MEMORIES

Memory Outline

- Overview
  - Array Memory View 1
  - Array Memory View 2: Types
    - Read/Write
    - ROM
    - NVRWM
  - Array Memory View 3
    - Combinational
    - Technically sequential
    - Sequential
- 6T SRAM cell
- SRAM circuits and layout
- Multi-port SRAM
- DRAM
- ROM circuits and layout
- Array Memory View 4: memories for ASICs and FPGAs
  - On-chip macros
    - ASICs
    - FPGAs
  - On-chip standard cell
    - ASICs
    - FPGAs
  - Off-chip
- Synthesized standard cell memories
  - Verilog
  - Timing
  - ROM: standard cell verilog
  - ROM: FPGA block RAM

Array Memory View 1: A Component of Digital Systems

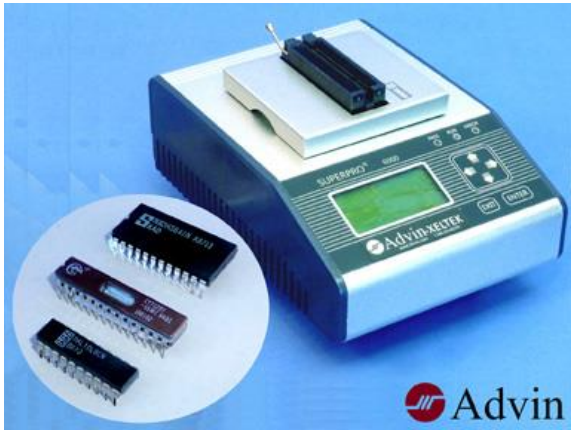
- Three primary components of digital systems
  - Datapath (does the work)
  - Control (manager)
  - **Memory (storage)**
    - "Single bit" ("foreground")
      - Clockless latches — e.g., SR latch
      - Clocked transparent latches — e.g., D latch
      - Clocked edge-triggered flip flops — e.g., D FF
    - **"Array" memories** ("background")

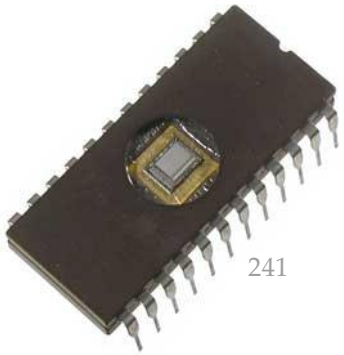
Memories

- Use in general digital processors
  - Instructions
  - Data
- Usage is more widespread in DSP, multimedia, embedded processors
  - Buffering input/intermediate/output data (e.g., rate matching)
  - Storing fixed numbers (e.g., coefficients)
  - Often relatively small (e.g., 8-64-256 words) and numerous (dozens are not unusual)
- Key design goal: density, especially for the memory cells. This means fitting the largest amount of memory storage into a certain amount of chip area

Array Memory View 2: Types of Memories

1. Read-write memories
   - SRAM: Static random access memory
     - Data is stored as the state of a bistable circuit typically using "back-to-back" inverters
     - State is retained without refresh as long as power is supplied
   - DRAM: Dynamic random access memory
     - Data is stored as a charge on a capacitor
     - State leaks away, refresh is required
2. ROM: Read-only memory, non-volatile
   - Basic ROM – mask programmed at design time
   - PROM: Programmable read-only memory; typically programmed at manufacture time by a "PROM burner"
     - Using fuse or anti-fuse circuits
   - Synthesized from standard cells
3. NVRWM: Non-volatile read-write memory
   - EPROM: Erasable ROM, erasable with UV light
   - Flash: ROM at low voltages, writable at high voltages



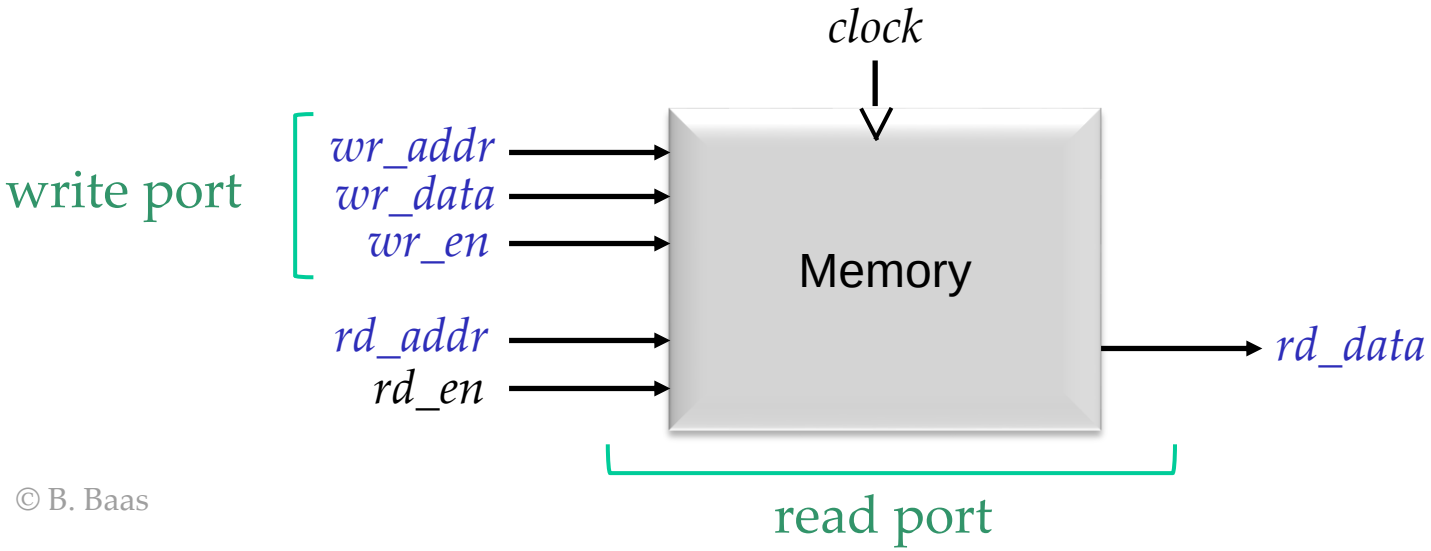


Memory View 3: Memory Logical Categories

- Combinational (output depends on present inputs only)
  - ROM: read-only memory
  - May be straight-through truly-combinational, or registered
- Feels like Combinational but technically Sequential
  - PROM: programmable read-only memory
  - EPROM: ROM, but erasable with UV light
- Sequential (output depends on present and past inputs)
  - SRAM: static memory
  - DRAM: dynamic memory
  - Flash: ROM at low voltages, writable at high voltages

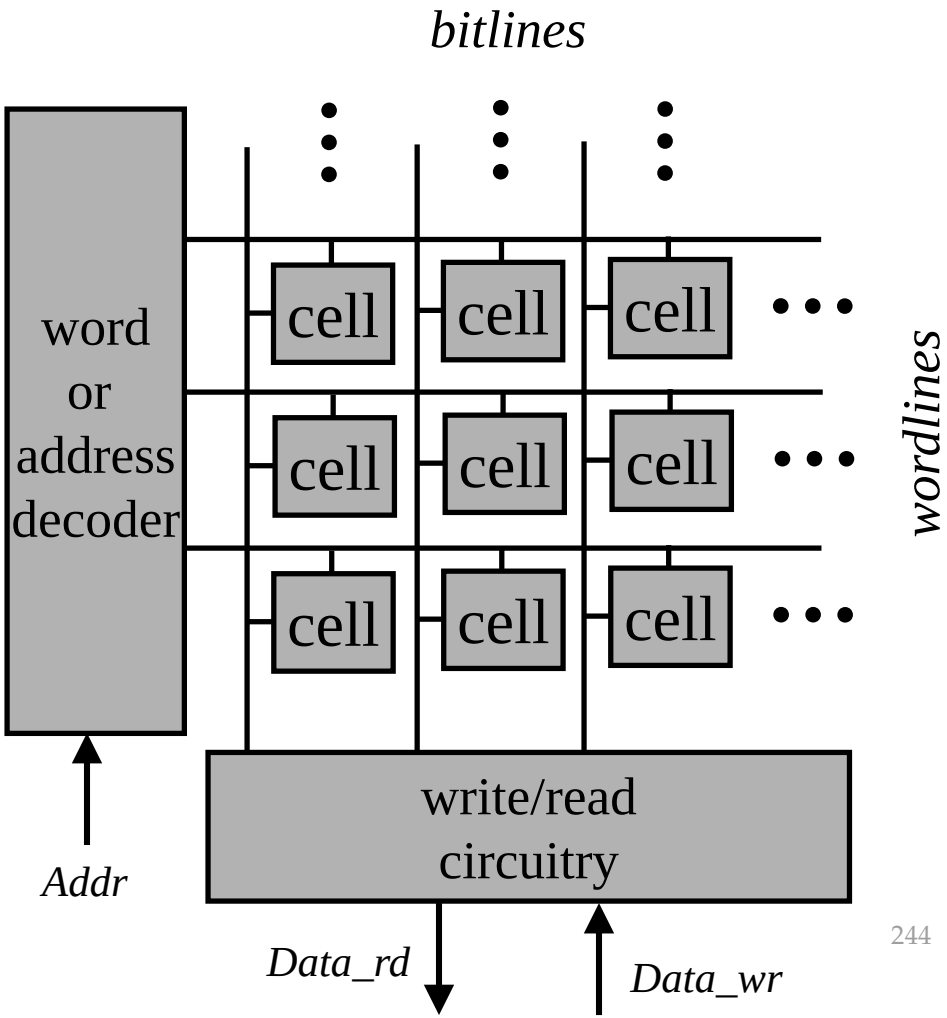
Basic Memory Inputs and Outputs

- The basic memory structure includes a write port and a read port as shown in the figure
  - Clocked or Synchronous memories include a *clock* input
  - A read-enable input (*rd_en*) is not needed for functionality but is often included to enable reduced power dissipation when read operations are not needed



