EECS 427
Lecture 20: Memory reliability and power / ROMS
Reading: 12.4, 12.5

Last Time

- Memory core
  - 6T SRAM and 1T DRAM cells are mainstays of the IC industry
- Memory peripherals (SRAM-oriented)
  - Decoders: For large memories, they take up a large fraction of total access time (speed-critical)
  - Sense amplifiers: required to speed up read times, reduce voltage swing on bit lines

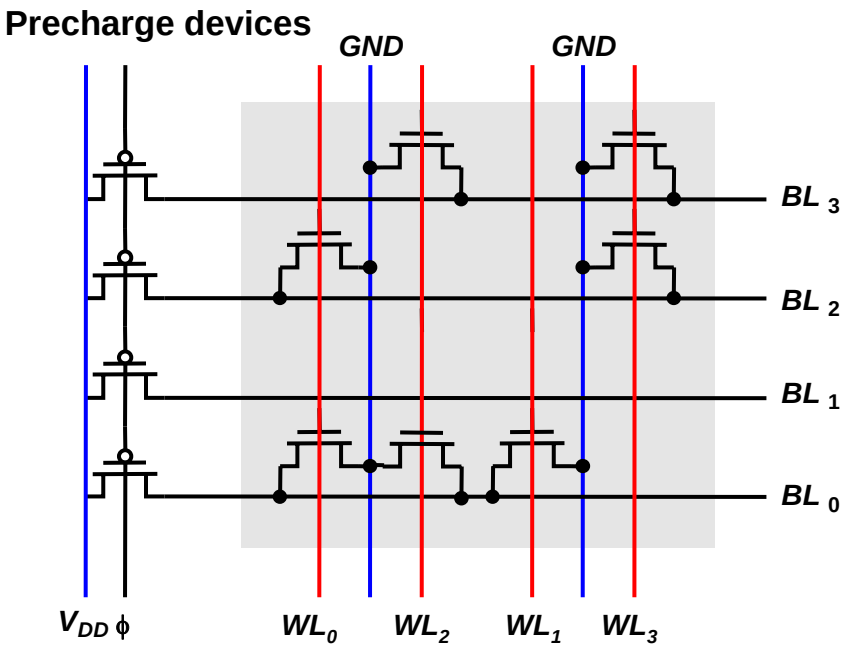
Lecture Overview

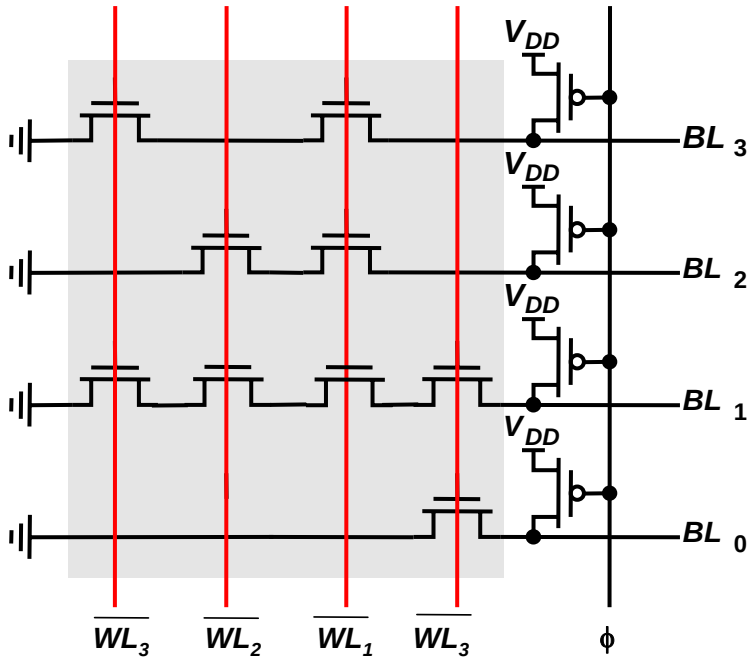
- Flash ROMs becoming big too, how do they work?
- Memories should be inherently robust/reliable by definition
- Memories comprise a growing portion of digital systems and hence their power consumption is appreciable

ROMS

- Decode and layout similar to SRAM
- Except : 1 transistor (or not) per cell
- Stronger cell drive + shorter wires + one-sided output = simpler sensing

ROMS : Nand and Nor

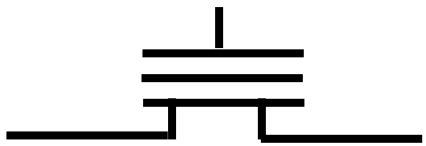




NOR ROM **NAND ROM**

NOR is faster, more SRAM-like, but takes more area

EEPROMS



- Replace ROM gate with floating gate : two poly layers separated by thin oxide
- Can program by placing large voltage on gate to produce tunneling effect.
- Tunneling is destructive, slow and inaccurate (might get a transistor that can’t be switched so use as erasable fuse instead of transistor)

Flash ROM

- Like EEPROM but better
- Monitor effect of programming on gate dynamically for better control so you get transistor or short (nand) / open (nor)
- Erase & Write in blocks for speed & efficiency via parallelism / shared control

