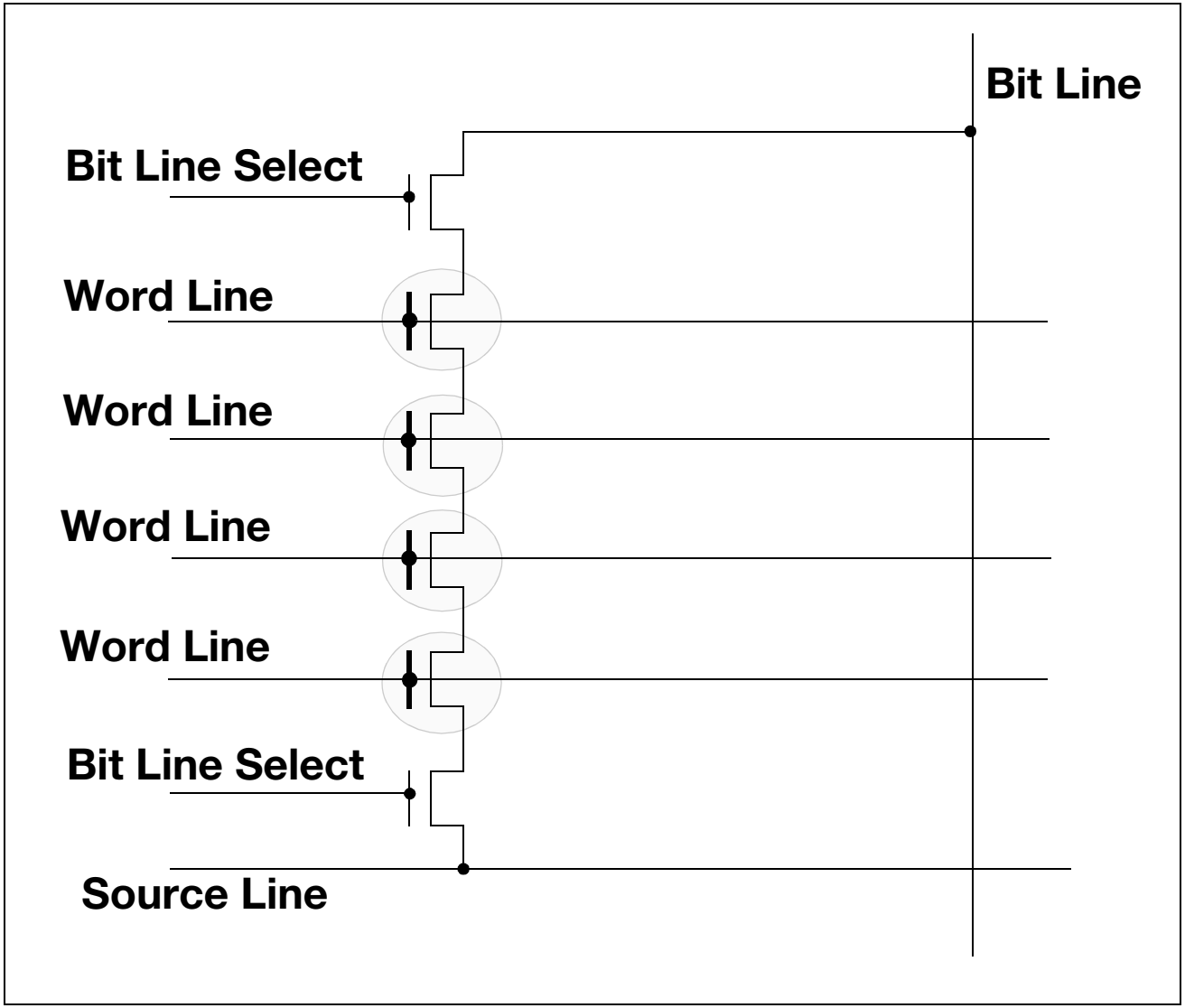


Idealized charge pumping circuit

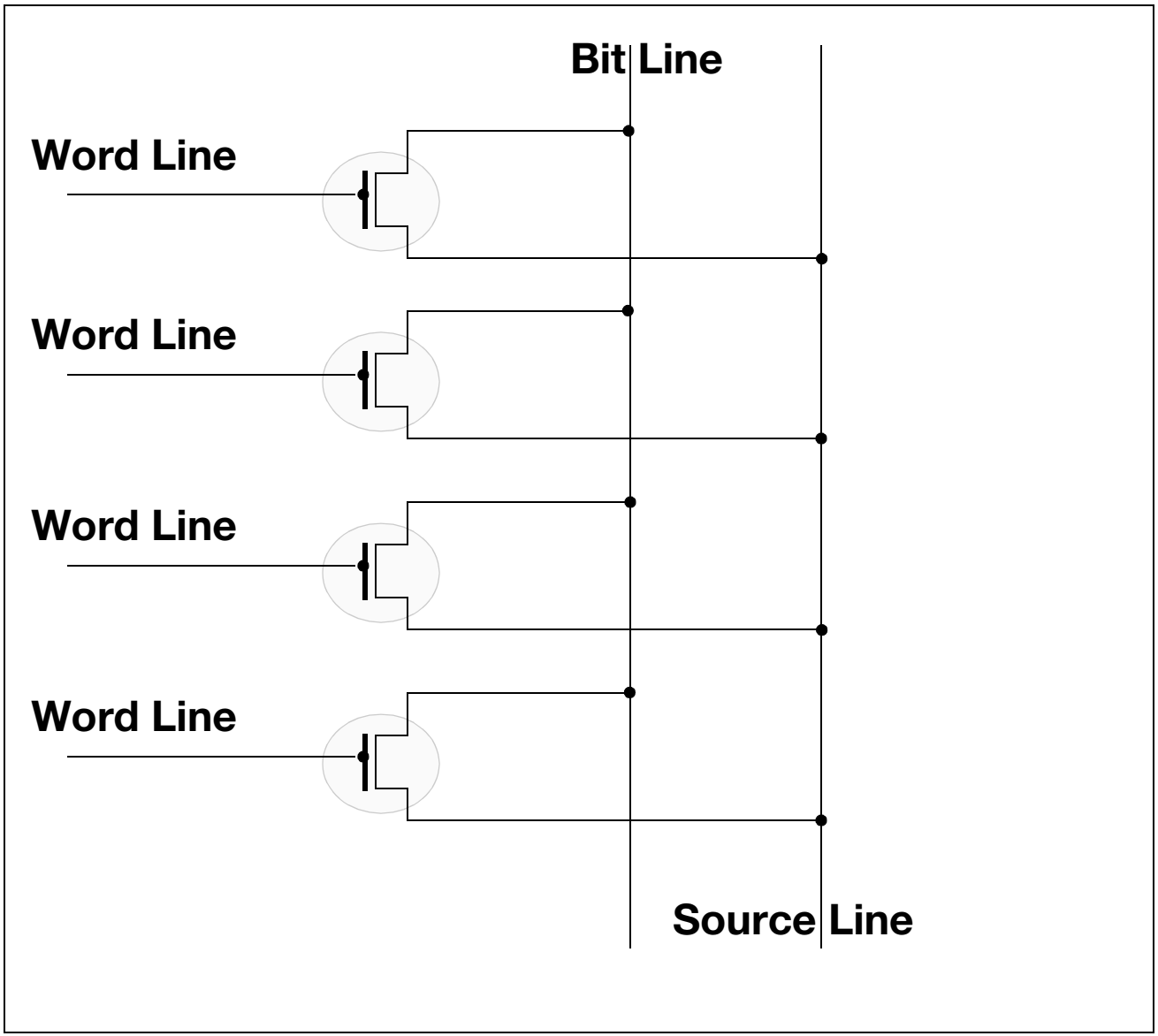
- Build up larger voltage for programming floating gate



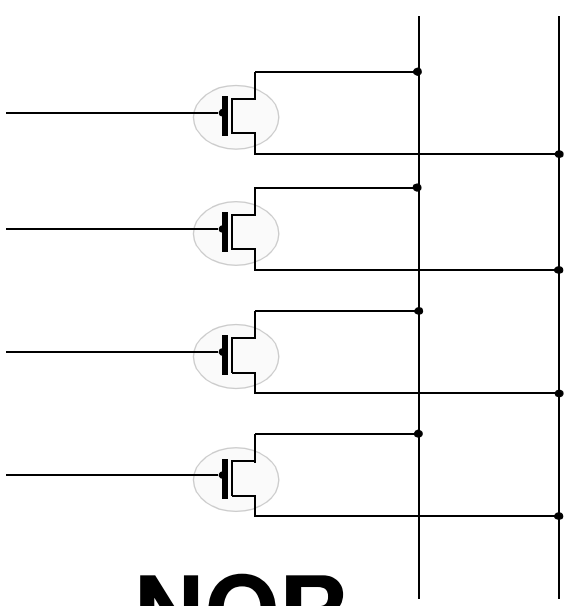
NAND Flash Array



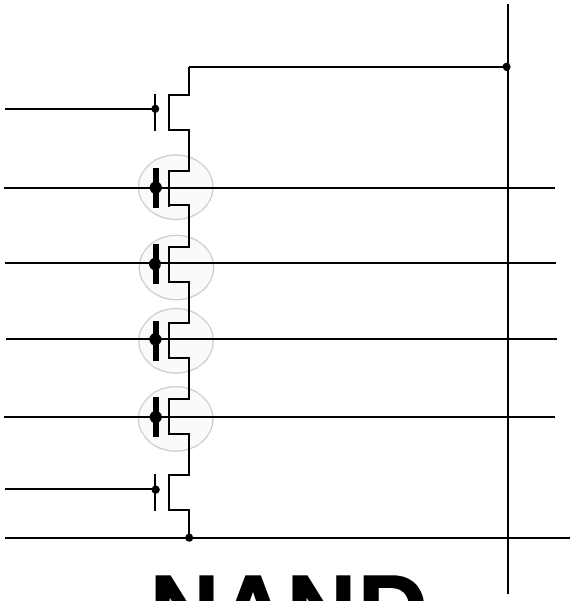
NOR Flash Array



NAND versus NOR



NOR



NAND

Better E/W Endurance (>100K vs >10K)

Smaller Cell Size (~40%)

Fast Read (~100ns)

Slow Read (~1 us)

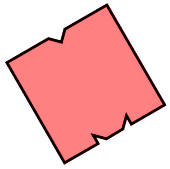
Slow Write (~10 us)

“Fast” Write (~1 us)