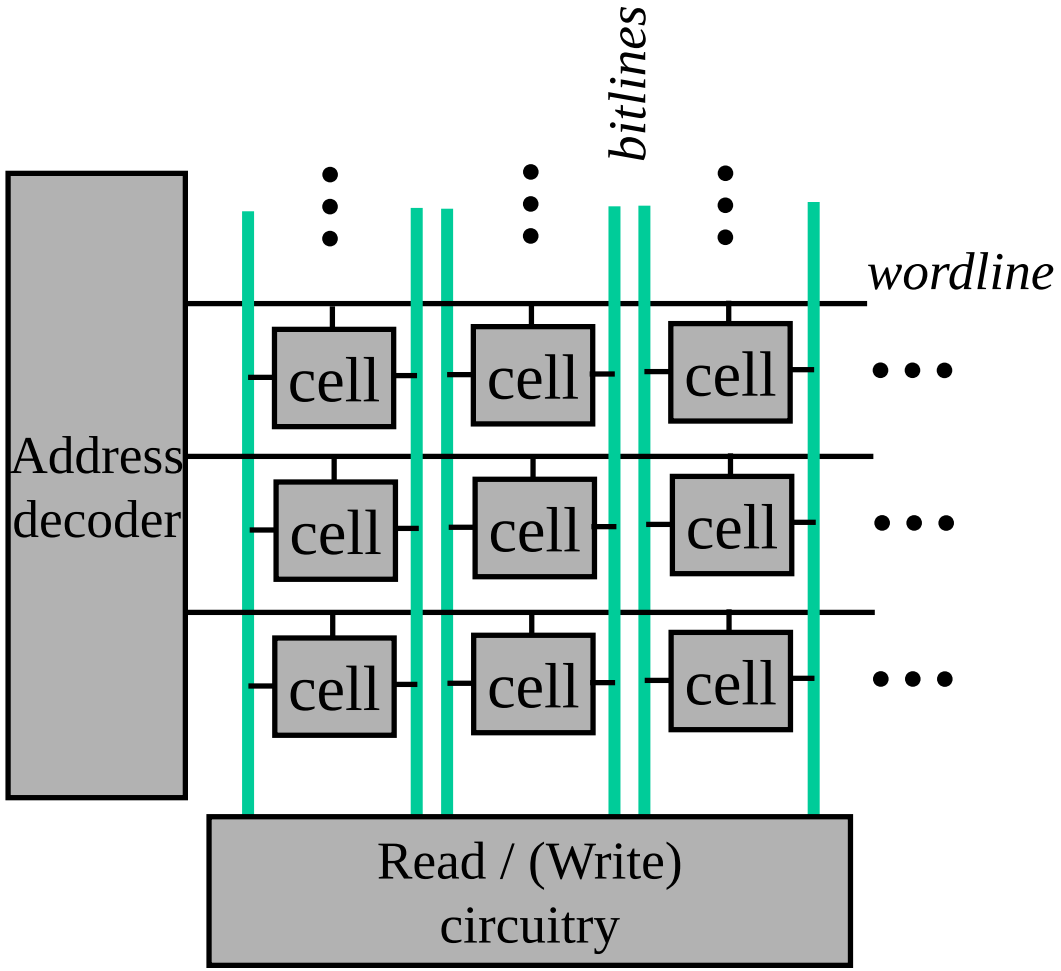
Memories

- Memories generally contain several components:
  - Array of cells
  - Address decoder
  - Write circuitry
  - Read circuitry (sense amplifiers)
  - *wordlines*
  - *bitlines*
- Interface signals
  - *Address* (one for each port)
  - *Data* (one for each port)
  - *Enable_write*
  - *Enable_read* (likely)
  - *Clock* (sometimes)

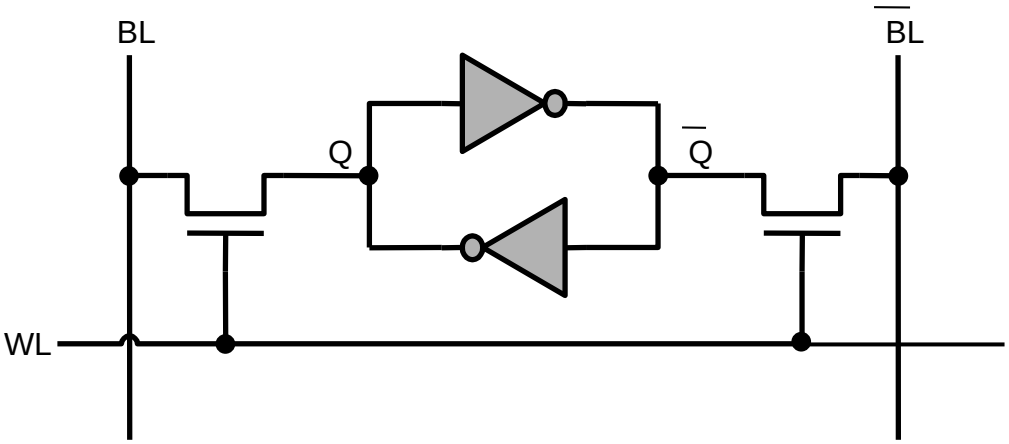


Memories—Differential bitlines

- Differential bitlines (*bitline* and *bitline_*) require more area but dramatically increase robustness and speed
  - Much smaller voltage differences can be detected
  - Much more noise can be tolerated



Six-Transistor (6T) SRAM Cell



- Cross-coupled inverters: a bistable element (two stable states)
- Density is critically important in memories
  - Single NMOS used for reading/writing
  - A lot of effort spent packing transistors and even pushing process design rules just for the 6T memory cell—the area of a 6T cell is typically one of the top critical parameters of a fabrication technology!

62nd International Electron Devices Meeting

IEEE International Electron Devices Meeting (IEDM) is the world's preeminent forum for reporting technological breakthroughs in the areas of semiconductor and electronic device technology, design, manufacturing, physics, and modeling. IEDM is the flagship conference for nanometer-scale CMOS transistor technology, advanced memory, displays, sensors, MEMS devices, novel quantum and nano-scale devices and phenomenology, optoelectronics, devices for power and energy harvesting, high-speed devices, as well as process technology and device modeling and simulation.

3:40 PM

2.6 A 7nm CMOS Platform Technology Featuring 4th Generation FinFET Transistors with a 0.027um2 High Density 6-T SRAM cell for Mobile SoC Applications (Late News), S.-Y. Wu, C.Y. Lin, M.C. Chiang, J.J. Liaw, J.Y. Cheng, S.H. Yang, C.H. Tsai, P.N. Chen, T. Miyashita, C.H. Chang, V.S. Chang, K.H. Pan, J.H. Chen, Y.S. Mor, K.T. Lai, C.S. Liang, H.F. Chen, S.Y. Chang, C.J. Lin, C.H. Hsieh, R.F. Tsui, C.H. Yao, C.C. Chen, R. Chen, C.H. Lee, H.J. Lin, C.W. Chang, K.W. Chen, M.H. Tsai, K.S. Chen, Y. Ku, and S.M. Jang, TSMC