Reliability and Yield

- Semiconductor memories trade-off noise margin for density and performance

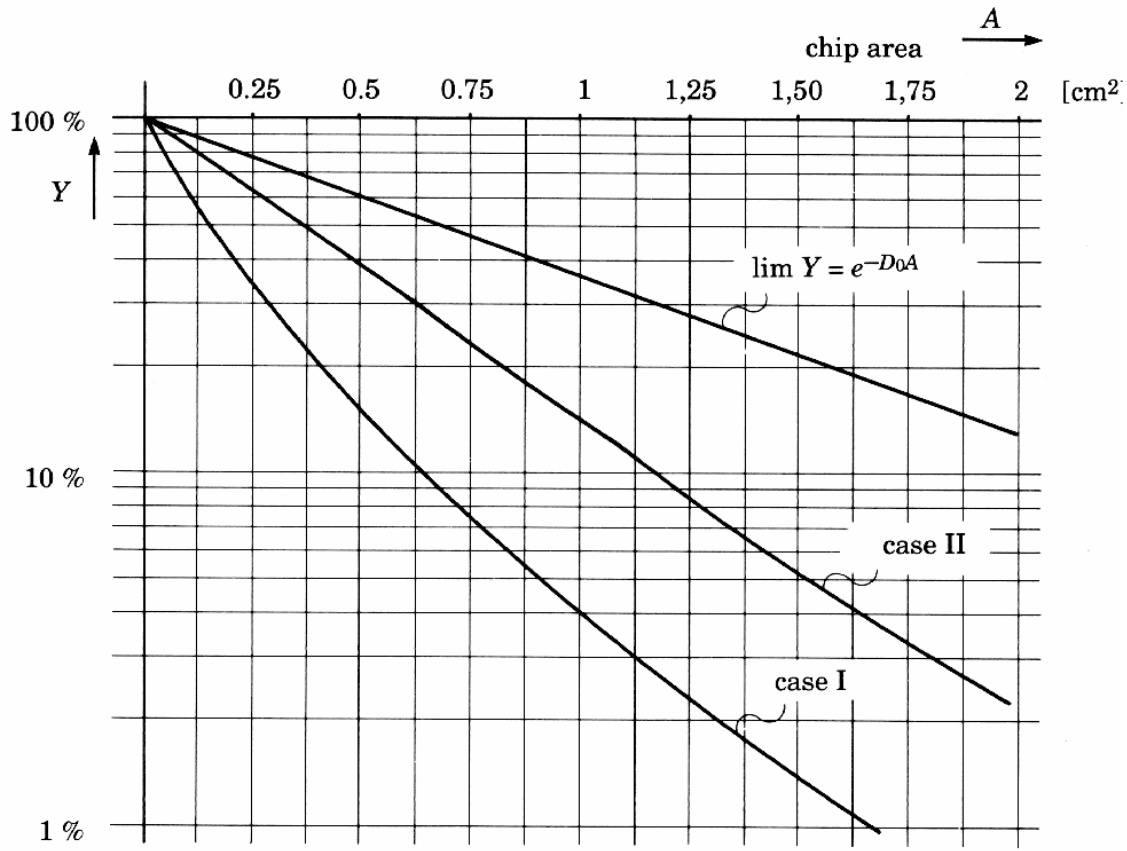
  Thus, they are highly sensitive to noise (cross talk, supply noise)
- High density and large die sizes cause yield problems (DRAM chips are the largest around)

- Yield = 100*(# of good chips/wafer)/(# of chips/wafer)