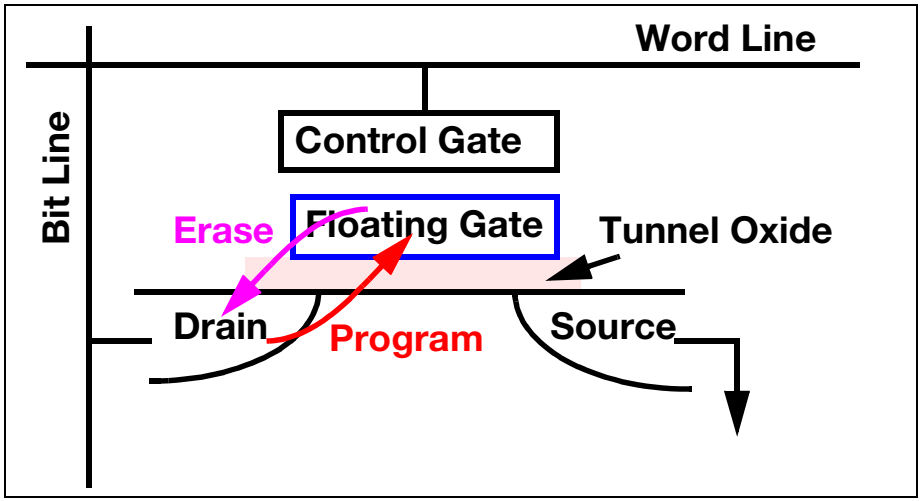
Used for Code

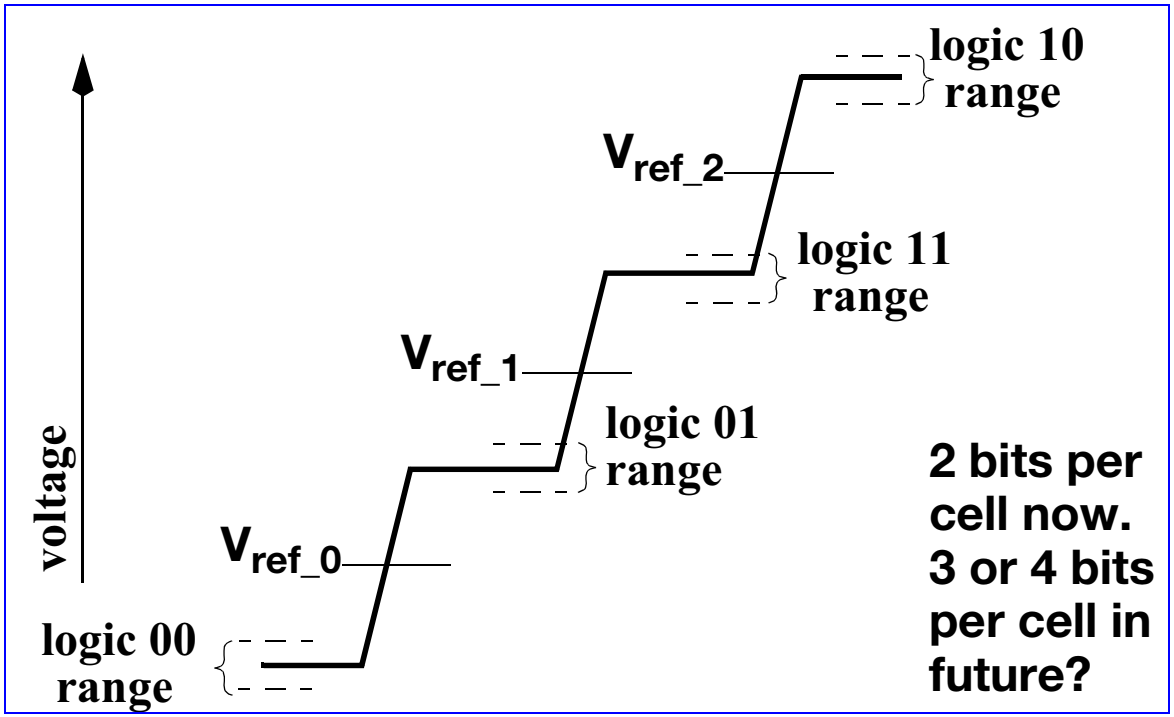
Used for Data

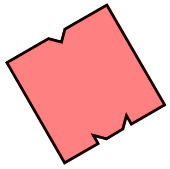
***** values accurate as of 2003 … updated table in a moment**



Multi (voltage) Level Cell

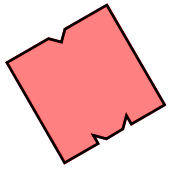




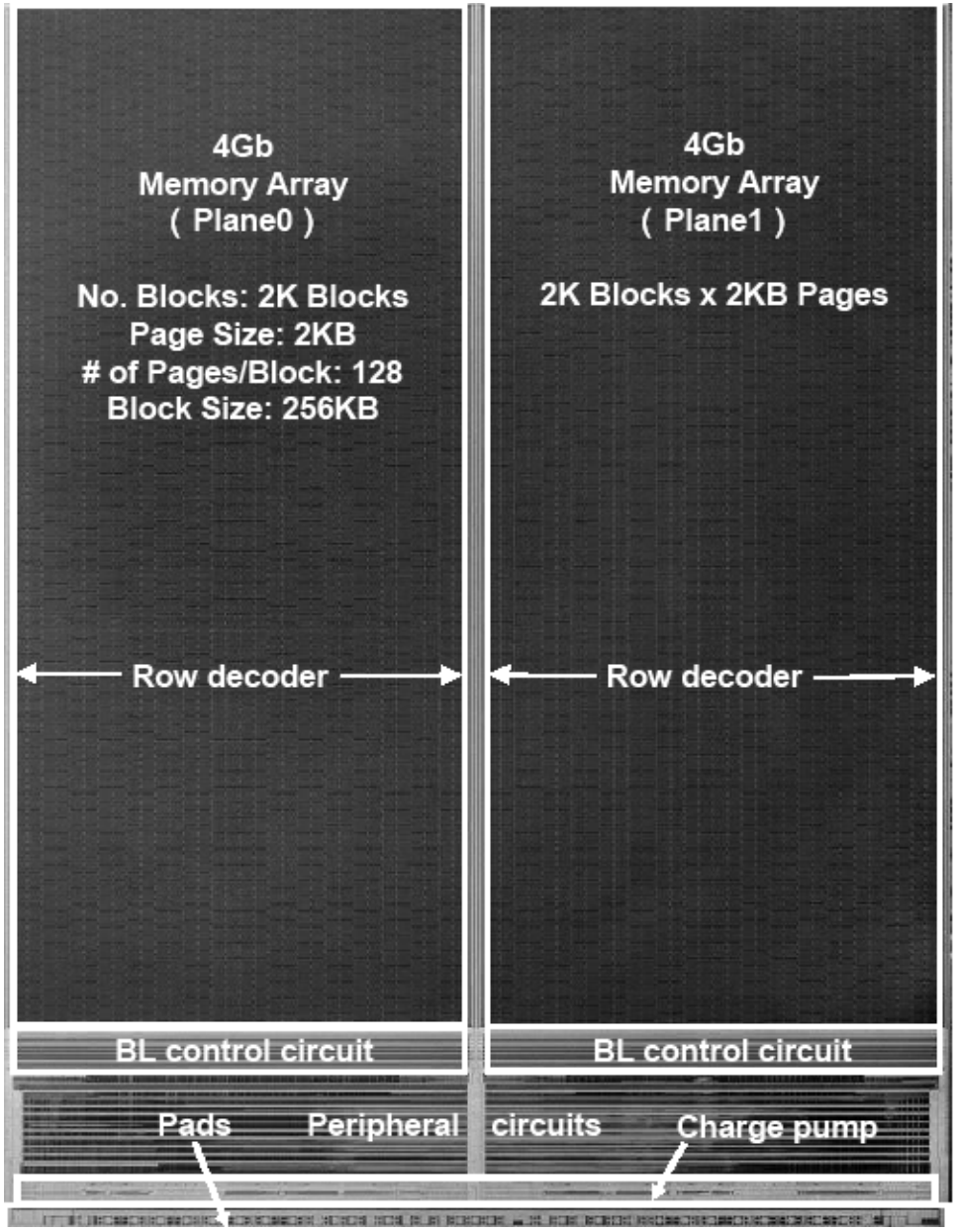


Reads and Writes

- Reads are relatively straightforward
- Writes are complex
- How long do we hold the reverse bias currents to “erase”?
- Did the cells erase properly?
- Did the write succeed?
  - If the write failed, recover, remap and re-write to another sector/block



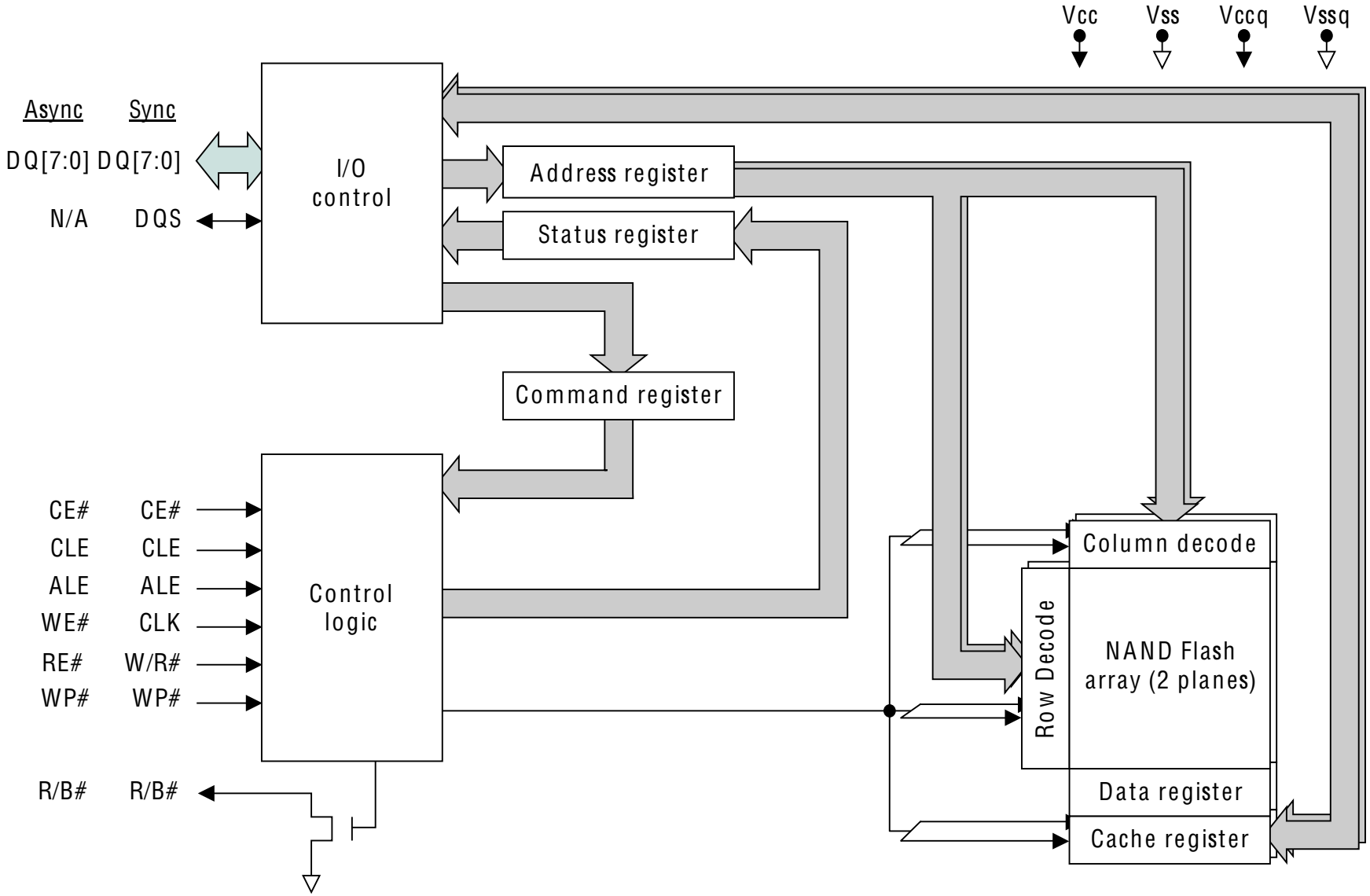
ISSCC 2005: Toshiba 8 Gb NAND



die size: 146 mm^2
70 nm CMOS
3M (2Al, 1 W)
cell size: 0.024um^2 (5 F^2)

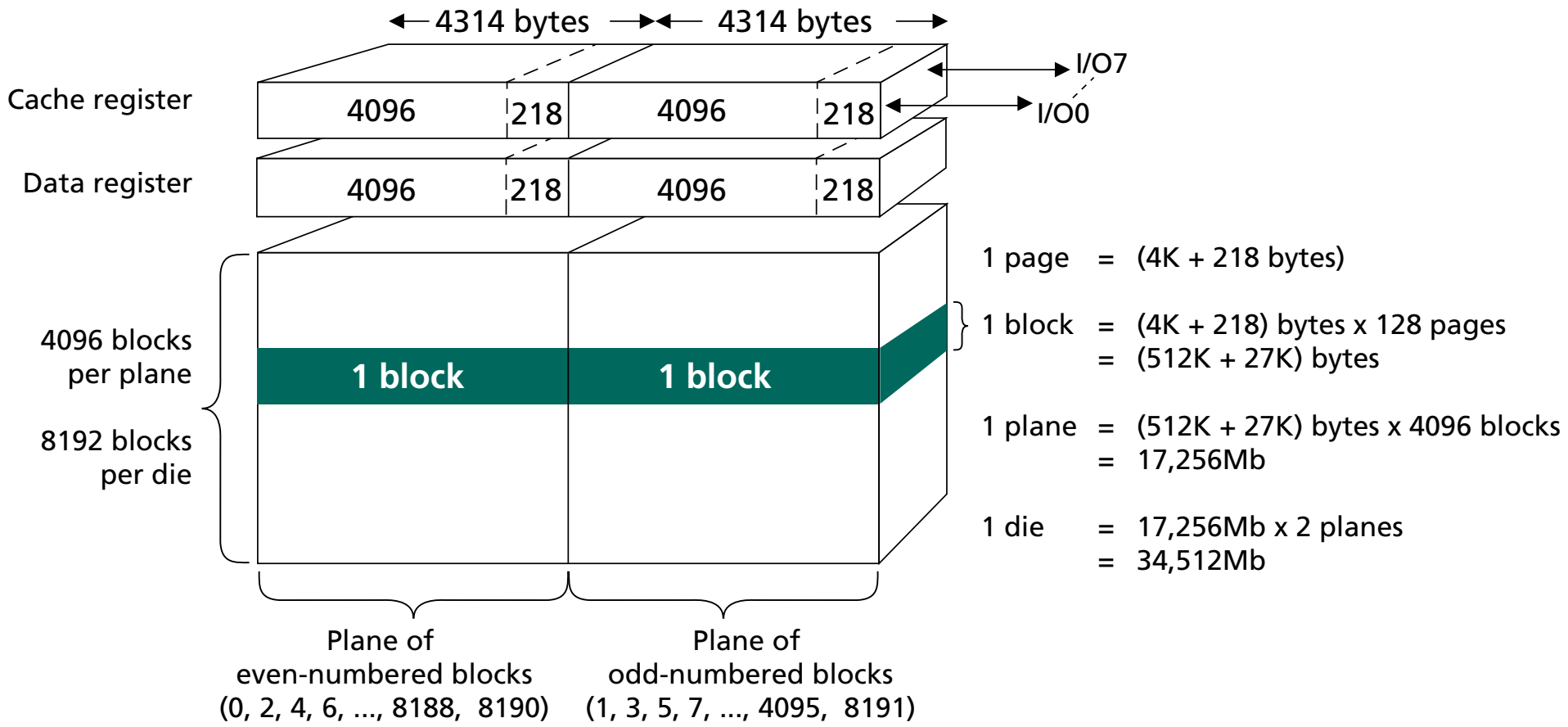
2 bits per cell
cycle time: 50ns
program time: 670us
programming throughput: 6 MB/s