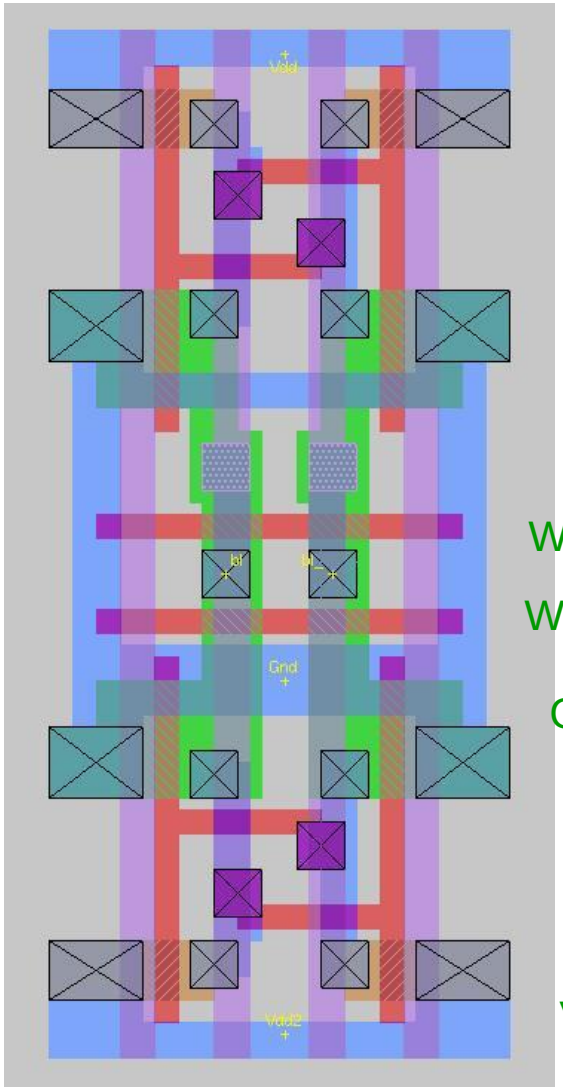
4:05 PM

2.7 A 7nm FinFET Technology Featuring EUV Patterning and Dual Strained High Mobility Channels (Late News), R. Xie, P. Montanini*, K. Akarvardar, N. Tripathi, B. Haran*, S. Johnson, T. Hook*, B. Hamieh*, D. Corliss*, J. Wang*, X. Miao*, J. Sporre*, J. Fronheiser, N. Loubet*, M. G. Sung, S. Sieg*, S. Mochizuki*, C. Prindle, S.-C. Seo*, A. Greene*, J. Shearer*, A. Labonte, S. C. Fan*, L. Liebmann, R. Chao*, A. Arceo*, K. Chung*, K. Y. Cheon**, P. Adusumilli*, H. Amanapu*, Z. Bi*, J. Cha**, H.-C. Chen*, R. Conti*, R. Galatage, O. Gluschenkov*, V. Kamineni, K. Kim**, C. Lee*, F. L. Lie*, Z. Liu*, S. Mehta*, E. Miller*, H. Niimi, C. Niu, C. Park, D.-H. Park**, M. Raymond, B. Sahu, M. Sankarapandian*, S. Siddiqui, R. Southwick*, L. Sun, C. Surisetty*, S. Tsai, S. Whang**, P. Xu*, Y. Xu*, C.-C. Yeh*, P. Zeitzoff, J. Zhang, J. Li*, J. Demarest*, J. Arnold*, D. Canaperi*, D. Dunn*, N. Felix*, H. Jagannathan*, S. Kanakasabapathy*, W. Kleemeier, C. Labelle, M. Mottura, P. Oldiges*, S. Skordas*, T. Standaert*, T. Yamashita*, M. Colburn*, M.-H. Na*, V. Paruchuri*, S. Lian**, R. Divakaruni*, T. Gow*, S. Lee**, A. Knorr, H. Bu*, and M. Khare*, GLOBALFOUNDRIES, *IBM, **Samsung

Retrieved November 17, 2016

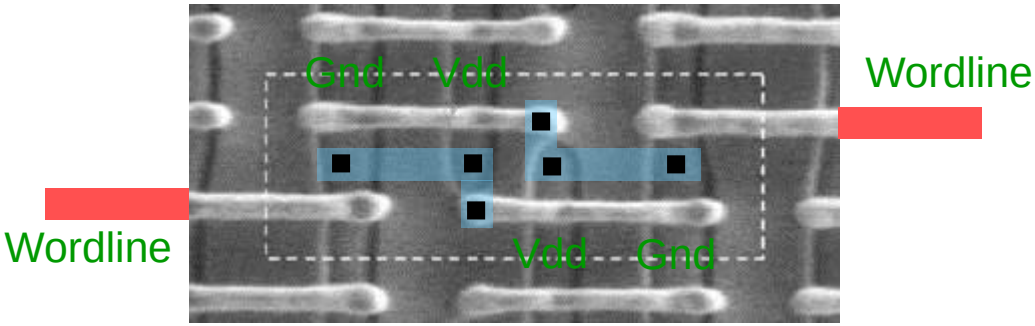
Layout: 6-Transistor (6T) Cell

- Two 6T SRAM cells
- Wordlines horizontal
- Bitlines vertical
- *Vdd* and *Gnd* horizontal

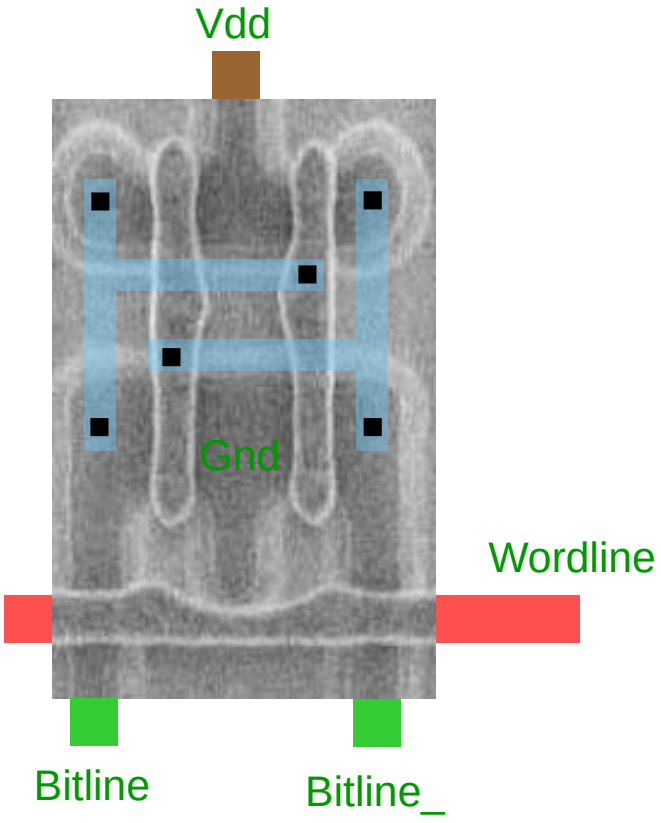


6-Transistor (6T) Cells

- Micro-photographs of fabricated 6T SRAM memory cells
- The devices are partially built with diffusion and polysilicon only—metal interconnect is shown where it might be located



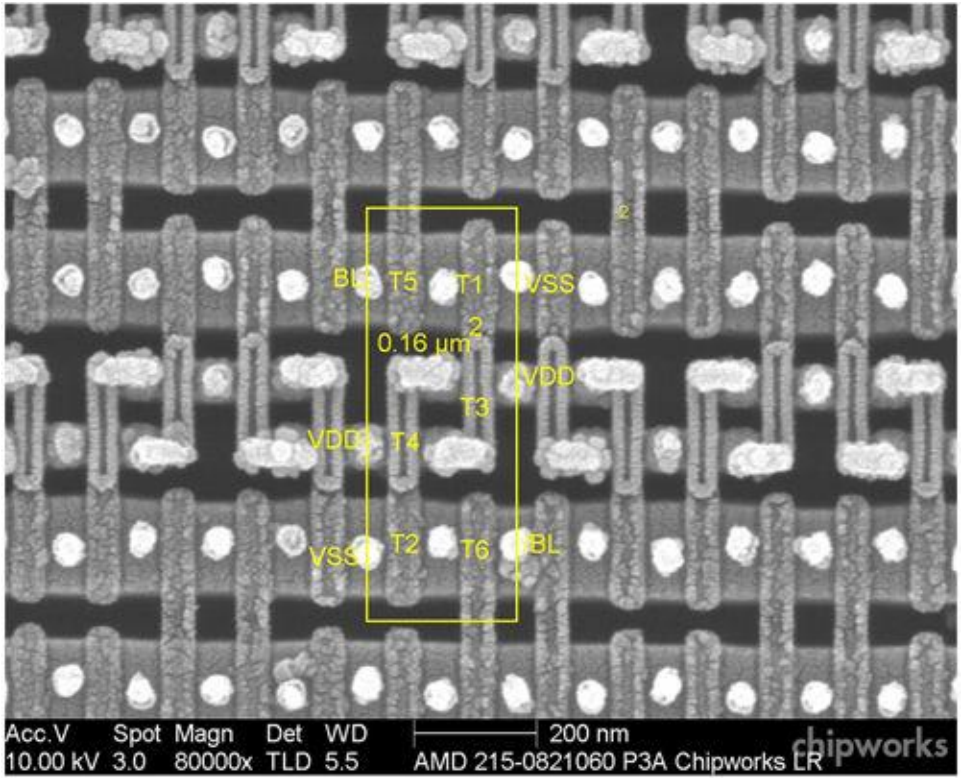
Intel 65 nm 0.57 μm^2 6T cell



[http://www.electroiq.com/articles/sst/print/volume-47/issue-2/departments/tech-news/technology-news.html]

6-Transistor (6T) Cells

- Micro-photograph of 0.16 μm^2 6T SRAM cells from an AMD Radeon graphics processor fabraicated in TSMC's 28 nm HP process