$$Y = [(1 - e^{-AD})/AD]^2$$

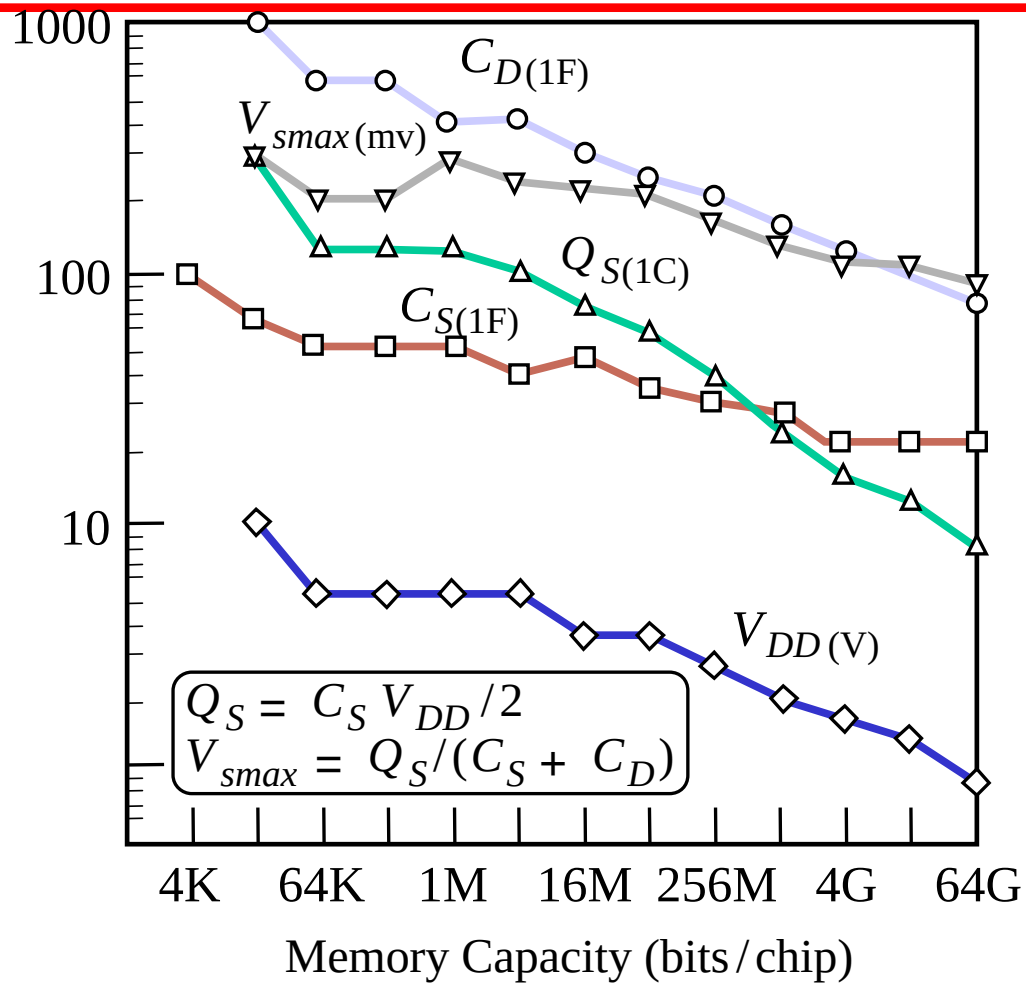
- A = Area, D = Defect Density
- Increase yield using error detection/correction and redundancy

Yield



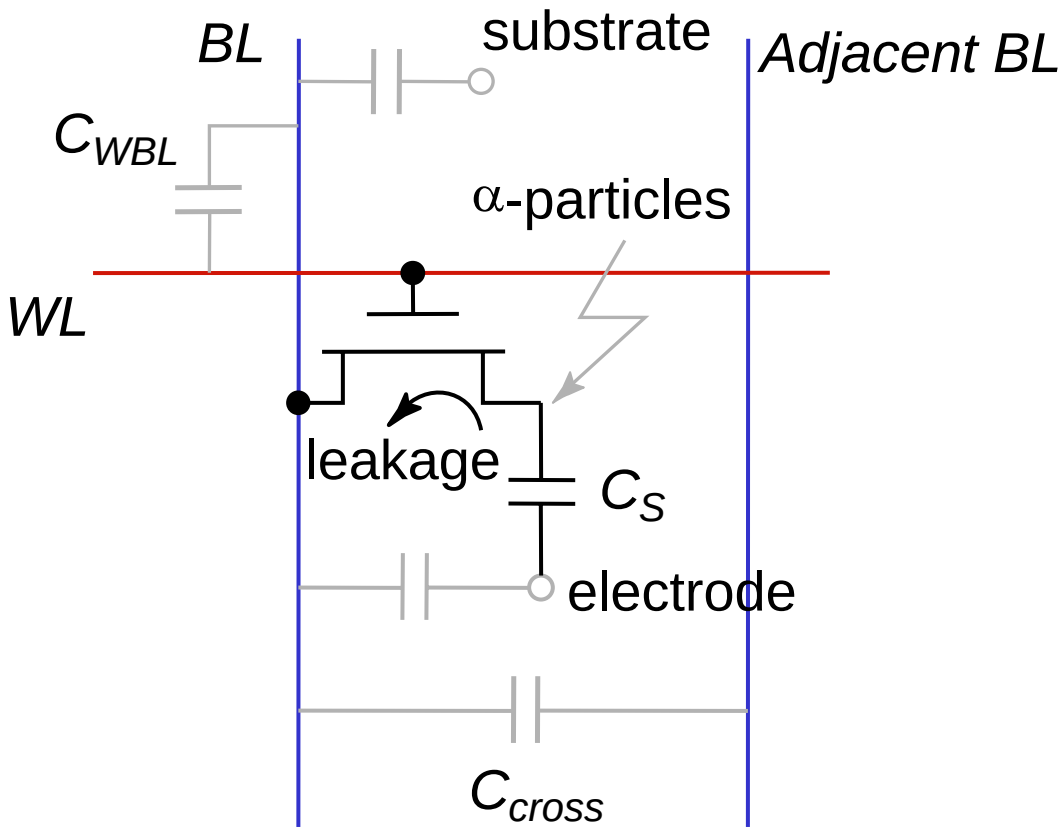
Yield curves at different stages of process maturity (from [Veendrick92])