Banks vs. Planes



Banks vs. Planes

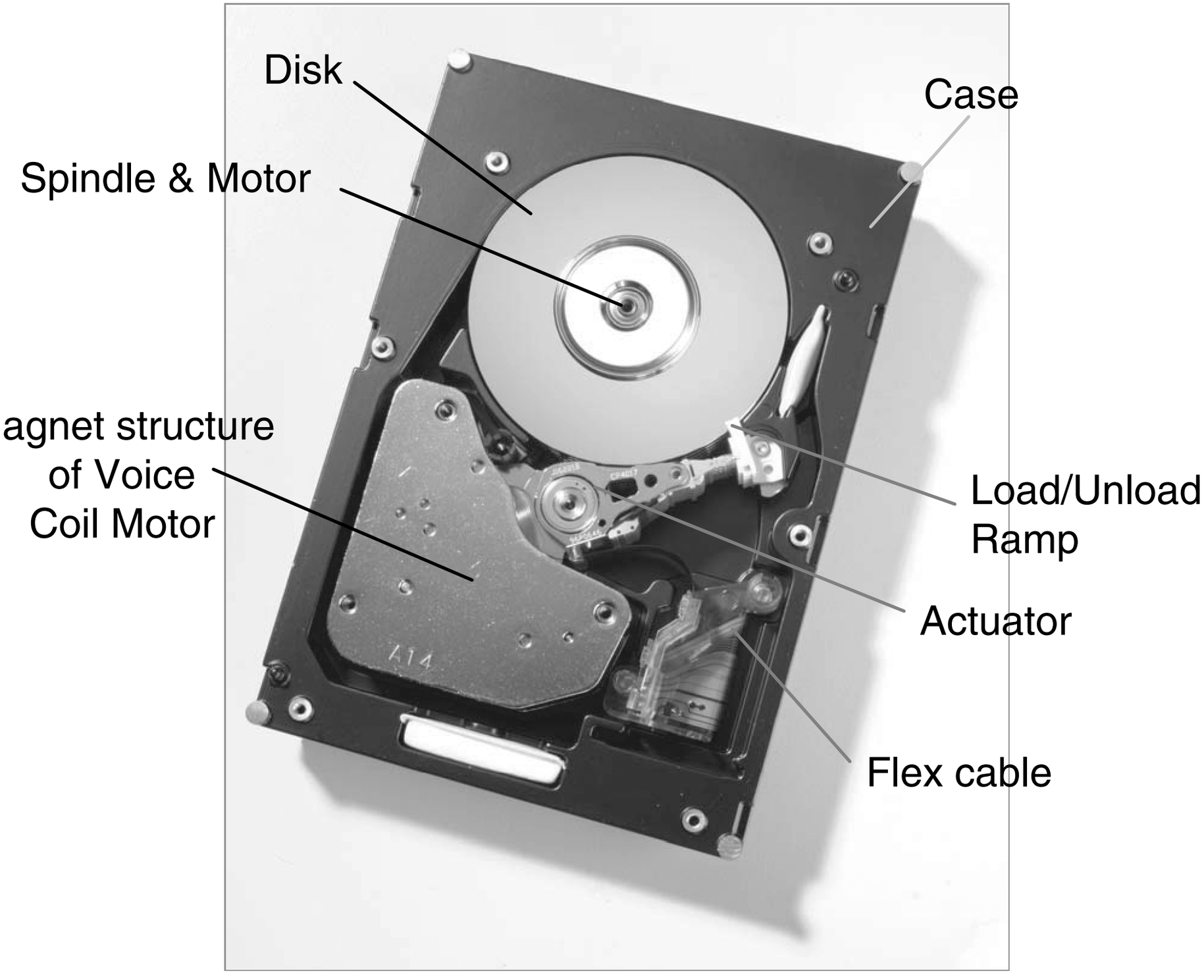
Figure 8: Array Organization: 32Gb and 64Gb Devices



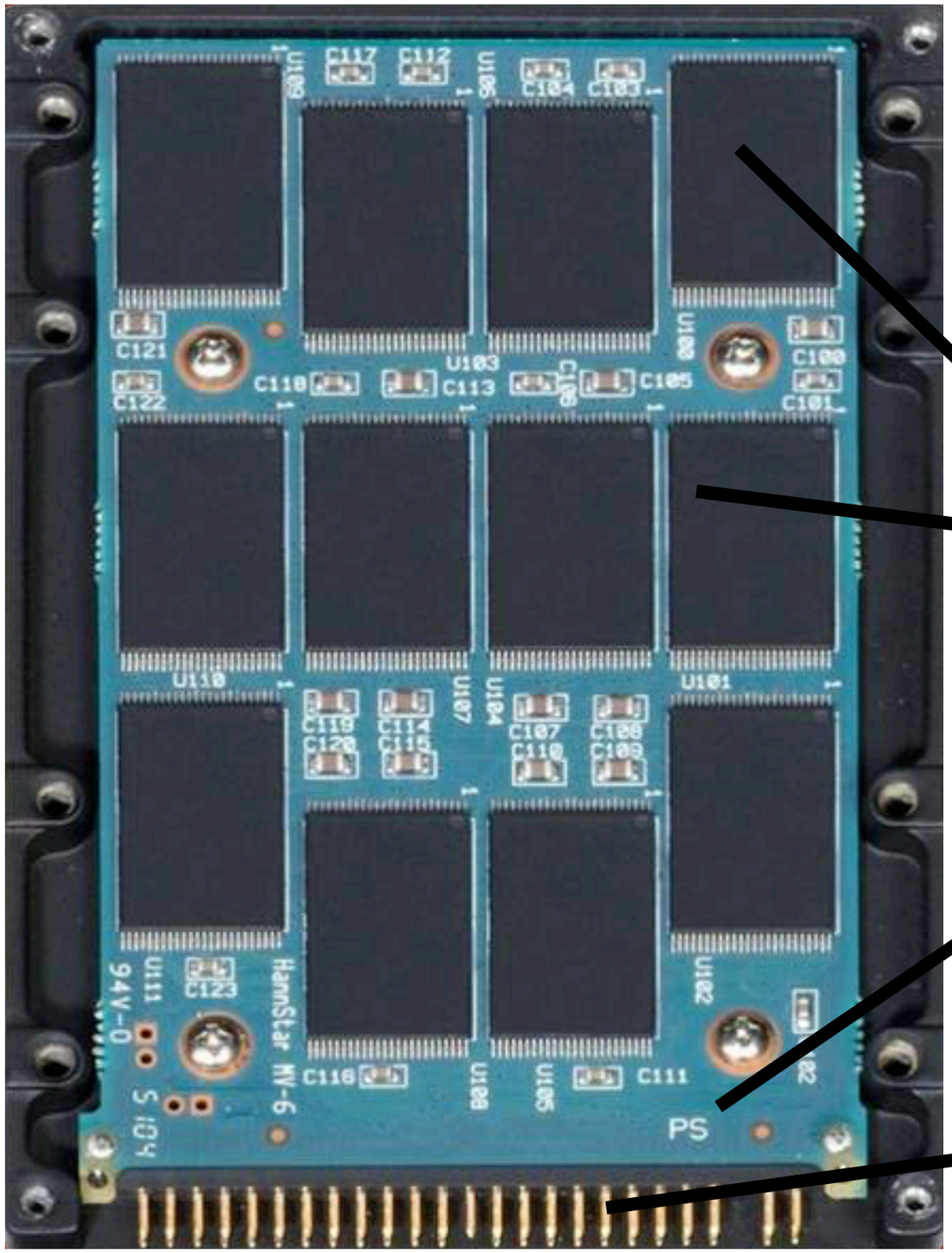
DISK & FLASH

A brief interlude

Disk


Disk
Case
Spindle & Motor
Magnet structure
of Voice
Coil Motor
Load/Unload
Ramp
Actuator
Flex cable

Flash SSD


C117
C112
U106
C104
C103
U109
C121
C122
C118
U103
C113
C105
C100
C101
U108
U110
U107
U104
U101
C119
C120
C114
C115
C107
C110
C108
C109
U111
C123
94V-0
HannStar MV-6
C116
U102
C111
PS
Flash memory arrays
Circuit board
ATA Interface

Disk Issues

- Keeping ahead of Flash in price-per-GB is difficult (and expensive)
- Dealing with timing in a polar-coordinate system is non-trivial
    - OS schedules disk requests to optimize both linear & rotational latencies; ideally, OS should not have to become involved at that level
- Tolerating long-latency operations creates fun problems
    - E.g., block-fill not atomic; must reserve buffer for duration; Belady's MIN designed for disks & thus does not consider incoming block in analysis
- Internal cache & prefetch mechanisms are slightly behind the times

Flash SSD Issues

- Flash does not allow in-place update of data (must block-erase first);
  implication is significant amount of garbage collection & storage management

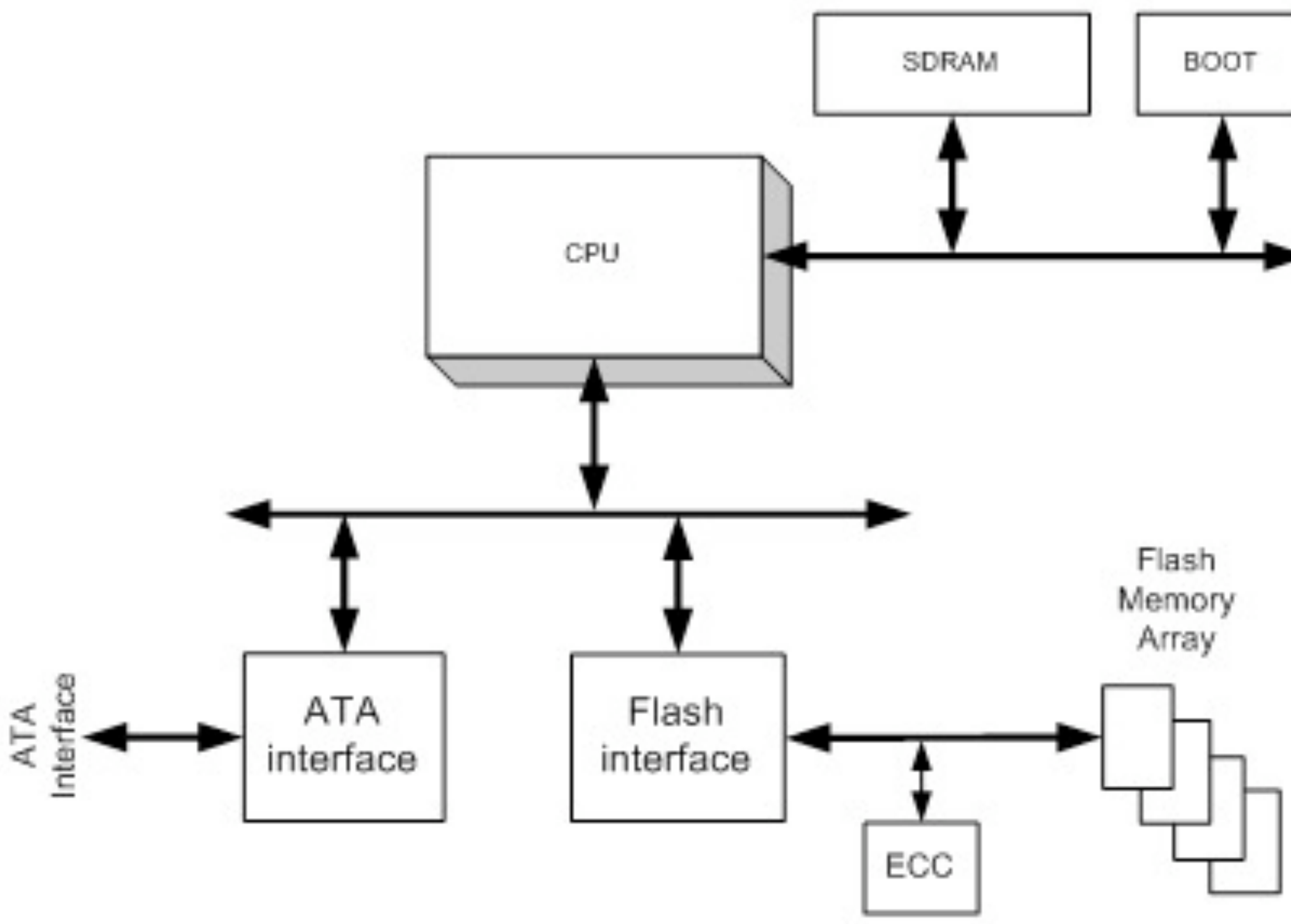
- Asymmetric read [1x] & program times [10x] (plus erase time [100x])