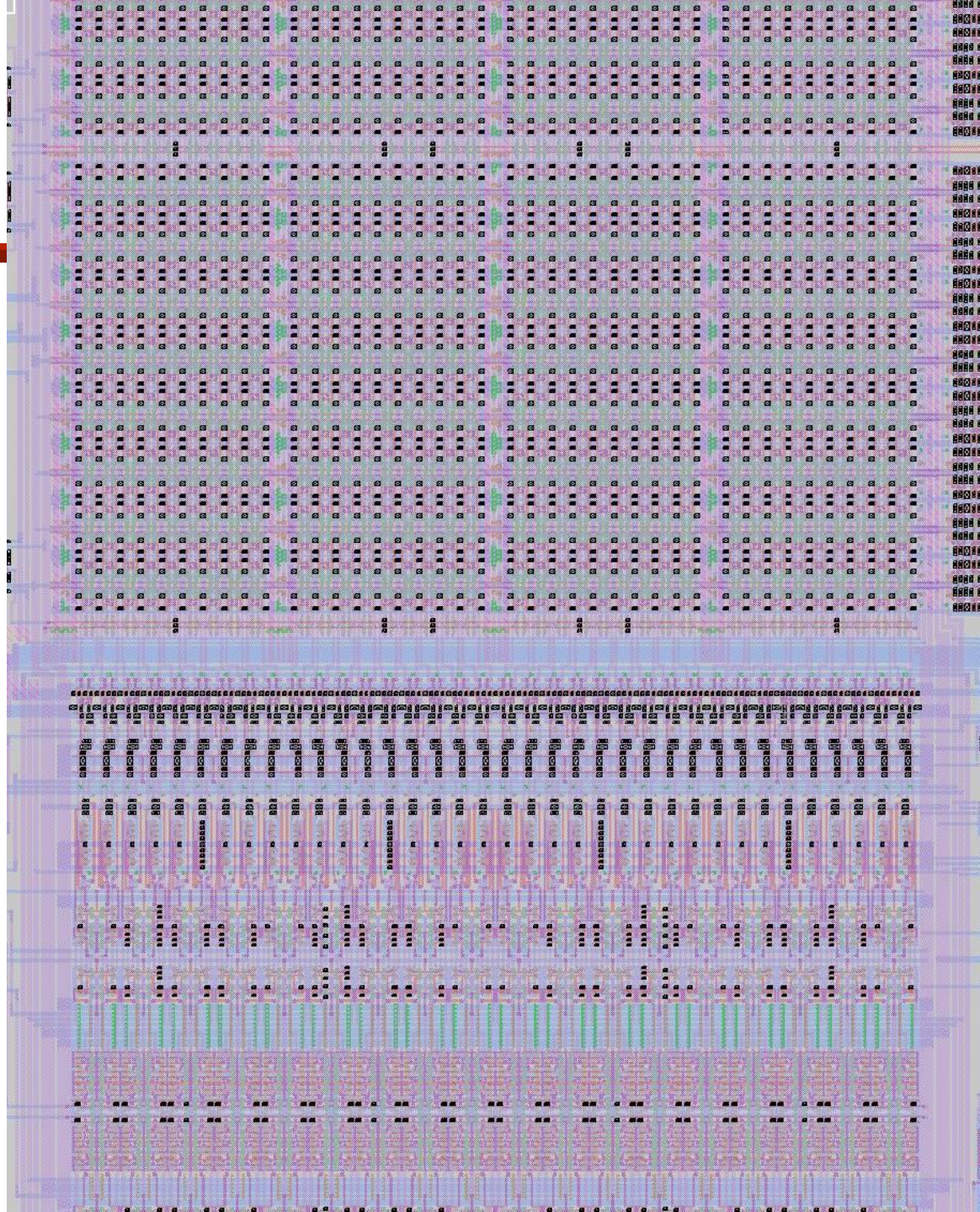


[Chipworks]

Layout: 6T cell array

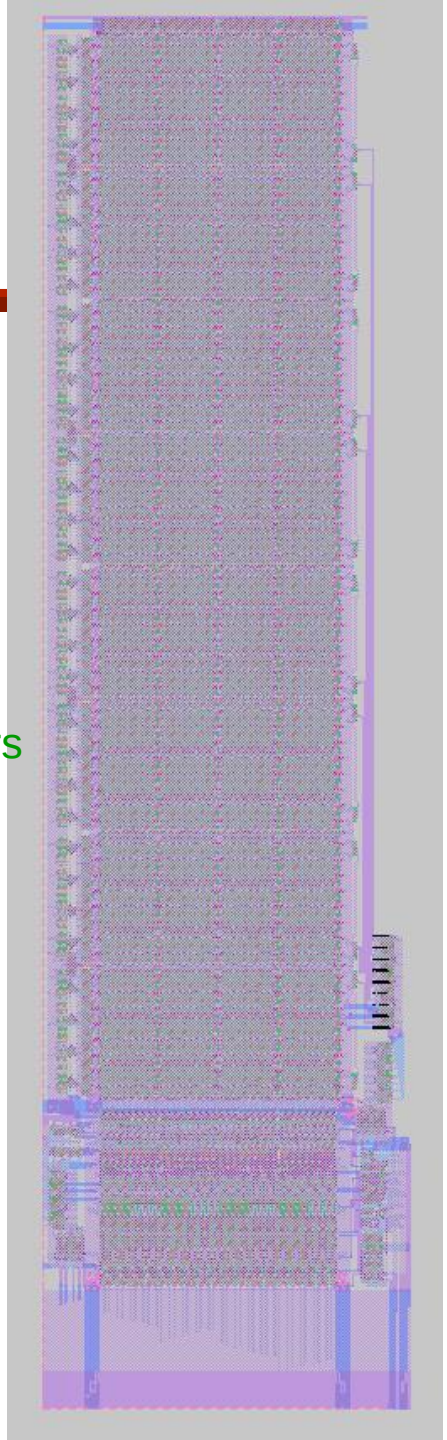
Cell array

Sense amplifiers and write circuits



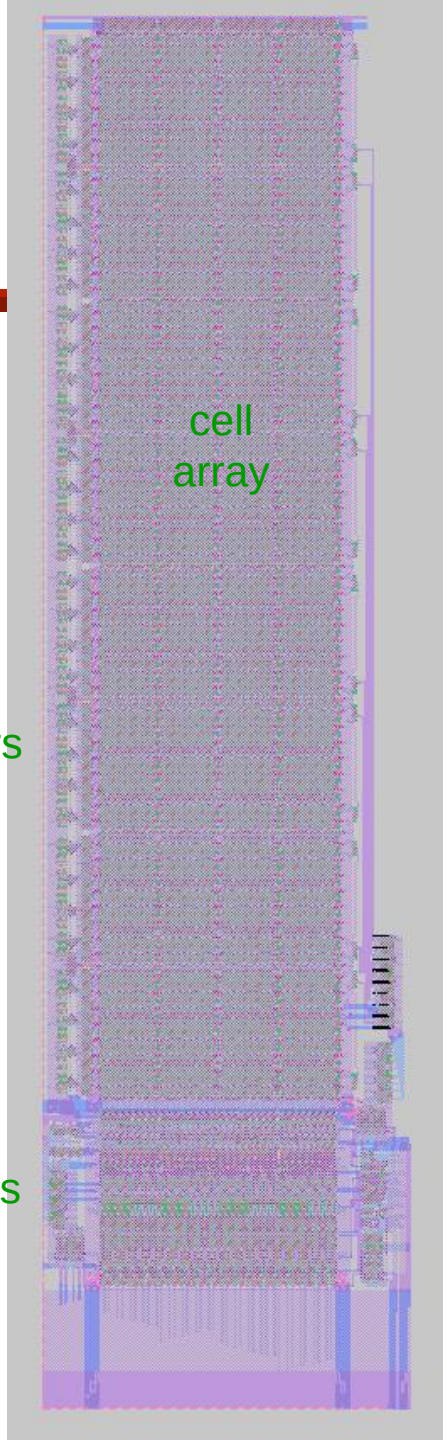
Layout: Memory Array

- 128 words x 36 bits
- Single-ported 6T SRAM
- Low-power hierarchical bitline structure