Sensing Parameters in DRAM

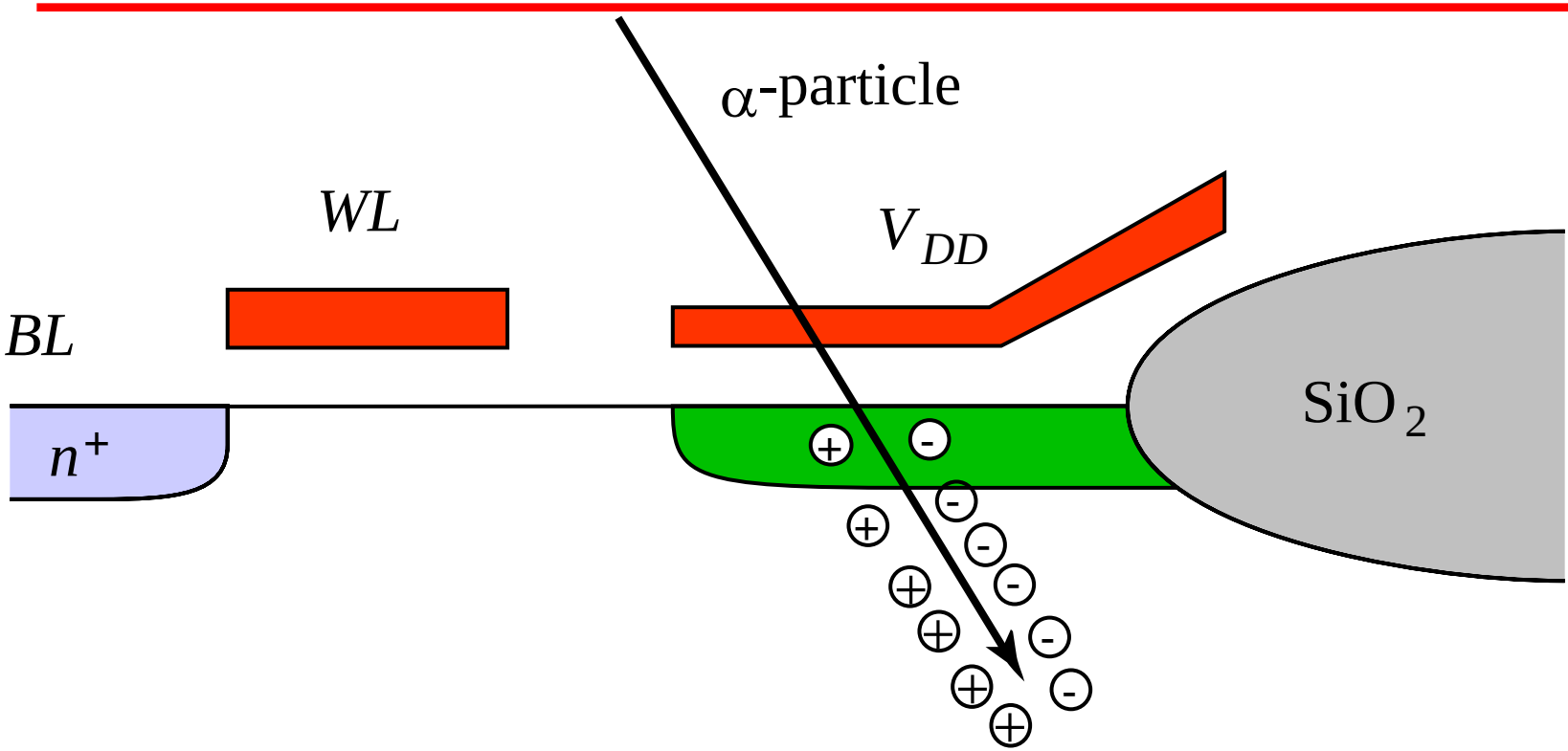


From [Itoh01]

Noise Sources in 1T DRAM



Alpha-particles (or Neutrons)