- Proprietary firmware (heavily IP-oriented, not public, little published)

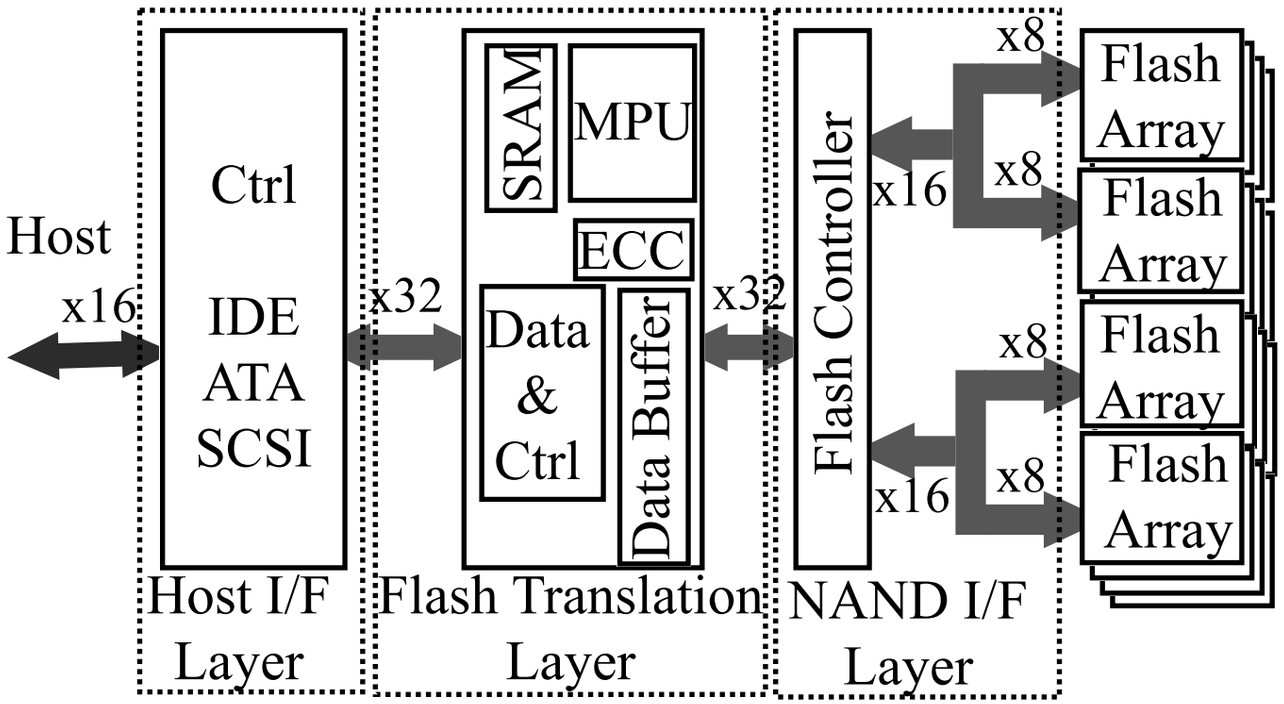
  - Lack of models: timing/performance & power, notably
    Flash Translation Layer is a black box (both good & bad)
    Ditto with garbage collection heuristics, wear leveling, ECC, etc.

  - Result: poorly researched (potentially?)
    E.g., heuristics? how to best organize concurrency? etc.

SanDisk SSD Ultra ATA 2.5” Block Diagram

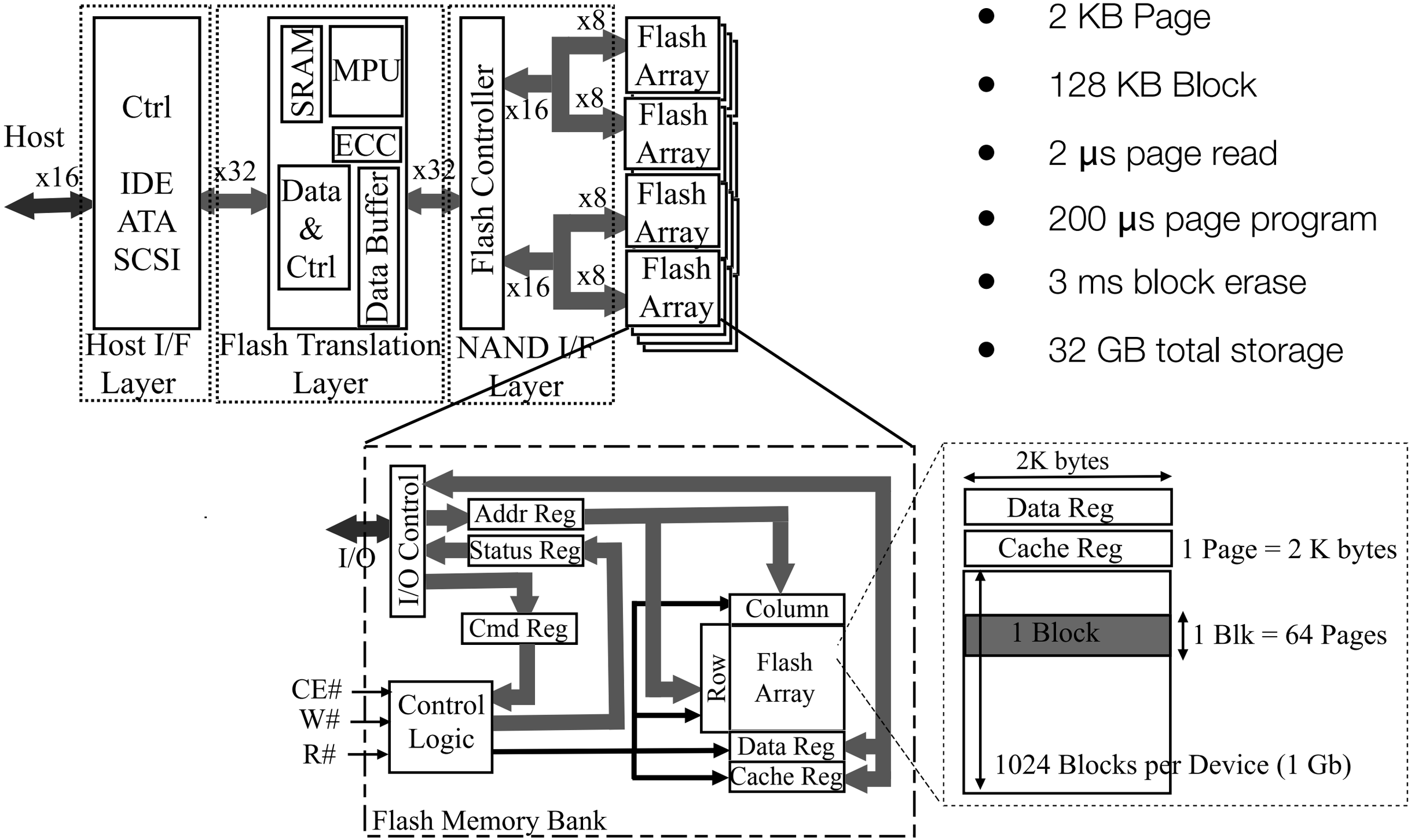


Flash SSD Organization & Operation



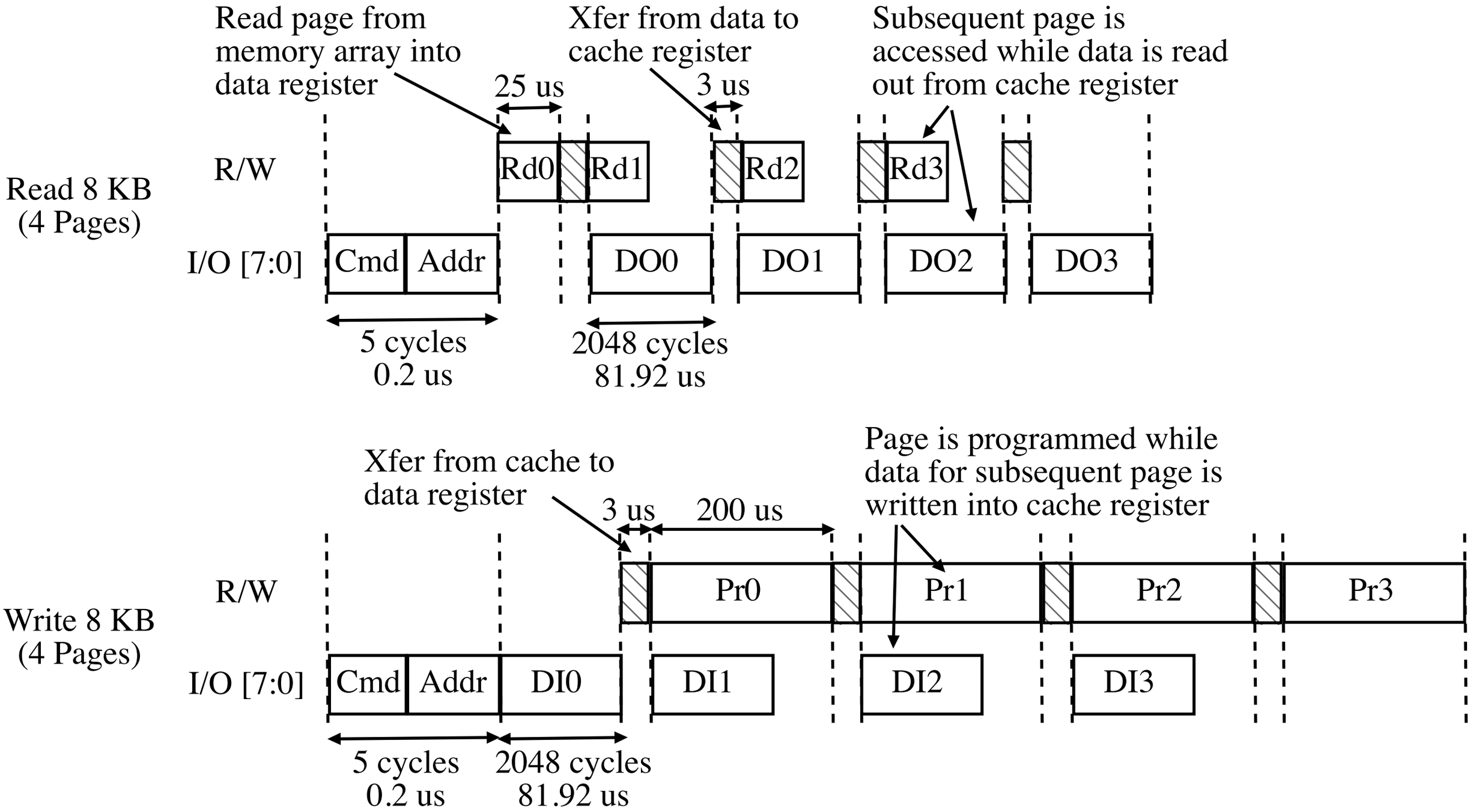
- Numerous Flash arrays
- Arrays controlled externally (controller rel. simple, but can stripe or interleave requests)
- Ganging is device-specific
- FTL manages mapping (VM), ECC, scheduling, wear leveling, data movement
- Host interface emulates HDD