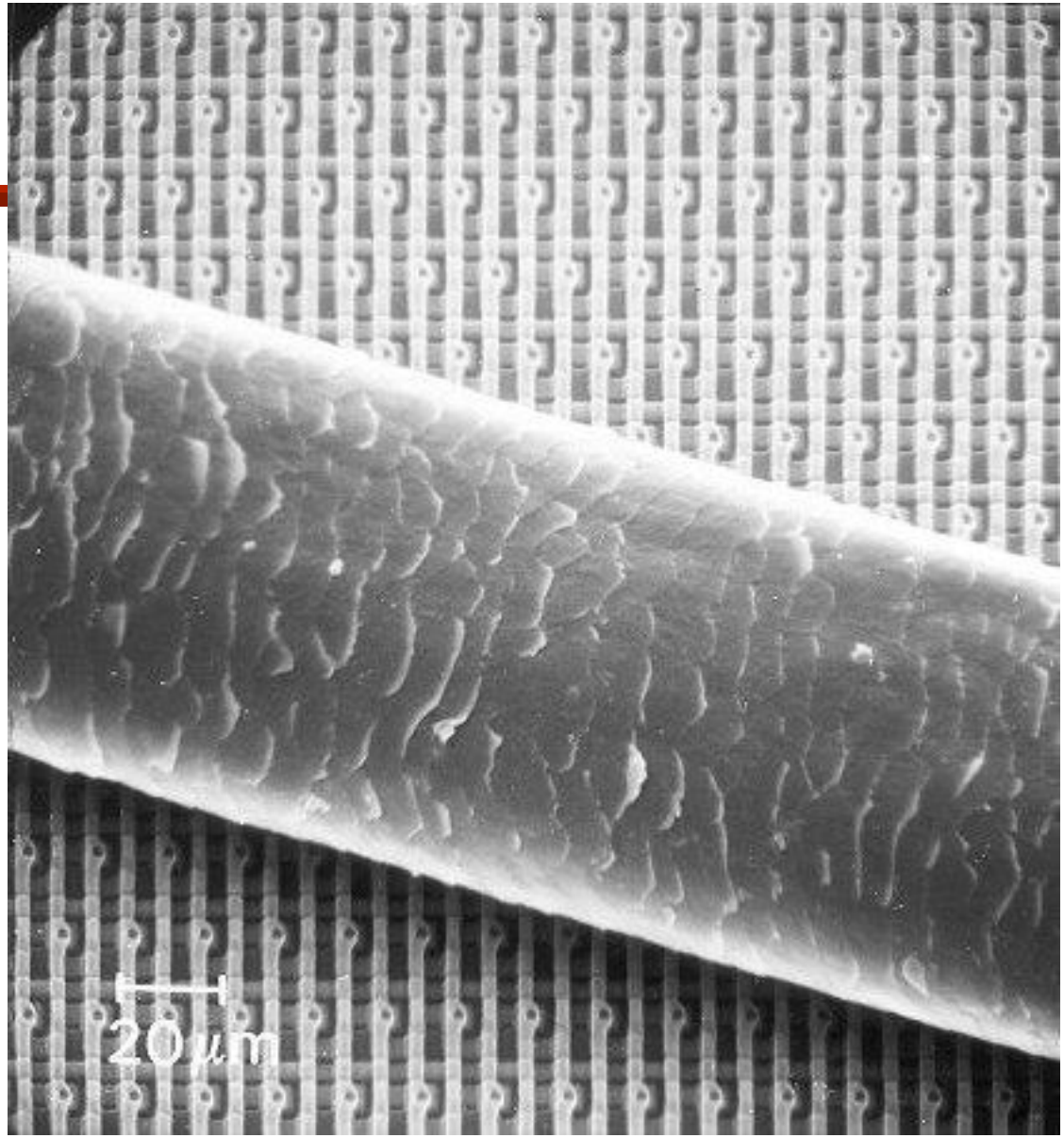
Layout: Memory array

- Layout issues
  - Cell density is critical
  - Decoder design should probably not use full *N*-word fanout
    - Predecoding address bits reduces loading of critical decoder signals
  - Routing of *Vdd* and *Gnd* to supply adequate current requires planning



Memory Array

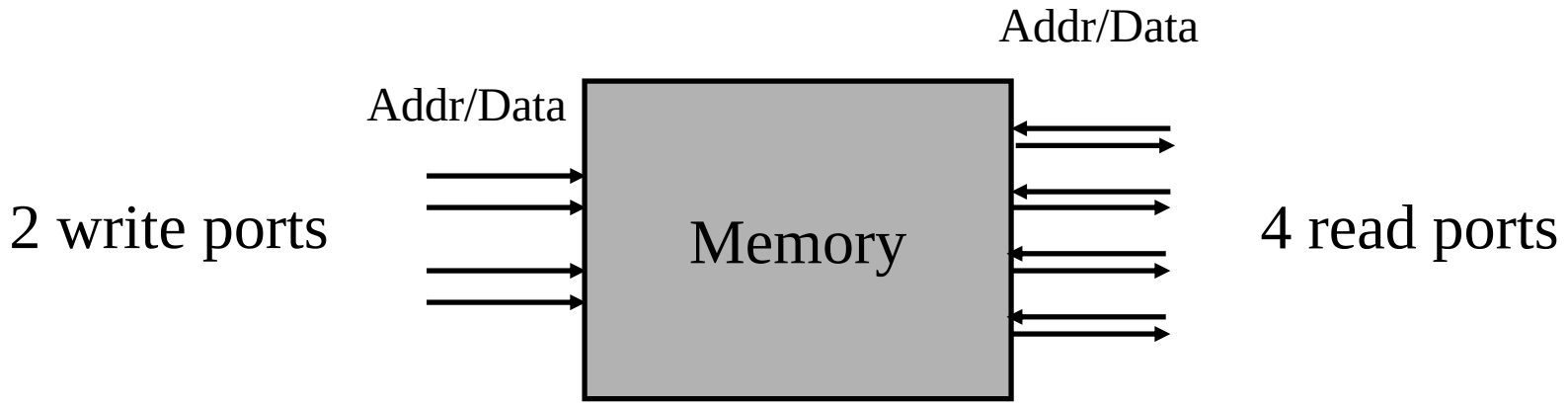
- Human hair on a 256 Kbit memory chip



Source: Helmut Föll

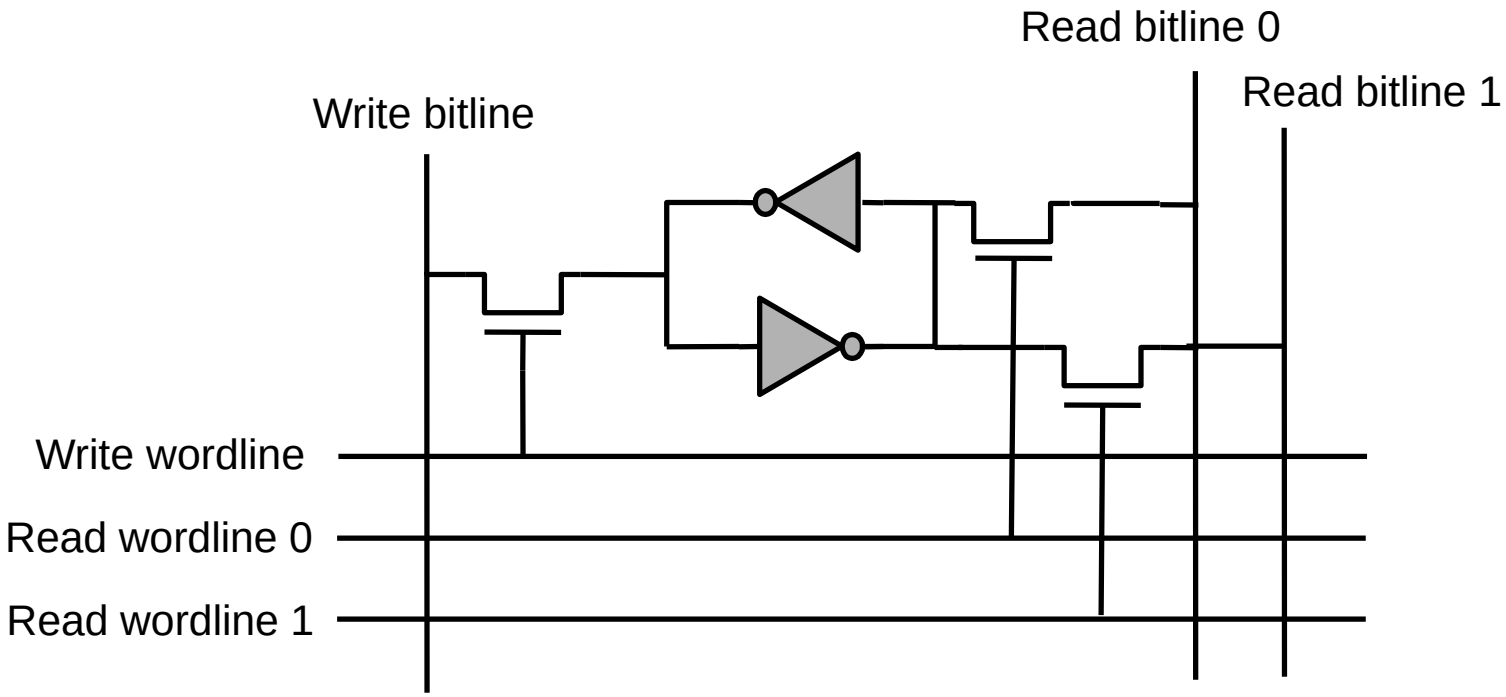
Multi-ported SRAM

- Frequently used in register files
  - Classic RISC computers have 1 write and 2 read ports
  - Modern multiple-instruction-issue computers can have many ports (22 (12 Rd, 10 Wr) in Itanium [ISSCC 05])
- More commonly use single-ended (non-differential) bitlines