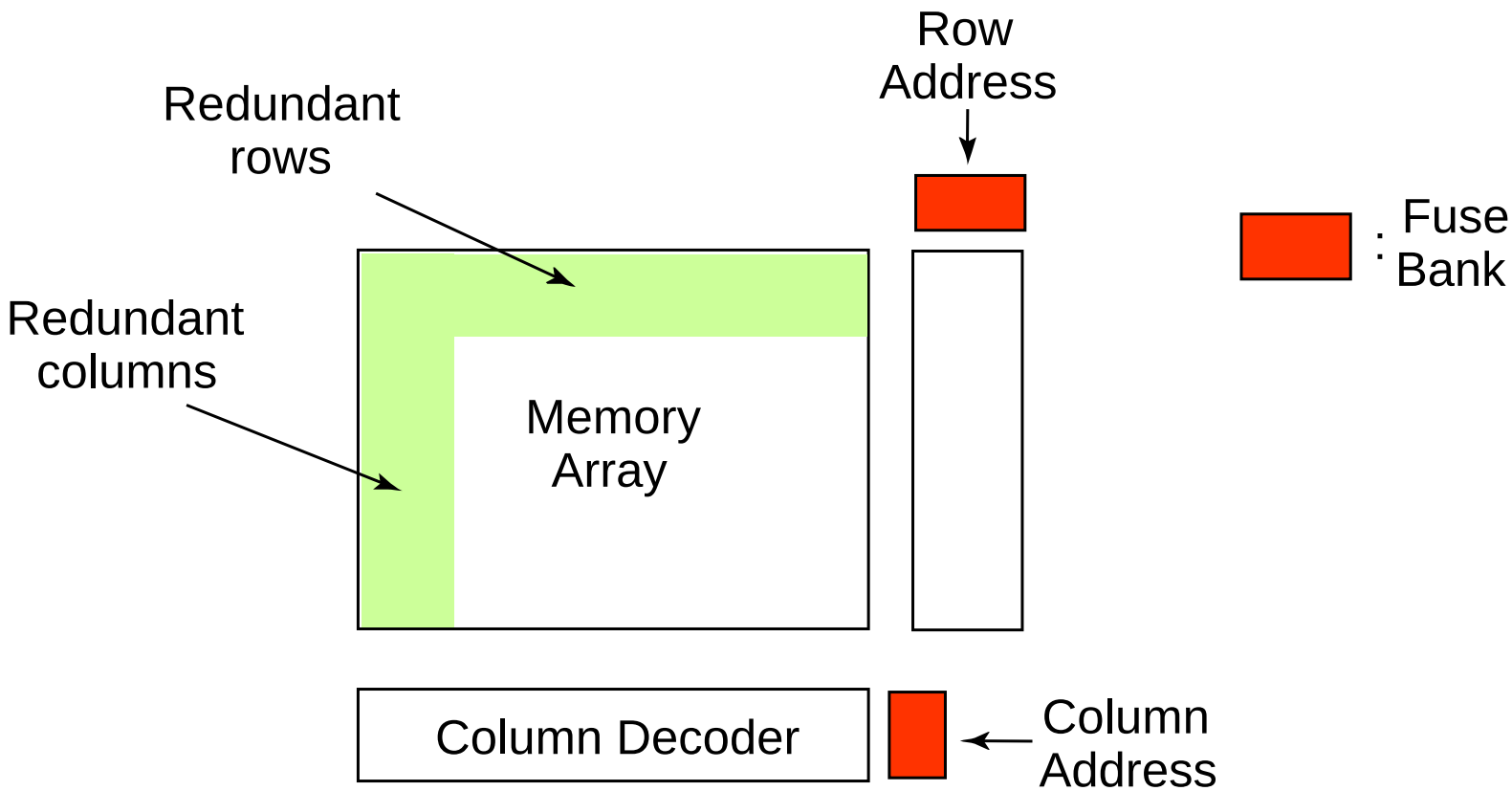


1 Particle ~ 1 Million Carriers

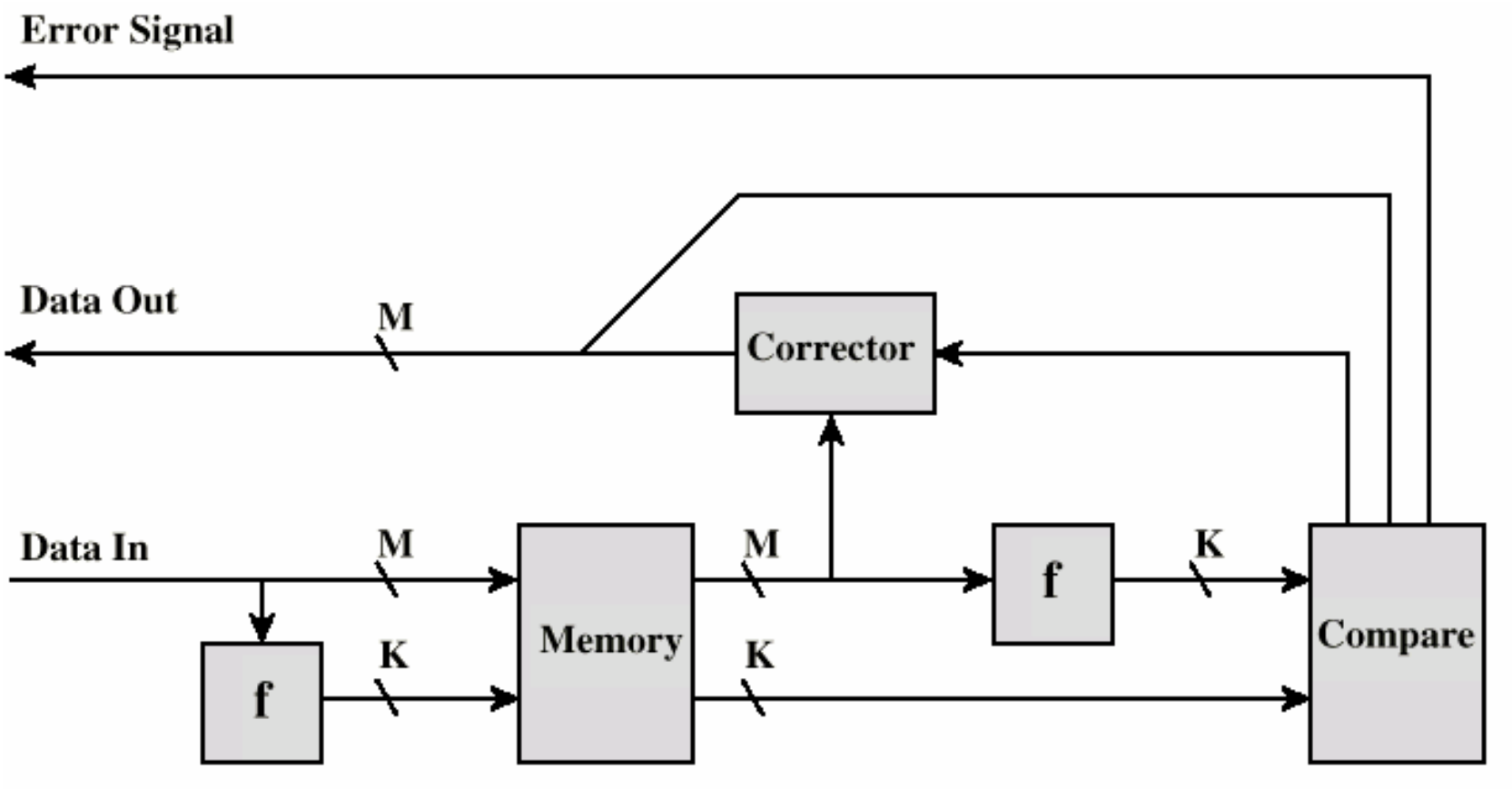
SRAM cell failure mode due to process variation