Flash SSD Organization & Operation

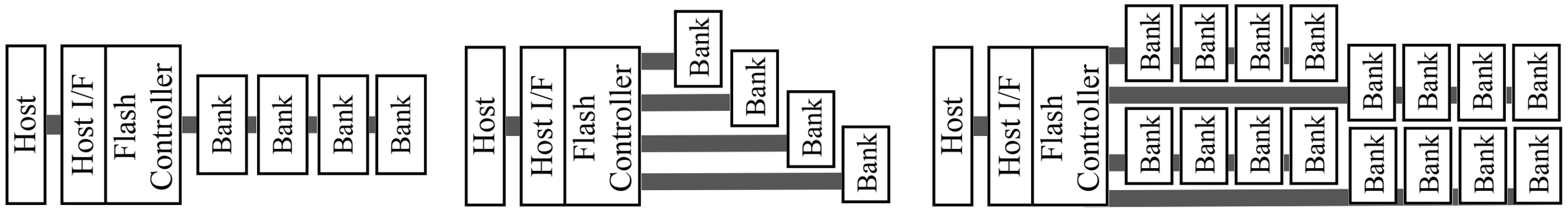


- 2 KB Page
- 128 KB Block
- 2 µs page read
- 200 µs page program
- 3 ms block erase
- 32 GB total storage

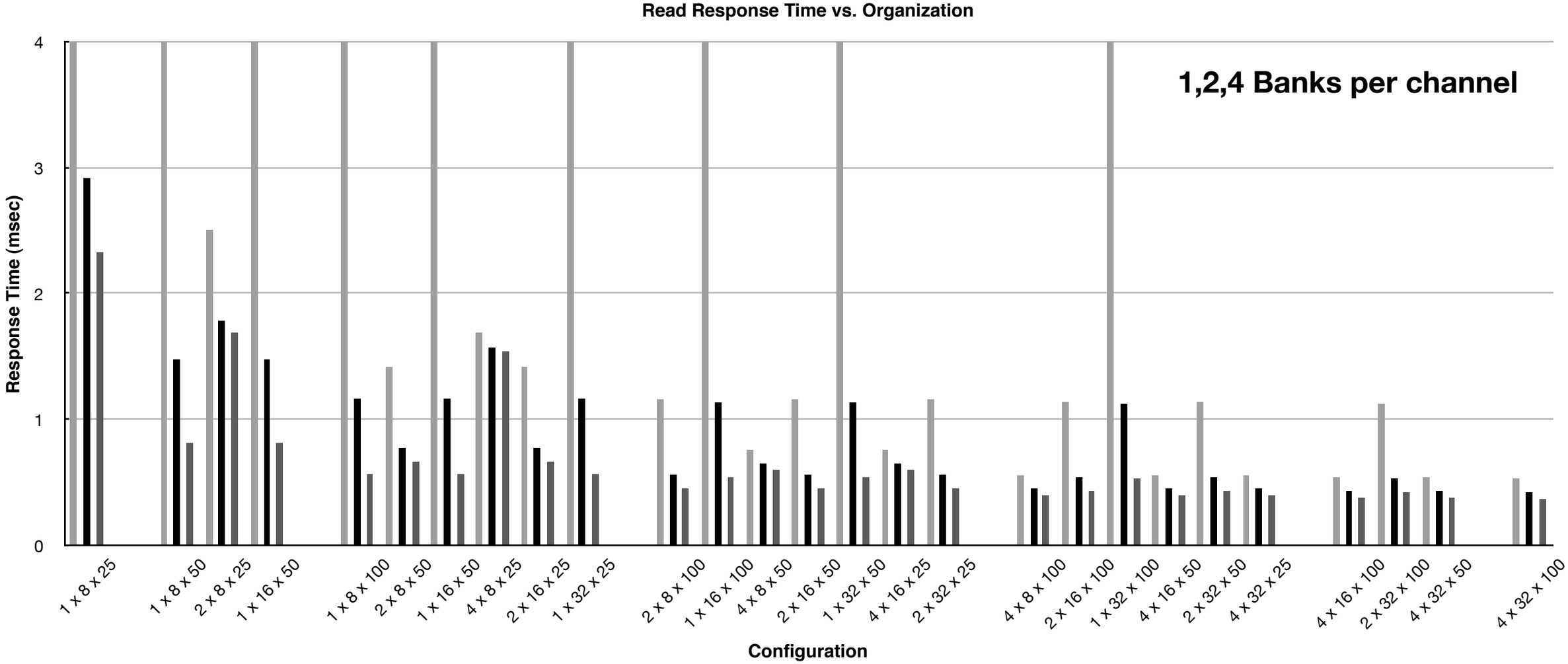
Flash SSD Timing



Some Performance Studies

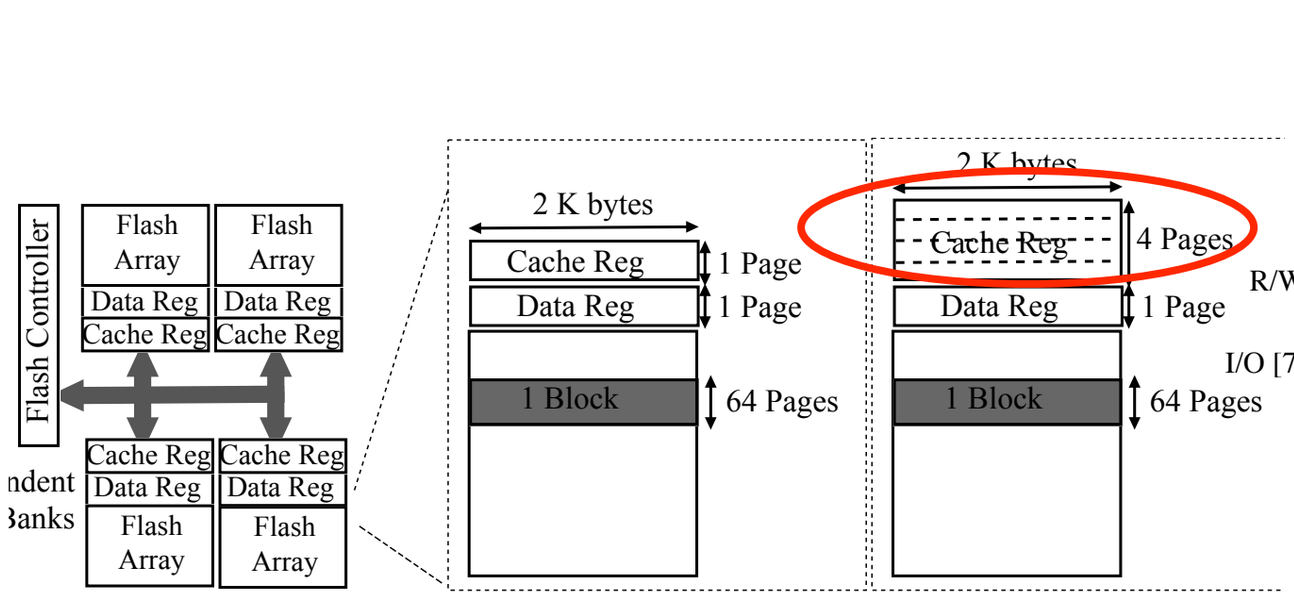


(a) Single channel (b) Dedicated channel for each bank (c) Multiple shared channels

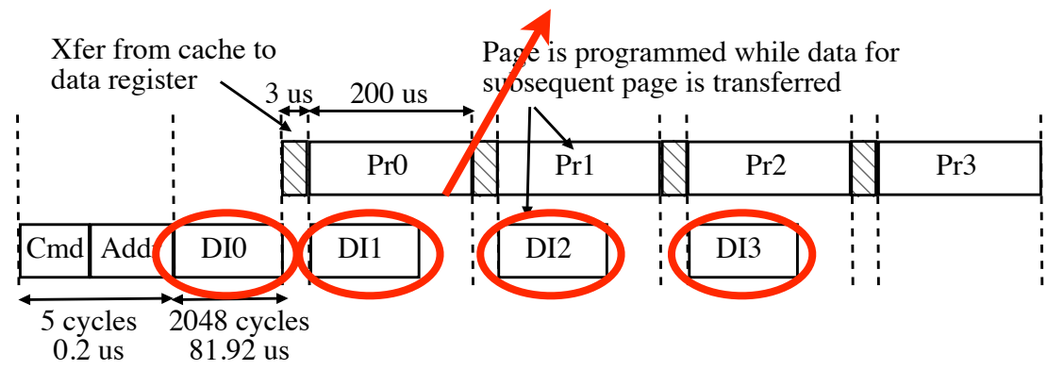


I/O Access Optimization

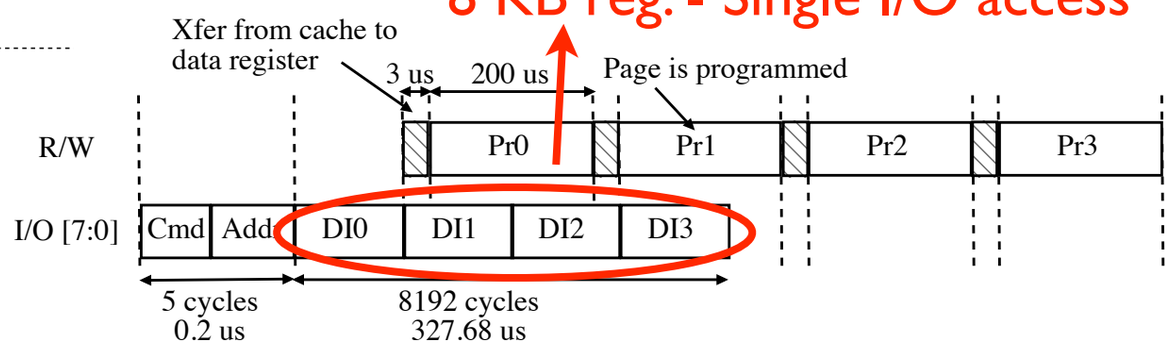
- Access time increasing with level of banking on single channel
- Increase cache register size



8 KB Write, 2 KB reg. - 4 I/O accesses



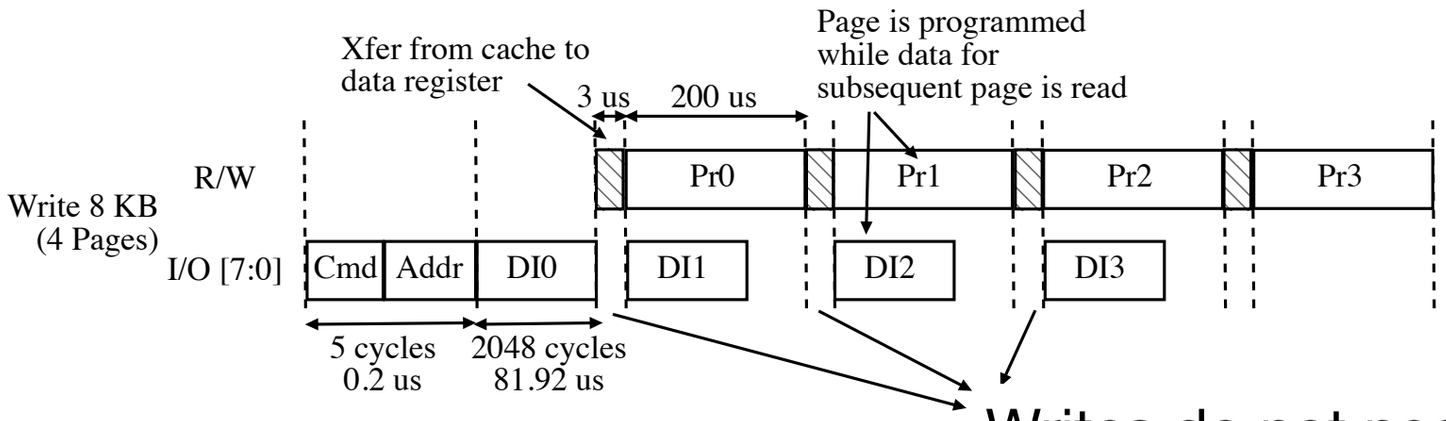
8 KB reg. - Single I/O access



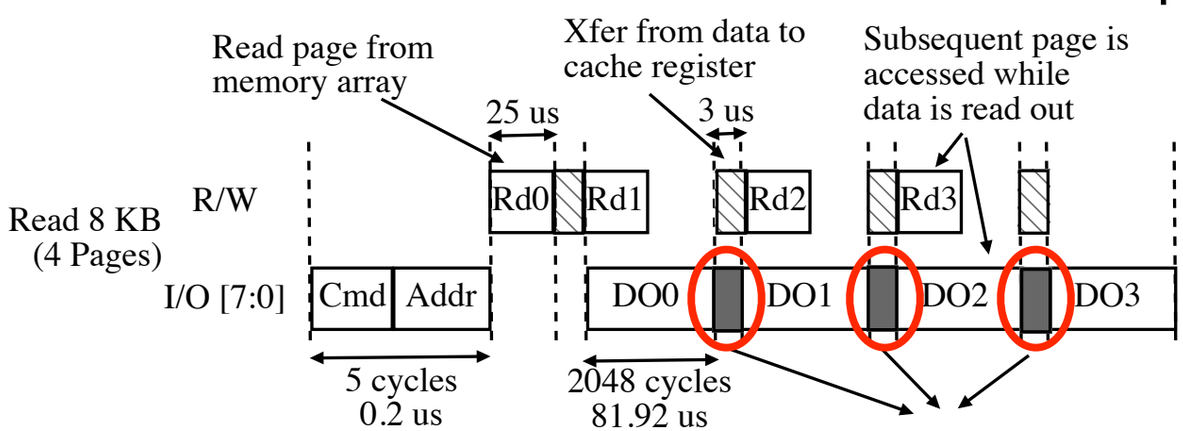
- Reduce # of I/O access requests

I/O Access Optimization

- Implement different bus-access policies for reads and writes



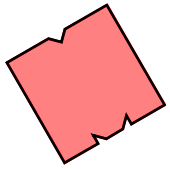
Writes do not need I/O access as frequently as reads