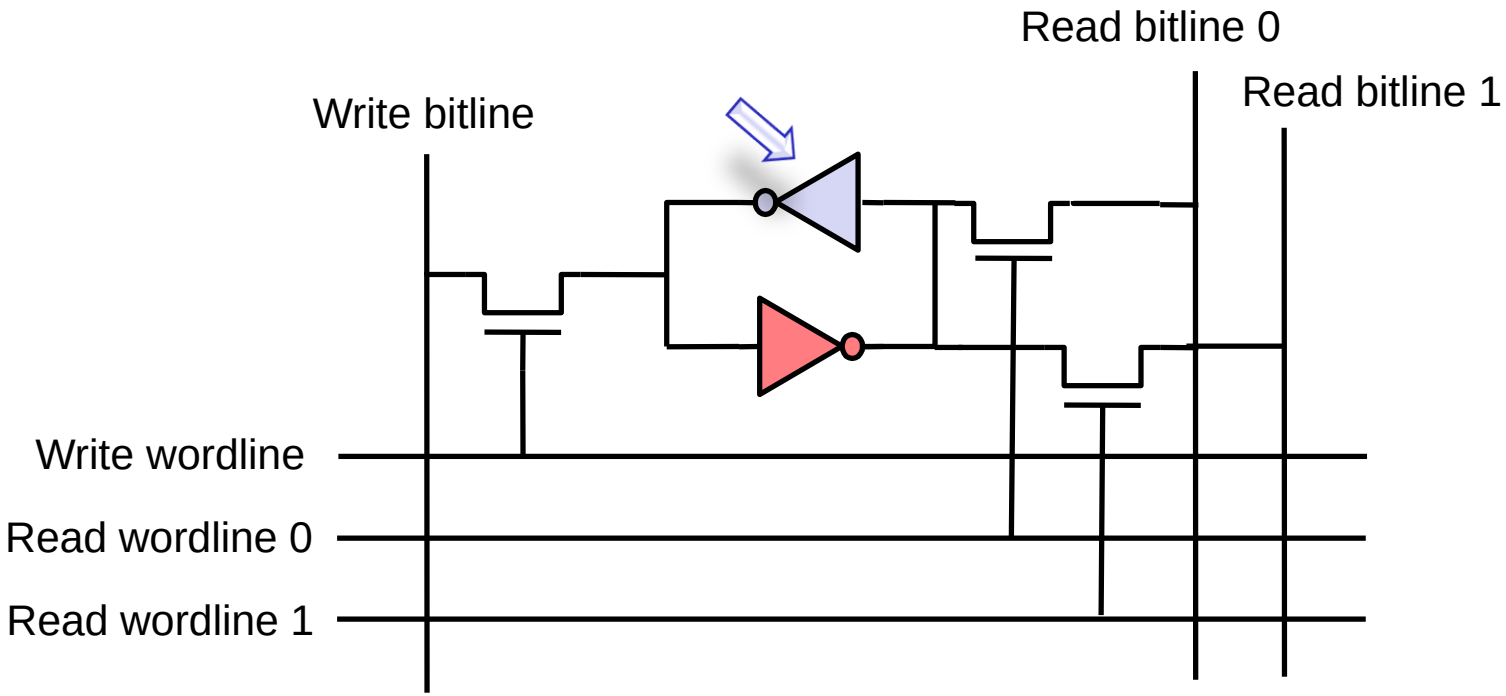
Multi-ported SRAM

- Example: one write port, two read ports



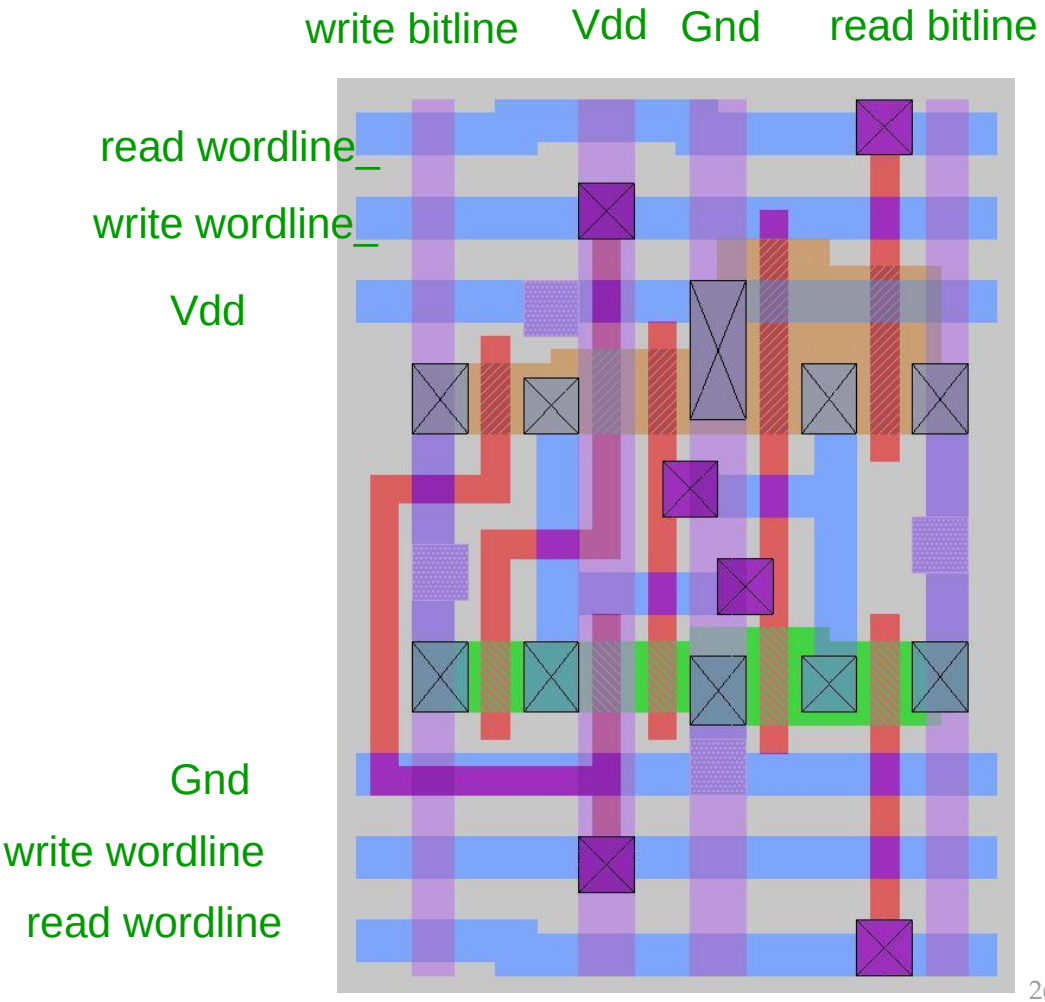
Multi-ported SRAM

- Example: one write port, two read ports
- If the feedback inverter is not tri-statable during writes, it must be made weak to permit writes, especially during a high value (*Vdd*) on the bitline