- Read failures are common
  - '0' node to be read will spike up when access transistor turns ON; if spikes too high, could toggle the held state
    - Depends on relative strengths of various transistors in the 6T cell

Redundancy


Row
Address
Redundant
rows
Redundant
columns
Memory
Array
Fuse
Bank
Column Decoder
Column
Address

Error Correcting Code Function



Hamming based codes

- M data bits
- K parity bits
- Must transmit M+K bits instead of just M (code overhead)
- Ex: 8 bit data (M=8) → 11000100

Bit positions:

1 2 3 4 5 6 7 8 9 10 11 12

P_1 P_2 1 P_4 1 0 0 P_8 0 1 0 0

Compute parity bits as XORs of various bit combinations

P_1 = XOR of bits 3,5,7,9,11 = 0

P_2 = XOR of bits 3,6,7,10,11 = 0

P_4 = XOR of 5,6,7,12 = 1

P_8 = XOR of 9,10,11,12 = 1

Checking of data

- 4 check bits (XOR parity bits with their original data):

C_1 = XOR of 1,3,5,7,9,11

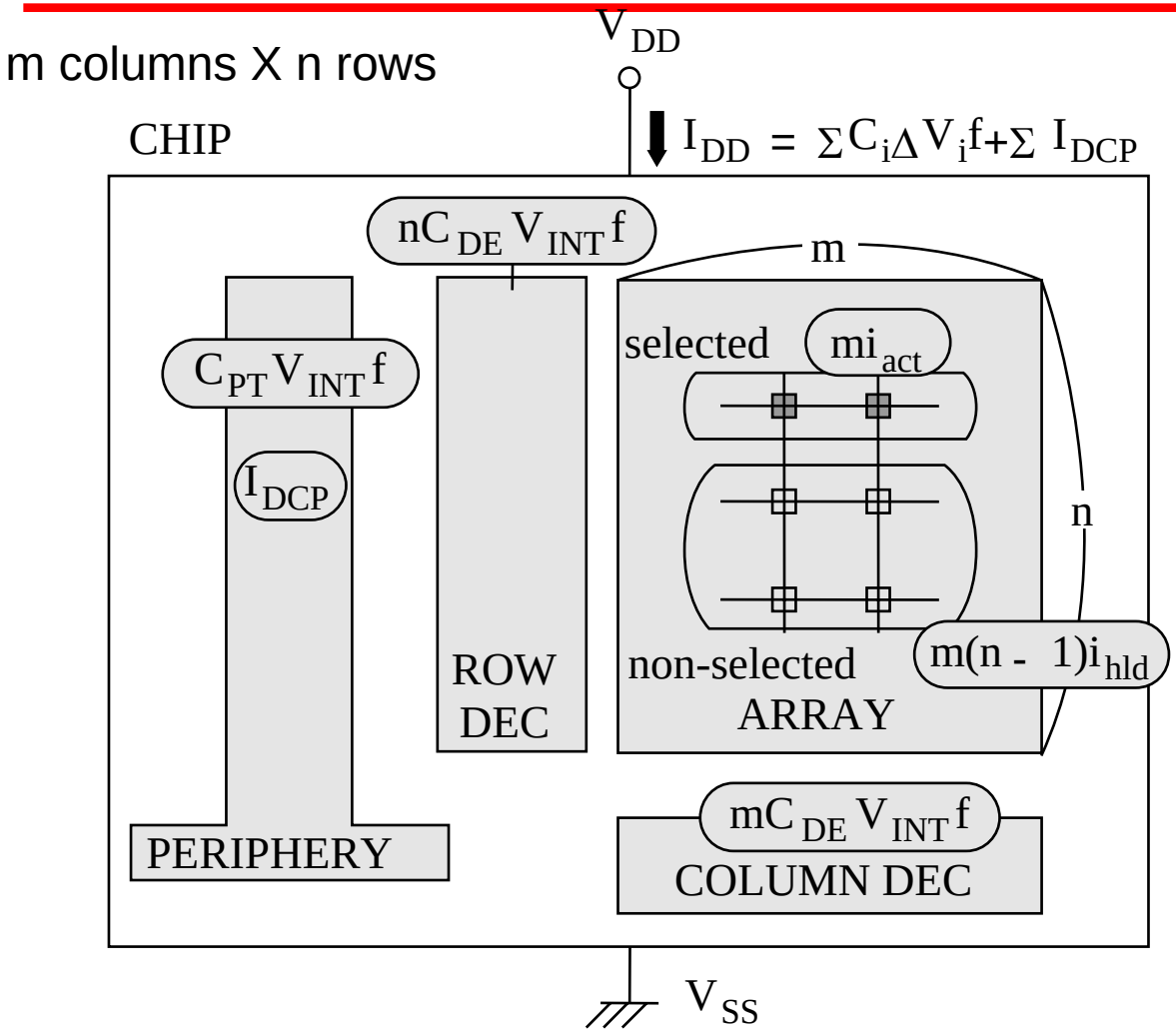
C_2 = XOR of 2,3,6,7,10,11

C_4 = XOR of 4,5,6,7,12

C_8 = XOR of 8,9,10,11,12

- If $C_1=C_2=C_4=C_8=0$, no errors
- Otherwise, bit C is in error ($C_8C_4C_2C_1$ encoded)
- Ex: if bit 5 is flipped erroneously in the memory array, C_4 and C_1 would be computed as high:
  - $C_8C_4C_2C_1$ = 0101 → bit location 5
  - Corrector circuit would then flip the state of bit 5 back to correct value

Sources of Power Dissipation in Memories



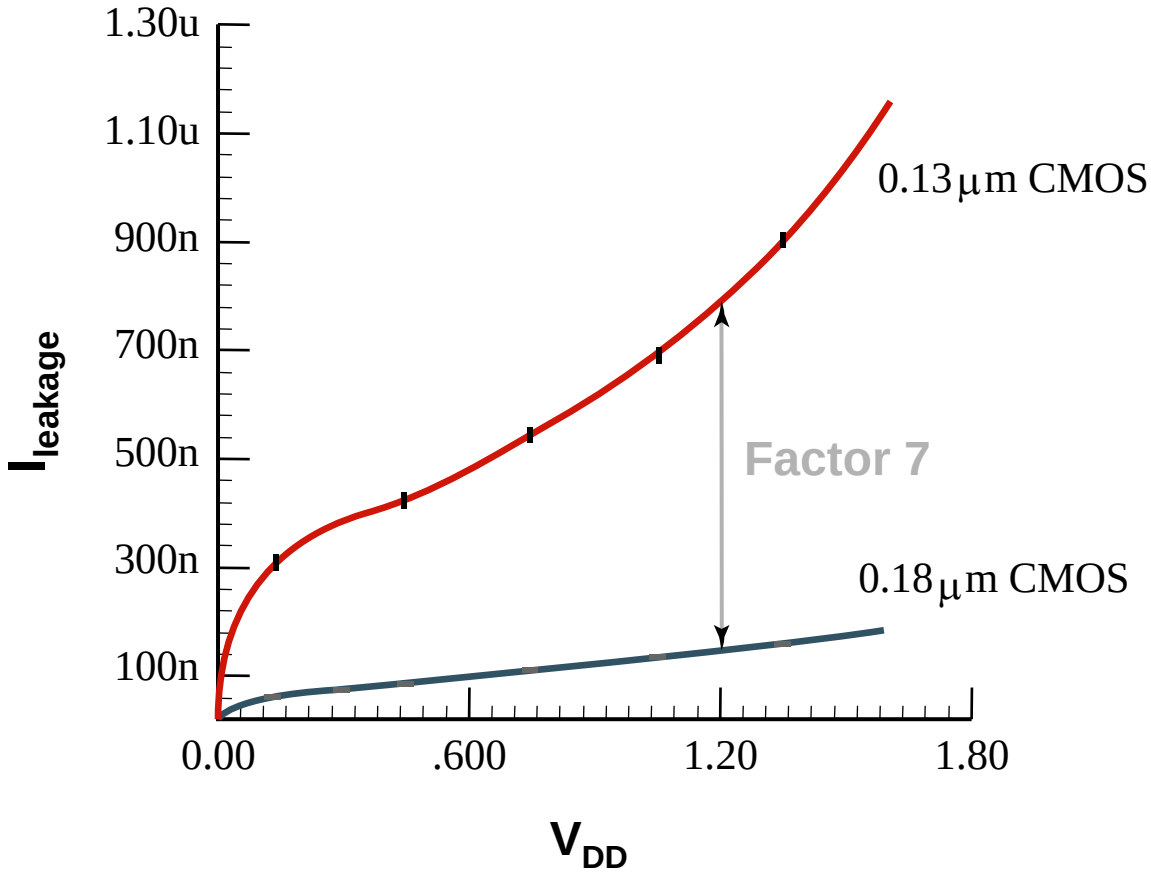
Leakage is a *major* issue in memories

Lots of transistor width, most of it isn't switching in a given access

Memory partitioning very important – can turn off unused blocks after decoding address