Reads: Hold I/O bus between data bursts

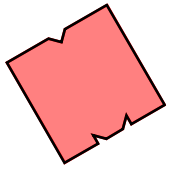
Flash Evolution

Parameter	2004 SLC	2007 SLC	2008 MLC	2010 SLC	2010 MLC
Generation	2Gb	2Gb	32Gb	32Gb	128Gb
# Blocks	2048	2048	8192	4096	8192
Block Size	128KB	128KB	512KB	1MB	2MB
Page Size +ECC	2KB +64	2KB +64	4KB +218	8KB +448	8KB +448
Pages/Blk	64	64	128	128	256
Planes/Dies	1/1	1/1	2/4	2/1	2/2
Read Time	25µs	25µs	50µs	35µs	75µs
Write Time	300µs	300µs	900µs	350µs	1600µs
Erase Time	2000µs	2000µs	3000µs	1500µs	5000µs
Longevity	100K	100K	10K	60K	30K
Max BW	33MT/s	40MT/s	50MT/s	200MT/s	200MT/s

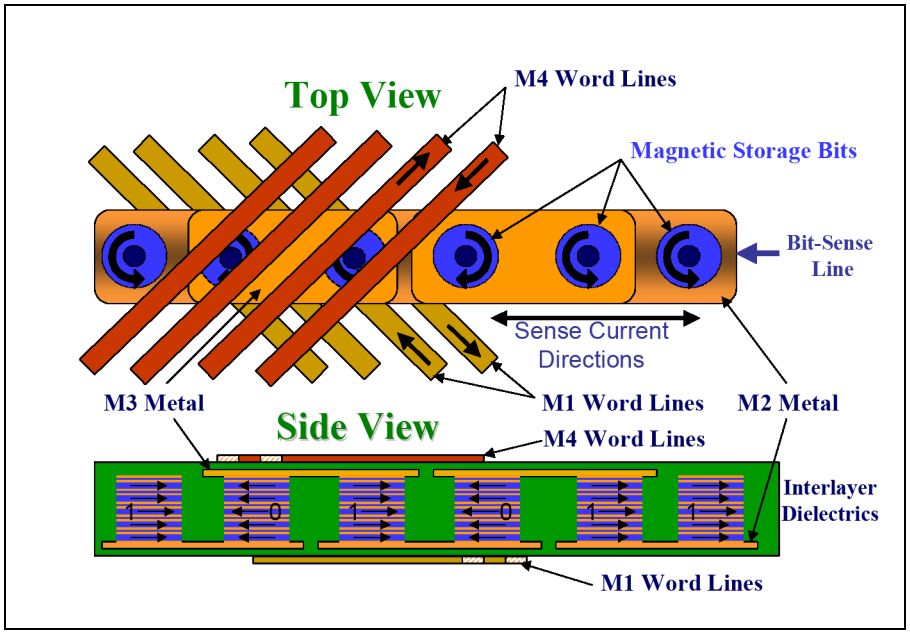


Flash Summary

- **Comparable read latency (to DRAM) and slow writes**
- **Non Volatile**
- **limited write cycles**
- **Very mature industry**
- **Very high density, but long term scalability? Oxide reliability, multi-level cell.**
- **Video applications now supported**



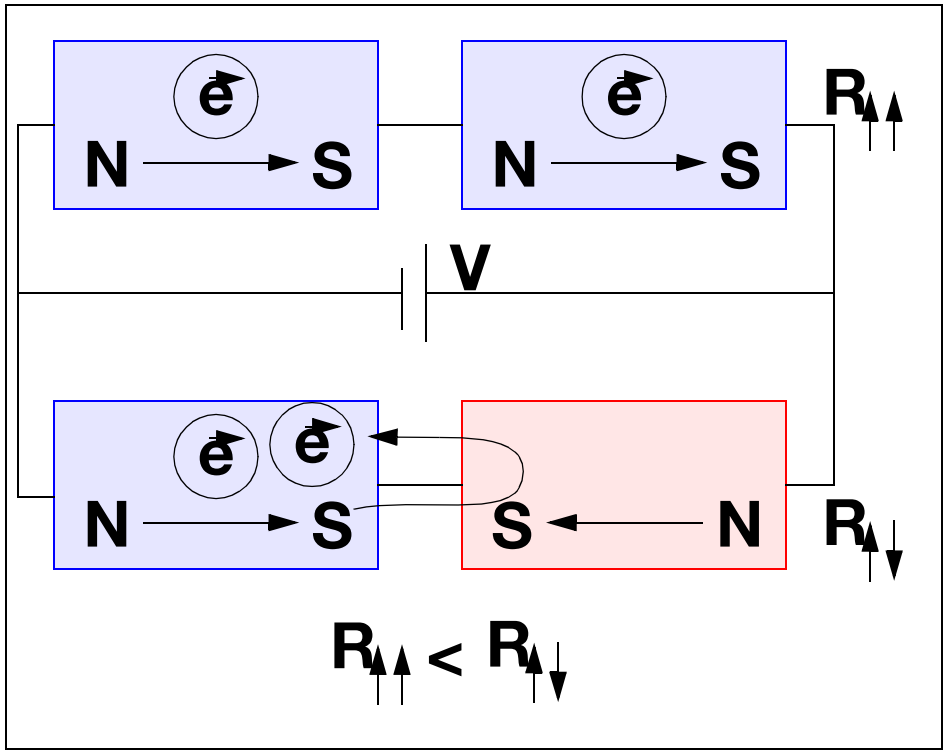
MRAM

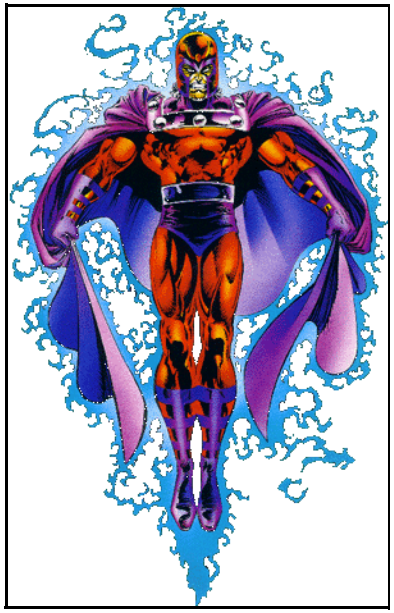


MRAM Memory Cell Structure

Magnetoresistance

- Change in electrical resistance with applied magnetic field

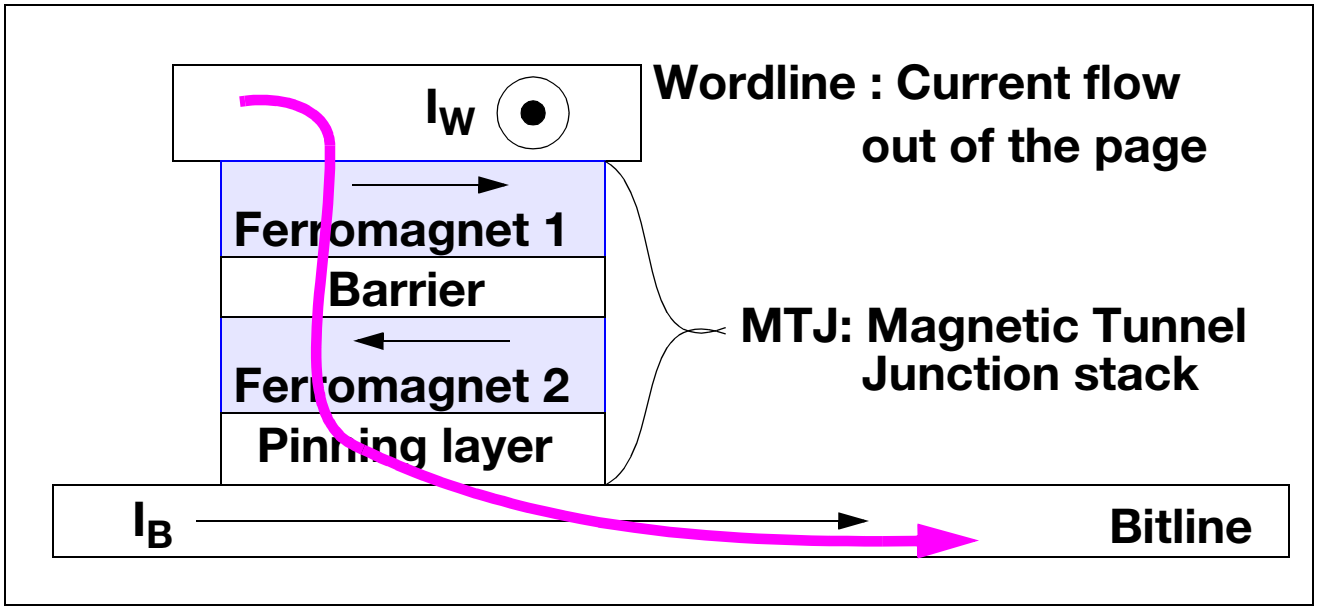




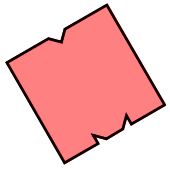
Magneto

- The resistance against the flow of electrons depend on polarization of electrons and availability of spin-up or spin-down states

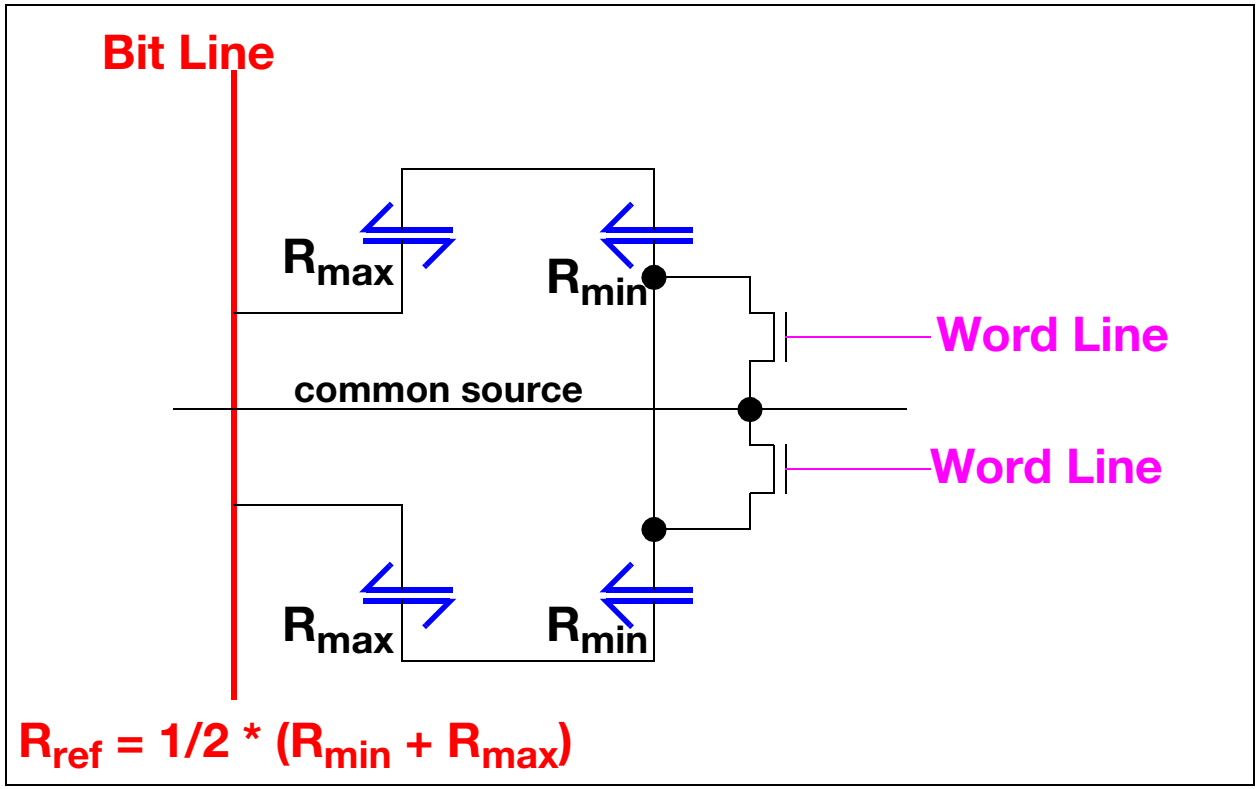
MRAM Cell Structure



Differences in resistance = “0” or “1”