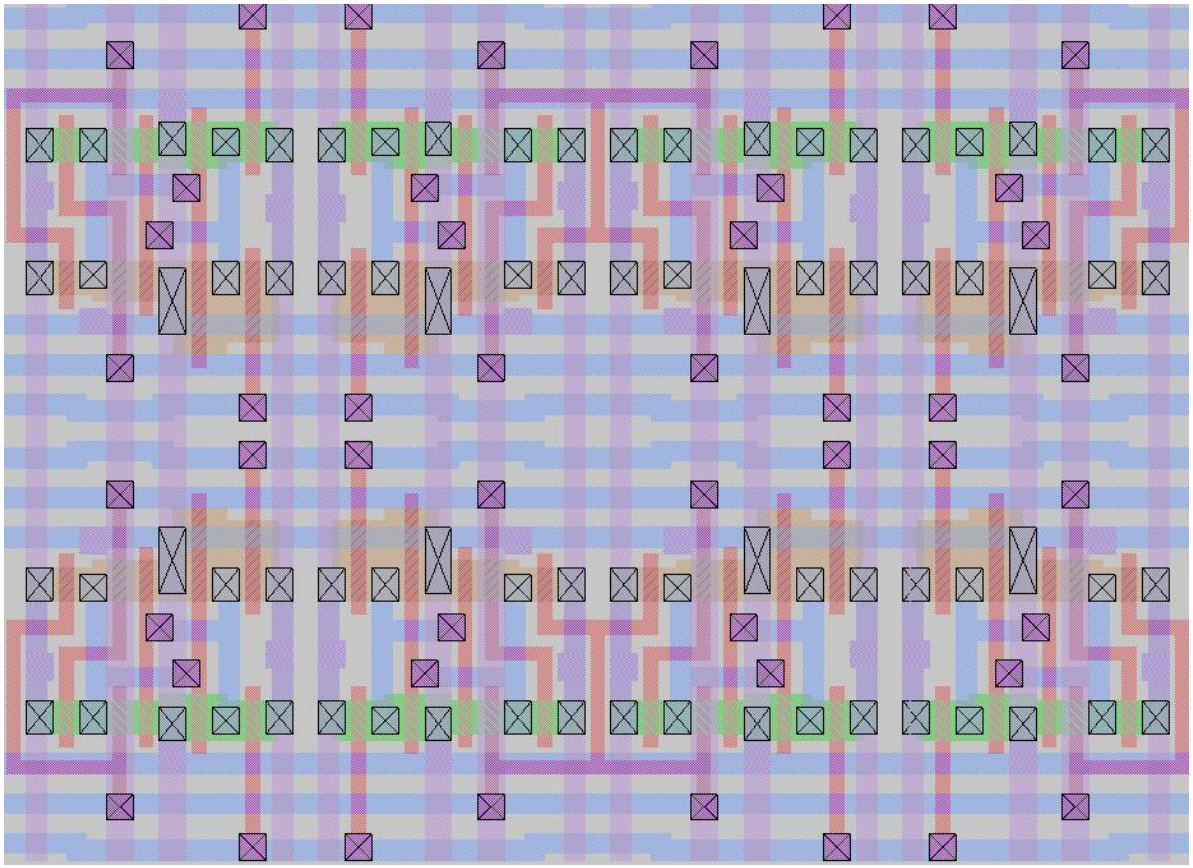
Layout: Dual-ported memory cell

- 10T one-read, one-write port
- Operates at very low supply voltages



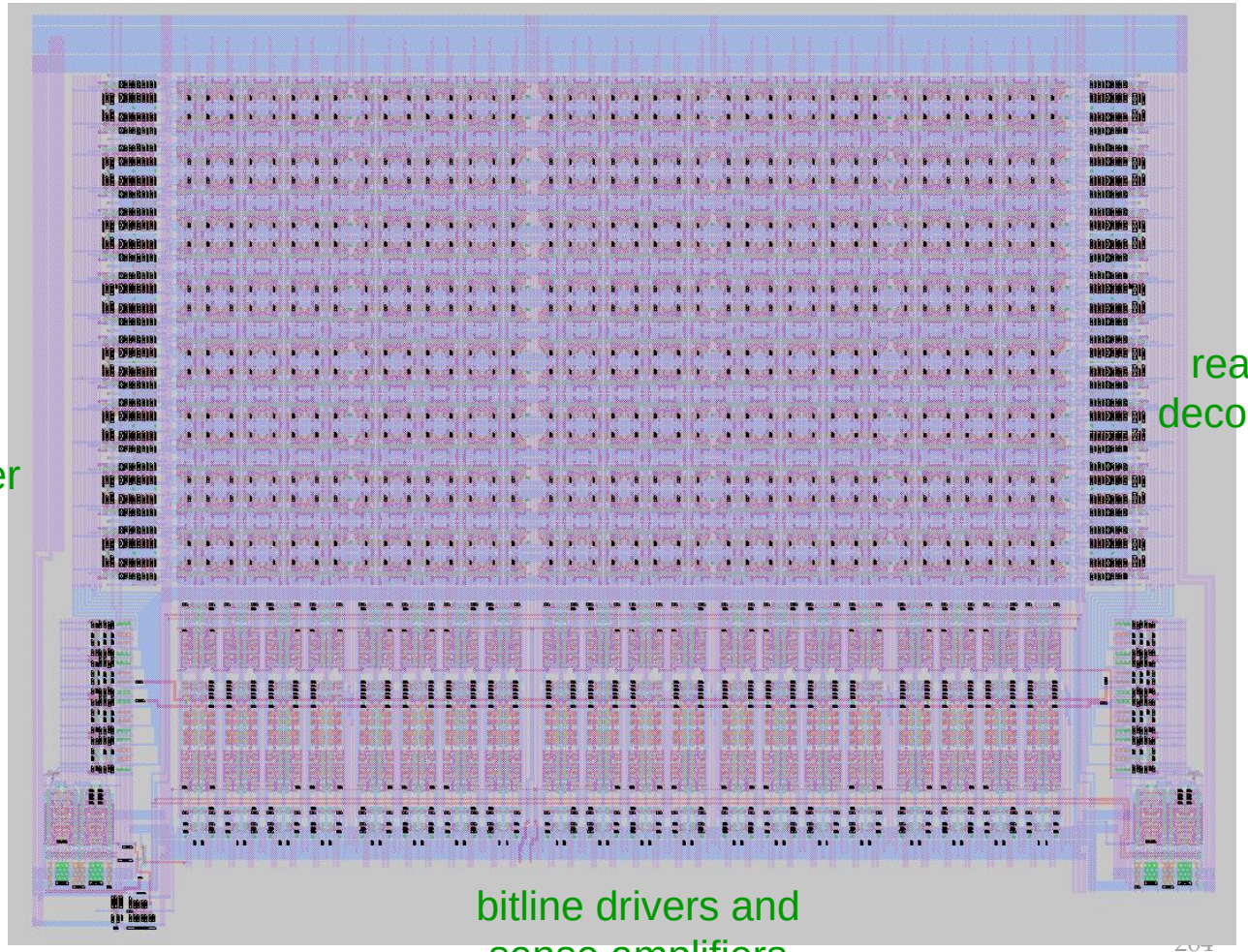
Layout: Dual-ported memory cell array

- Eight 10T dual-ported memory cells



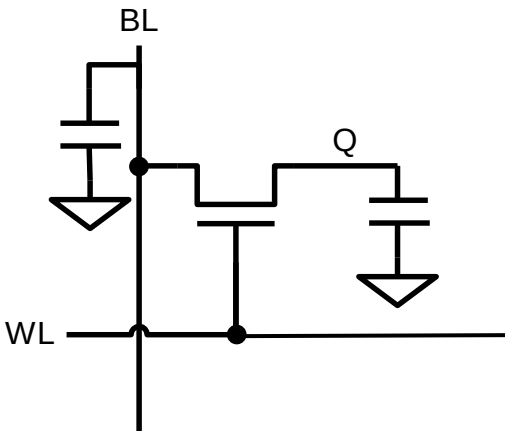
Layout: Dual-ported memory array

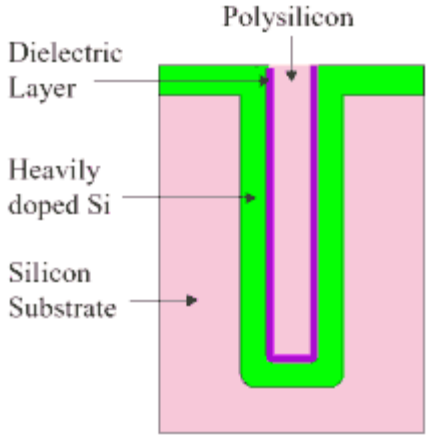
- 10T one-read, one-write port
- 16 words x 40 bits



DRAM

- Smaller cell size (1T cell)
  - One transistor to access cell
  - Often has special structure for cell capacitor (trench capacitor) that is not available in standard fabrication technologies
- Single bitline to read and write the memory cell
- Must be periodically refreshed
- Write operation:
  - Wordline at *Vdd*
  - Bitline value stored on cell capacitor



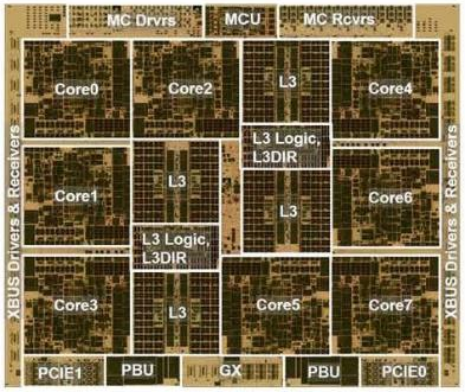


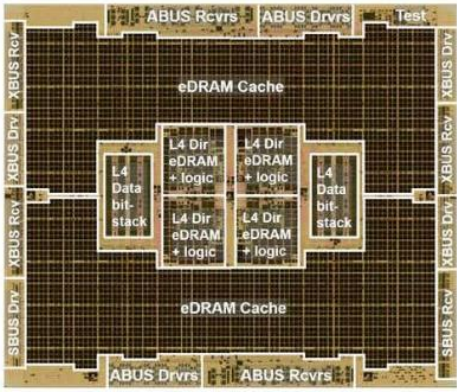
"Trench" capacitor

DRAM

- Classic DRAM
  - High density dedicated DRAM chips typically packaged in DIMMs
  - Top manufacturers include: Samsung (47%), Hynix (26%), Micron (19%), and others.
    [Source: Statista, for Q2 2016]

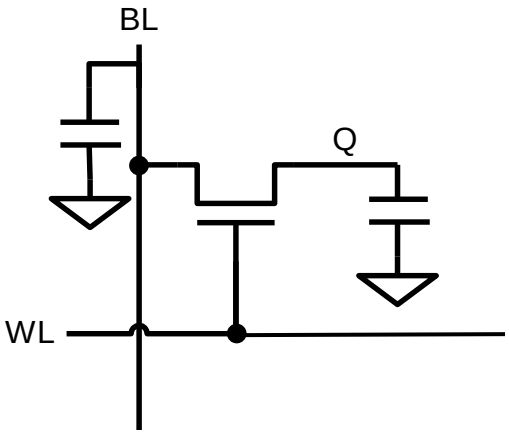
- Embedded DRAM (eDRAM)
  - DRAM arrays are also available as a process option for standard "logic"-type CMOS fabrication technologies
  - Ex: IBM 22 nm z Processor
    - "CP" Microprocessor chip on left: 8 cores, 64MB of eDRAM Level-3 cache, 678 mm2 die area, 4.0 billion transistors, 17 metal layers.
      eDRAM is used for Level-3 cache and memory controller; and also inside each processor for Level-2 and Branch Target Buffer caches.
    - "SC" System Controller chip on right: 480 MB of eDRAM Level-4 cache, 678 mm2 die area, 7.1 billion transistors, 15 metal layers
    - *[Source: "22nm Next-Generation IBM System z Microprocessor", ISSCC, 2015.]*