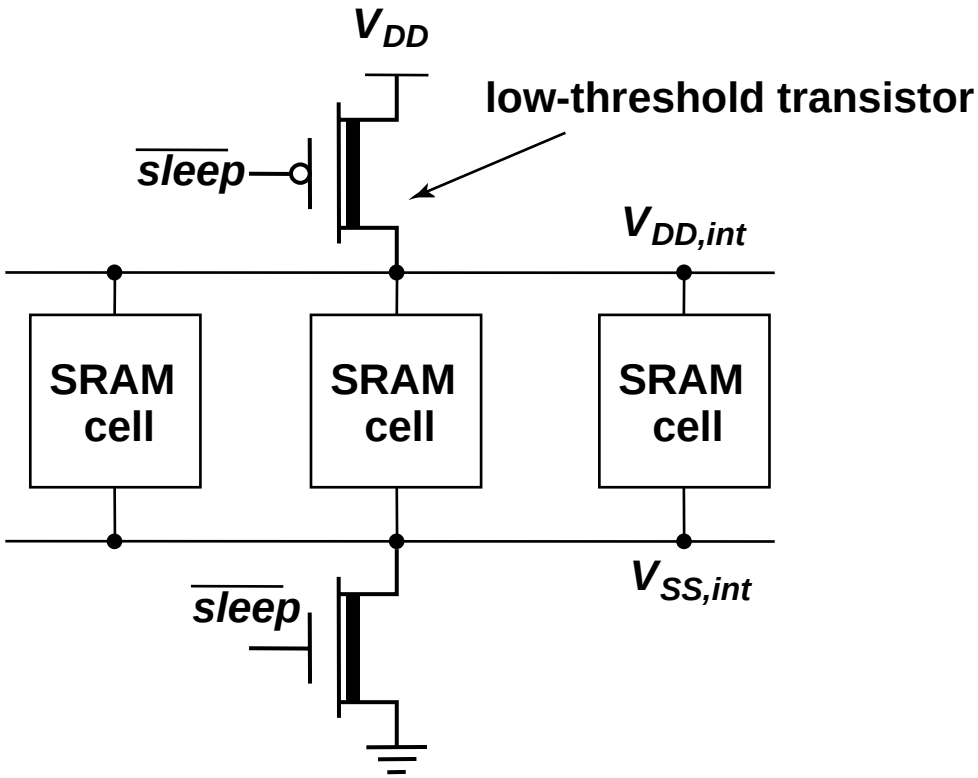
From [Itoh00]

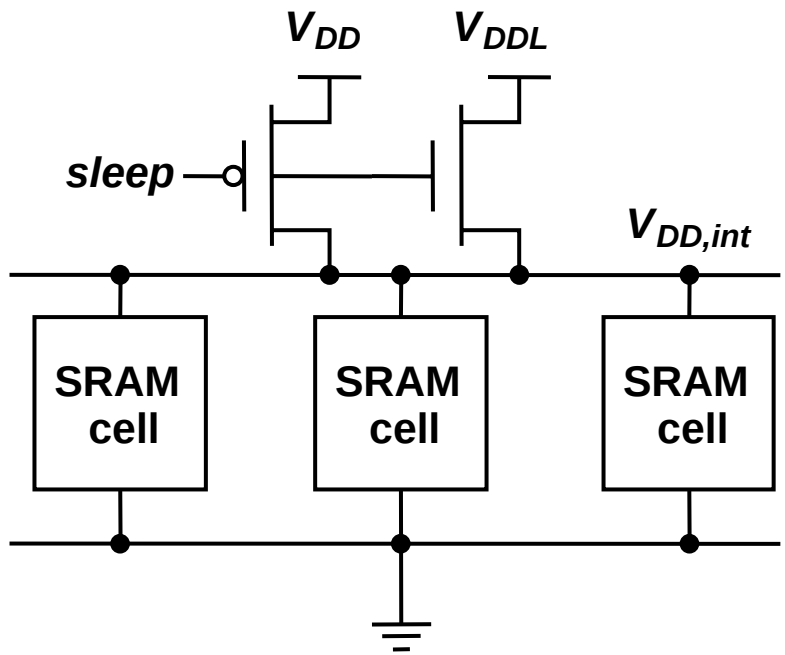
Data Retention in SRAM



SRAM leakage increases with technology scaling

Suppressing Leakage in SRAM





Inserting Extra Resistance

Reducing the supply voltage

Additional Techniques

- Turn off unused blocks in an array
- Apply body bias to sections of the array not being used to increase Vth and reduce leakage
- Just use a higher Vth in the memory cells (speed penalty amortized over entire access time)

Conclusions

- Memory reliability is extremely critical
  - Build in redundancy, error detection (parity) and correction
- Leakage power is even more significant in memories than in logic
  - Due to low switching activity in most of the array, huge amount of transistor width