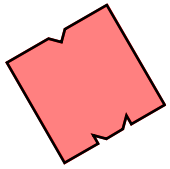


MRAM Reference Circuit

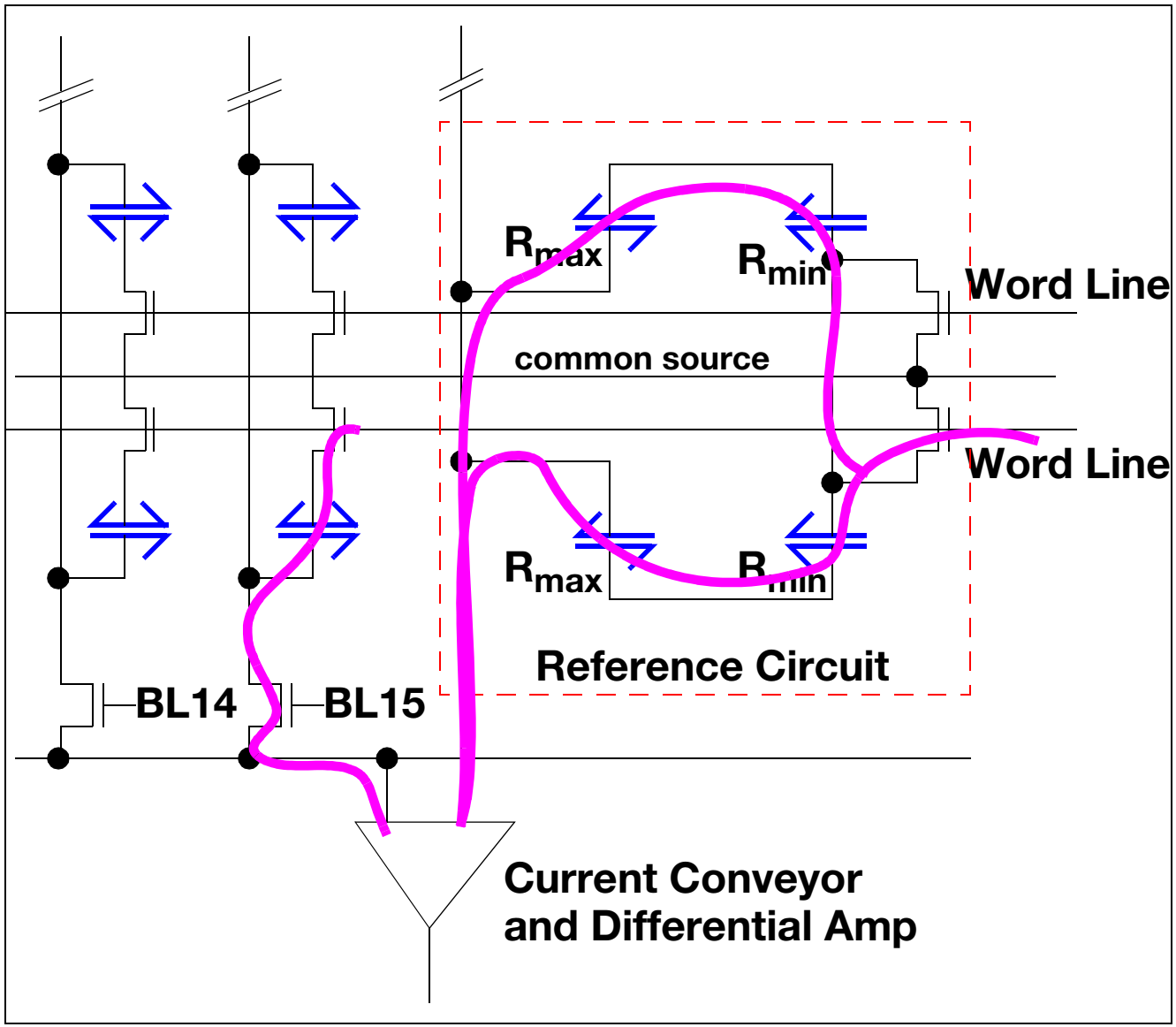


Reference Cell

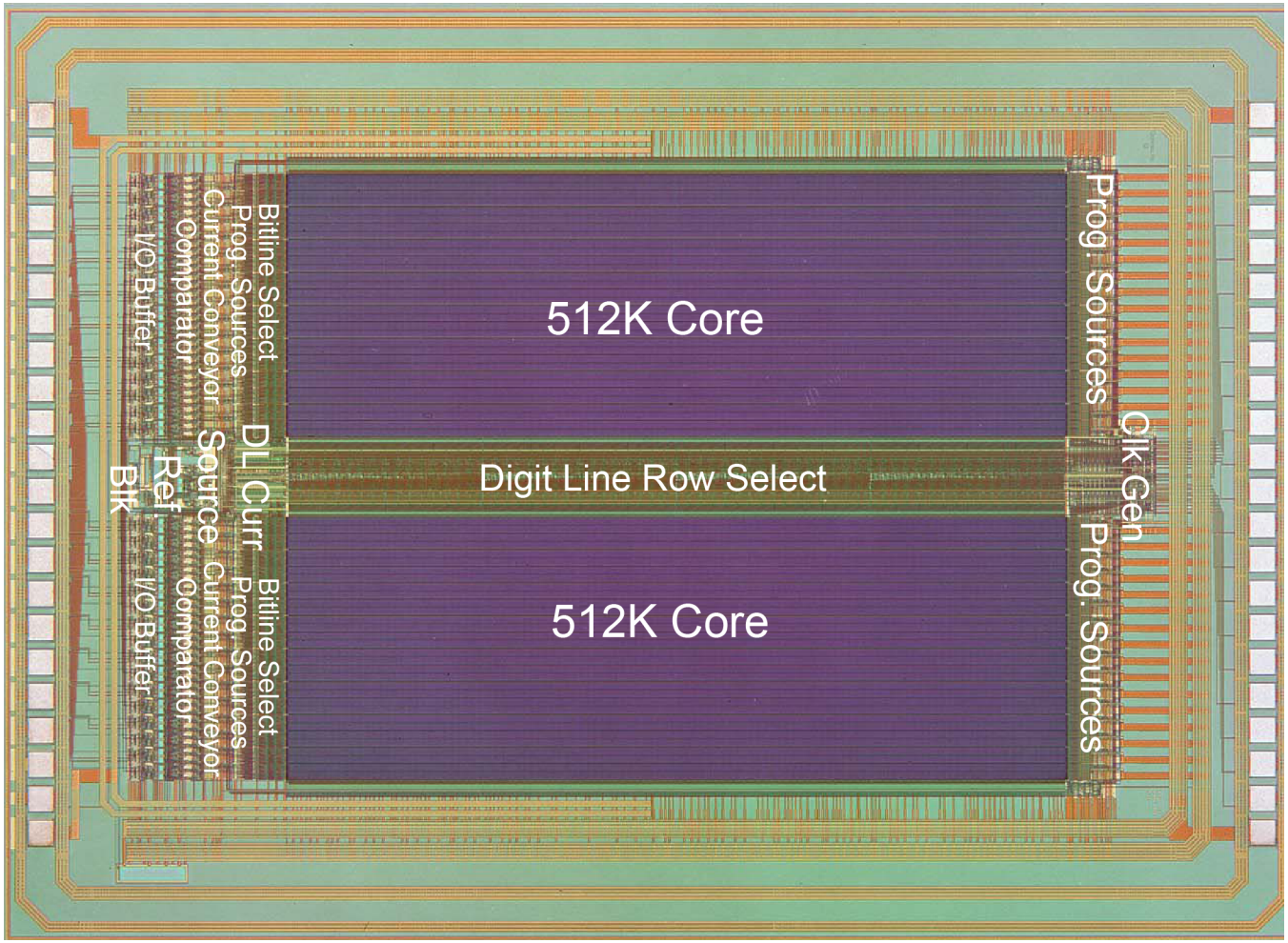
Reference Cell uses Parallel/Serial combination of MTJ's in two memory states to generate "mid resistance" reference between those two states



MRAM Segment

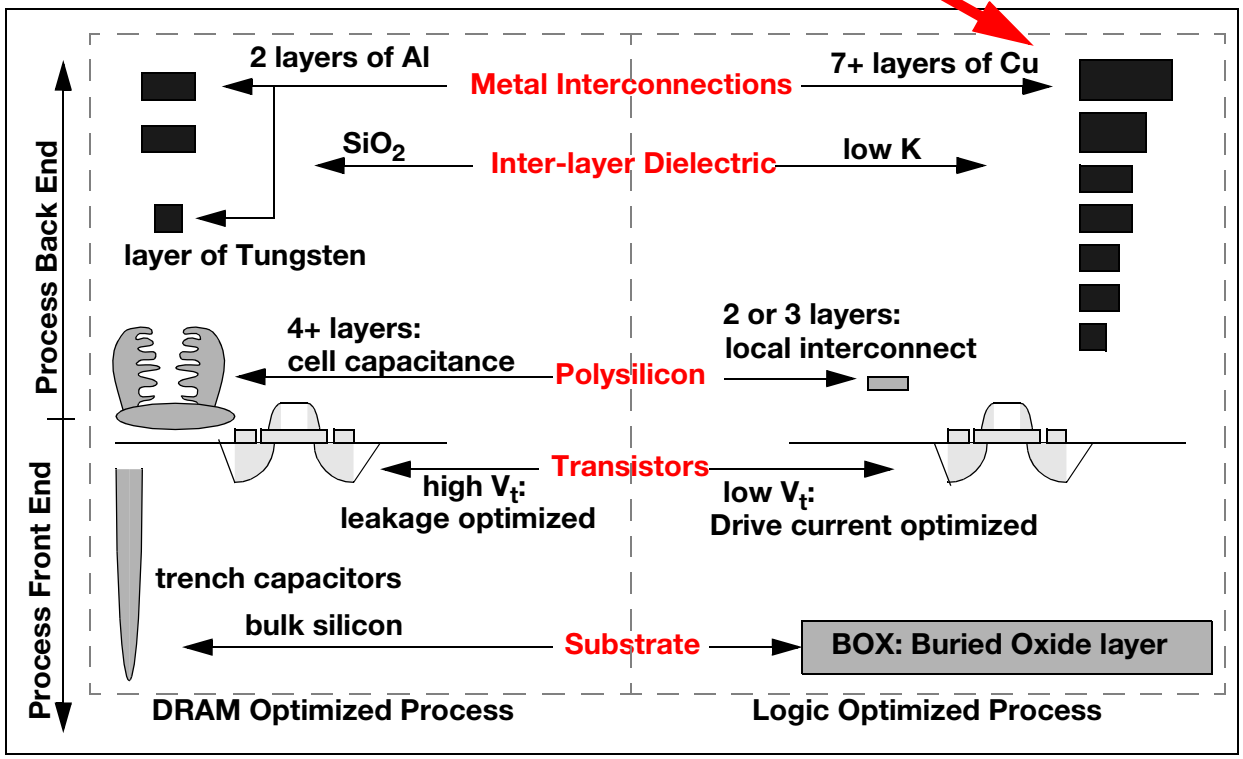


Motorola 1 MBit MRAM Chip



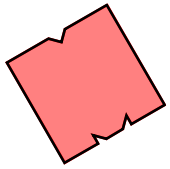
Process Compatibility

Can be built between metal layers above active silicon

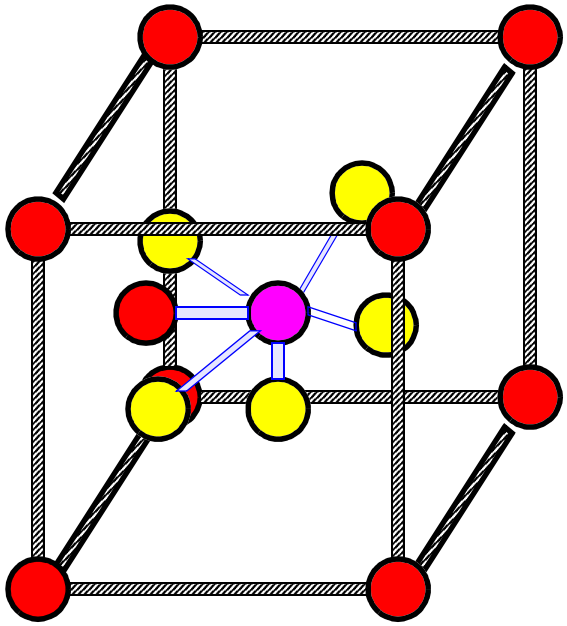


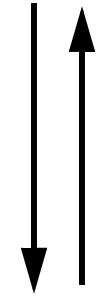
MRAM Summary

- **Non Volatile**
- **No need to refresh**
- **(potentially) High density**
- **Non destructive read**
- **High write current.**
- **Read speed = write speed; < 50ns**
- **Unlimited R/W endurance**
- **Soft error immunity**
- **Material compatibility with CMOS logic and DRAM?**
- **Currently, large cells, but *may* scale down to DRAM/Flash levels of 6 F^2.**



FeRAM Cell





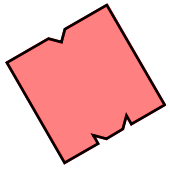
Tetra/Pentavalent Atom

Di/Monovalent Metal Atoms

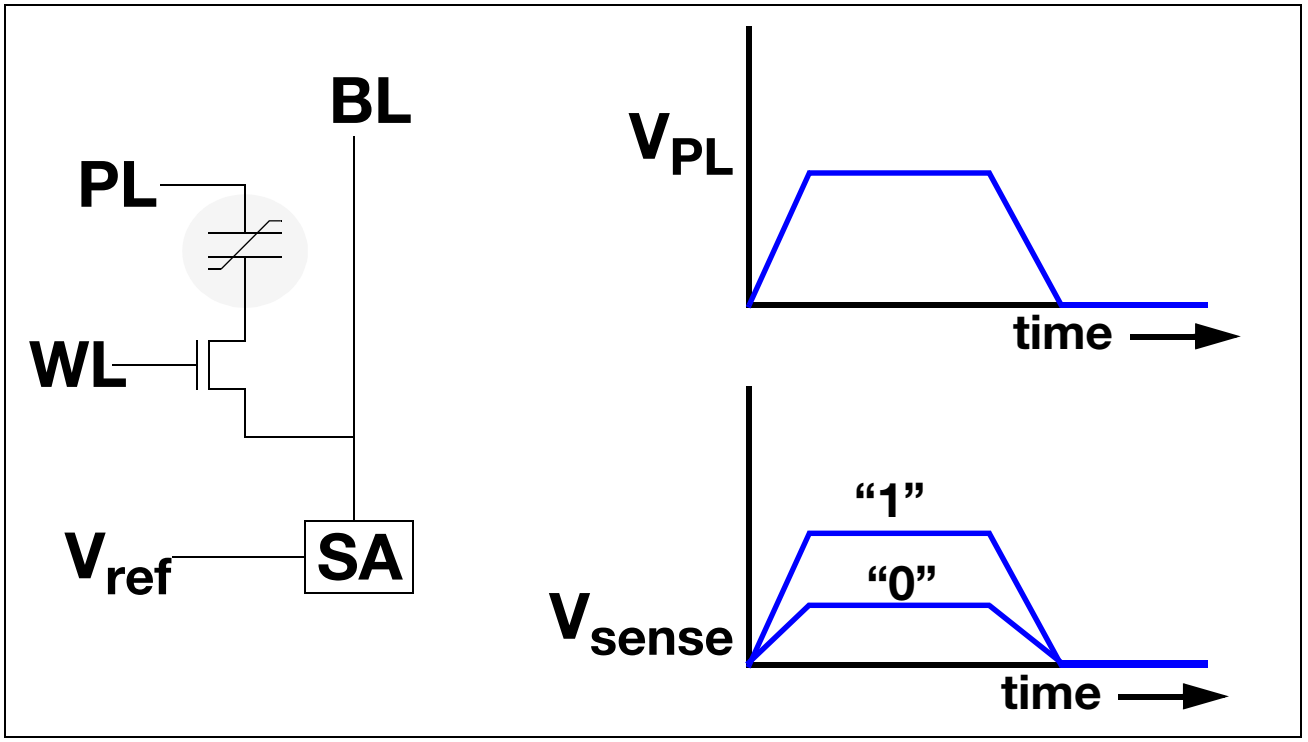
Oxygen Atoms

PZT (PbO, ZrO_2, TiO_2) Lead-Zirconate-Titanate unit cell

Applied Electric Field Moves Center Atom

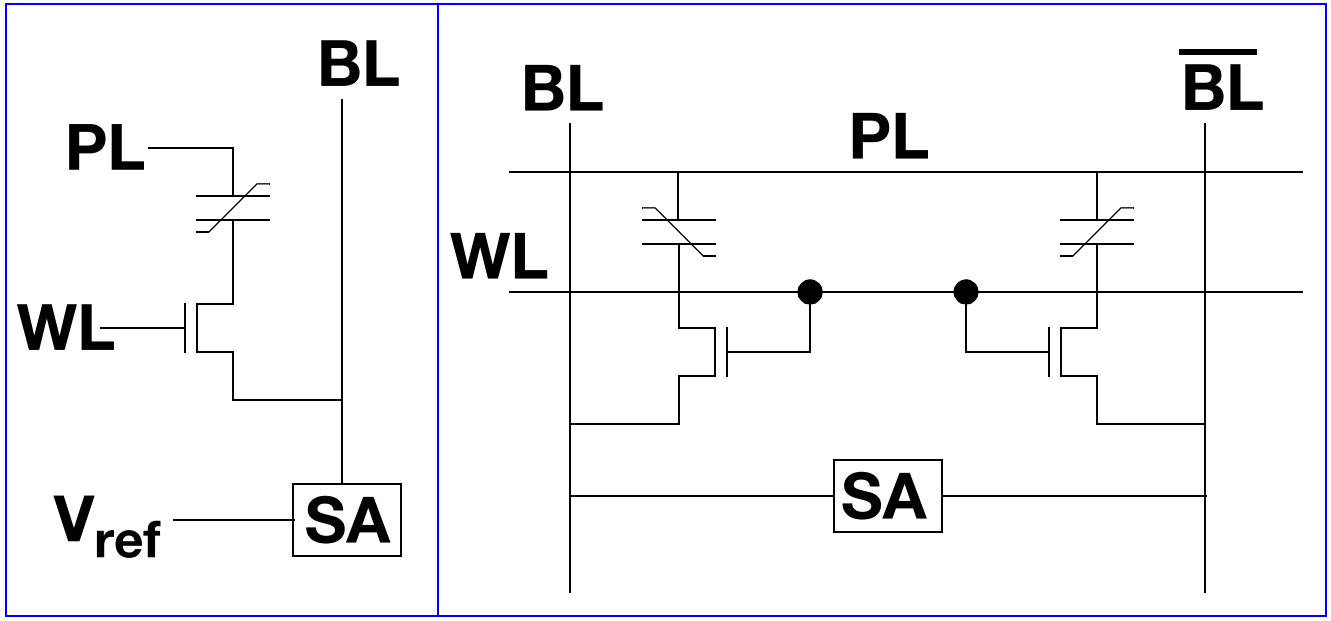


FeRAM Cell Operation



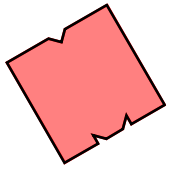
- Ferroelectric material can be polarized into two stable states. States can be maintained without power

FeRAM Circuit Structure



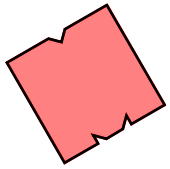
1T1C: DRAM-like

2T2C: Built-in reference

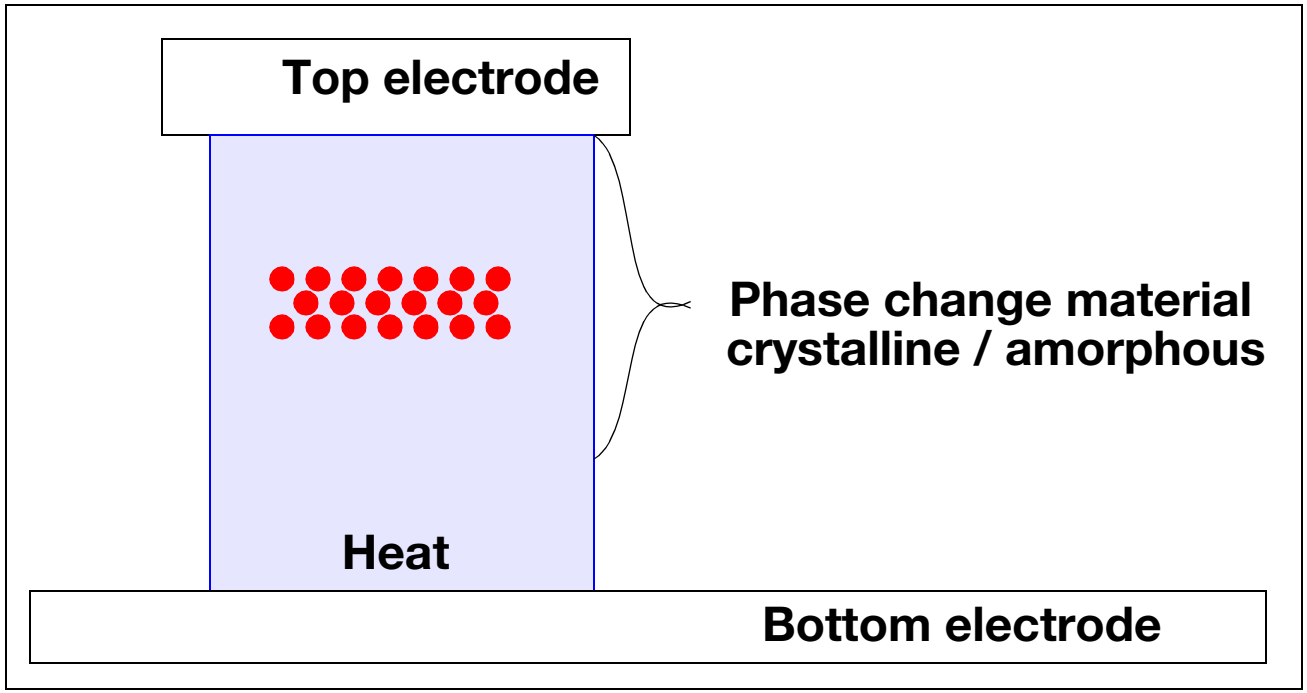


FeRAM Summary

- **Non-volatile (no refresh)**
- **Destructive read**
- **Low voltage and low power**
- **Fast read and fast write (compared to NVM)**
- **limited R/W endurance (< 10^9 reads)**



Phase Change Memory

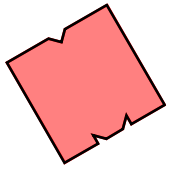


Heat, then . . .

Fast cool down = amorphous. High resist.
Slow cool down = crystalline. Low resist.

Differences in resistance = “0” or “1”

Same material used in re-writeable CD/DVD optical disks (GeSbTe)



PCRAM Summary

- Rad Hard (SEU has limited/no effect)
- Non destructive read
- Direct write (no need for DRAM Sense Amp style Read-Modify-Write)
- Low voltage and low power
- Fast read and medium speed write (fast compared to NVM)
- Unlimited read cycles
- Limited write cycles ($< 10^{12}$); Large write current can overheat element, element can be stuck at low resistance state
- Can scale to 22nm tech (source: Intel)