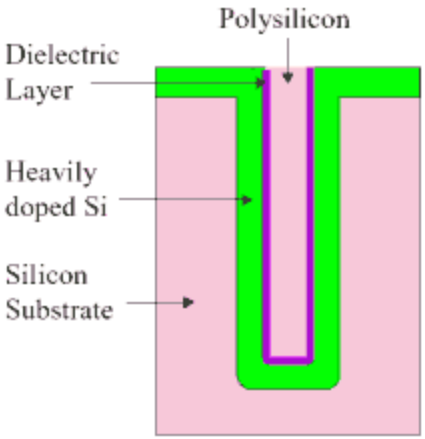
DRAM

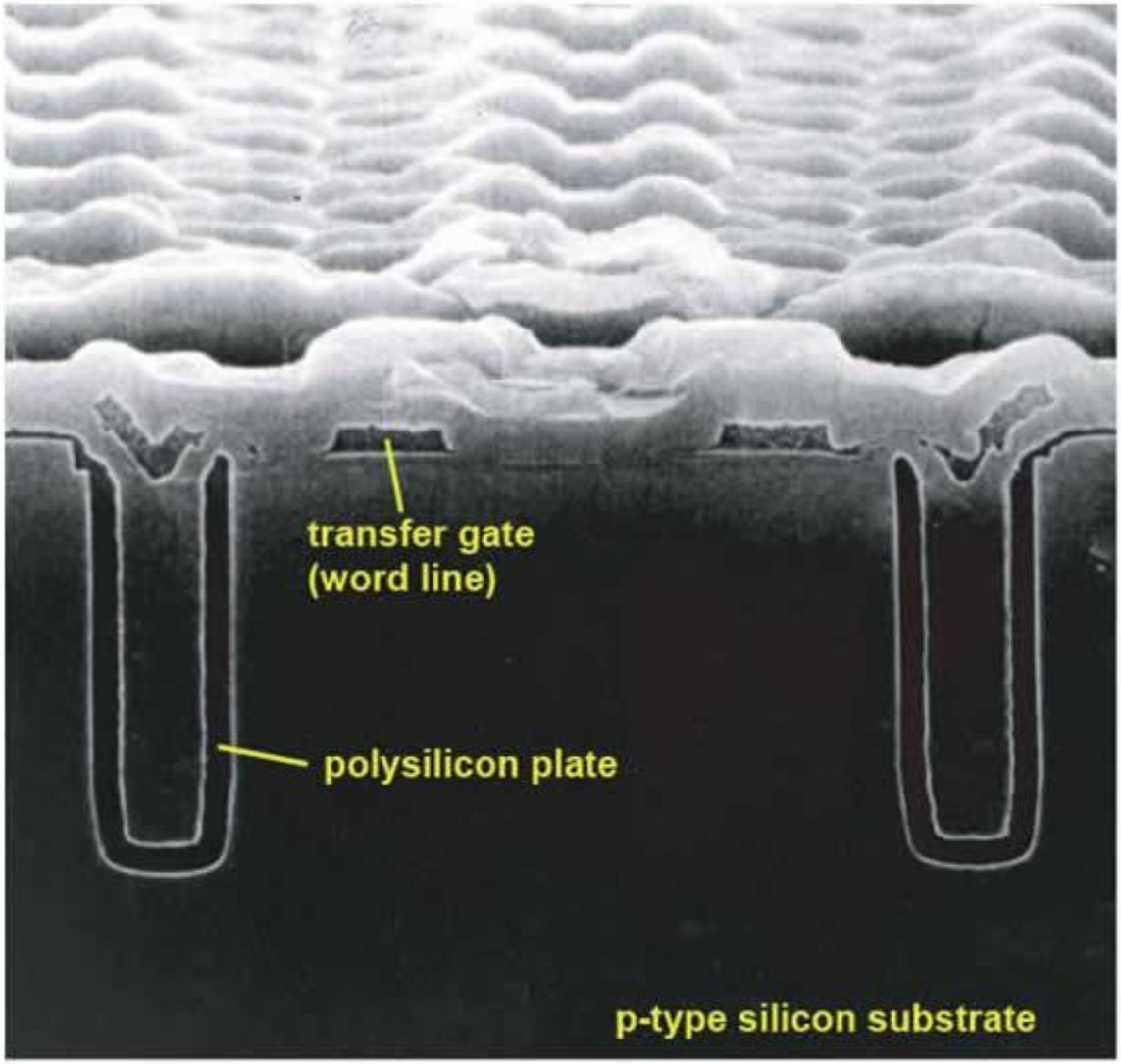
- Read operation:
  - Wordline at Vdd
  - Bitline precharged to intermediate voltage level (e.g., Vdd/2)
  - Read bitline voltage is the "charge shared" result of storage capacitor and bitline capacitance
    - C_cell ~ C_bitline / (10-100)
    - ΔV_bitline ~ 250 mV
- Read is a "destructive read" and value must be re-written to be read again
- Circuit challenges:
  - More complex sense amplifiers
  - Many use differential sense amplifier with dummy cell
  - Higher noise susceptibility
- *Vdd* not needed in array



Trench Capacitor

- Deep trench capacitor provides high capacitance per area
- Example to the right shows a DRAM cross section

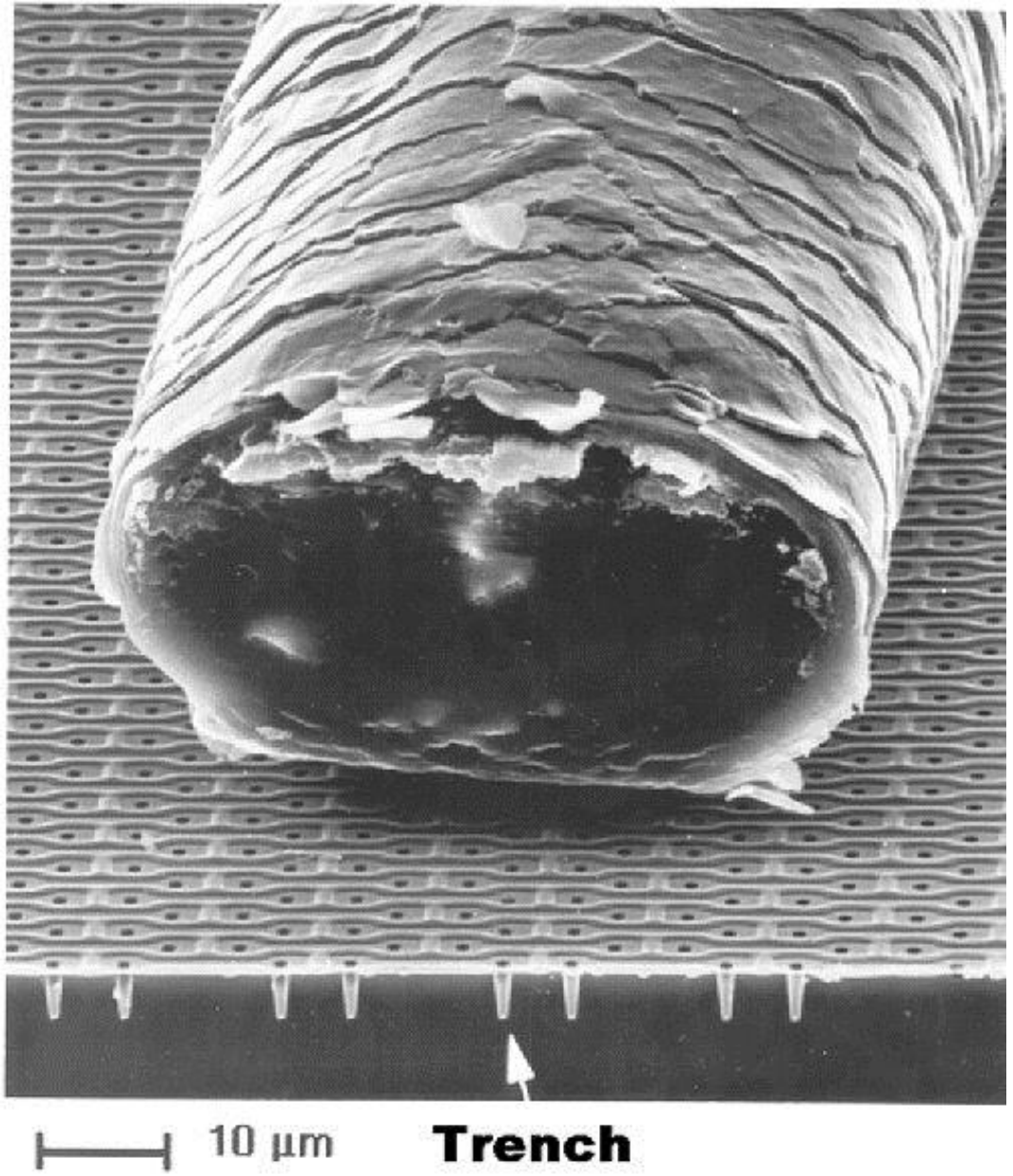




http://www.ieee.org/portal/cms_docs_sscs/sscs/08Winter/sunami-fig3.jpg
http://electron9.phys.utk.edu/phys136d/modules/m5/trench.gif

Memory Array

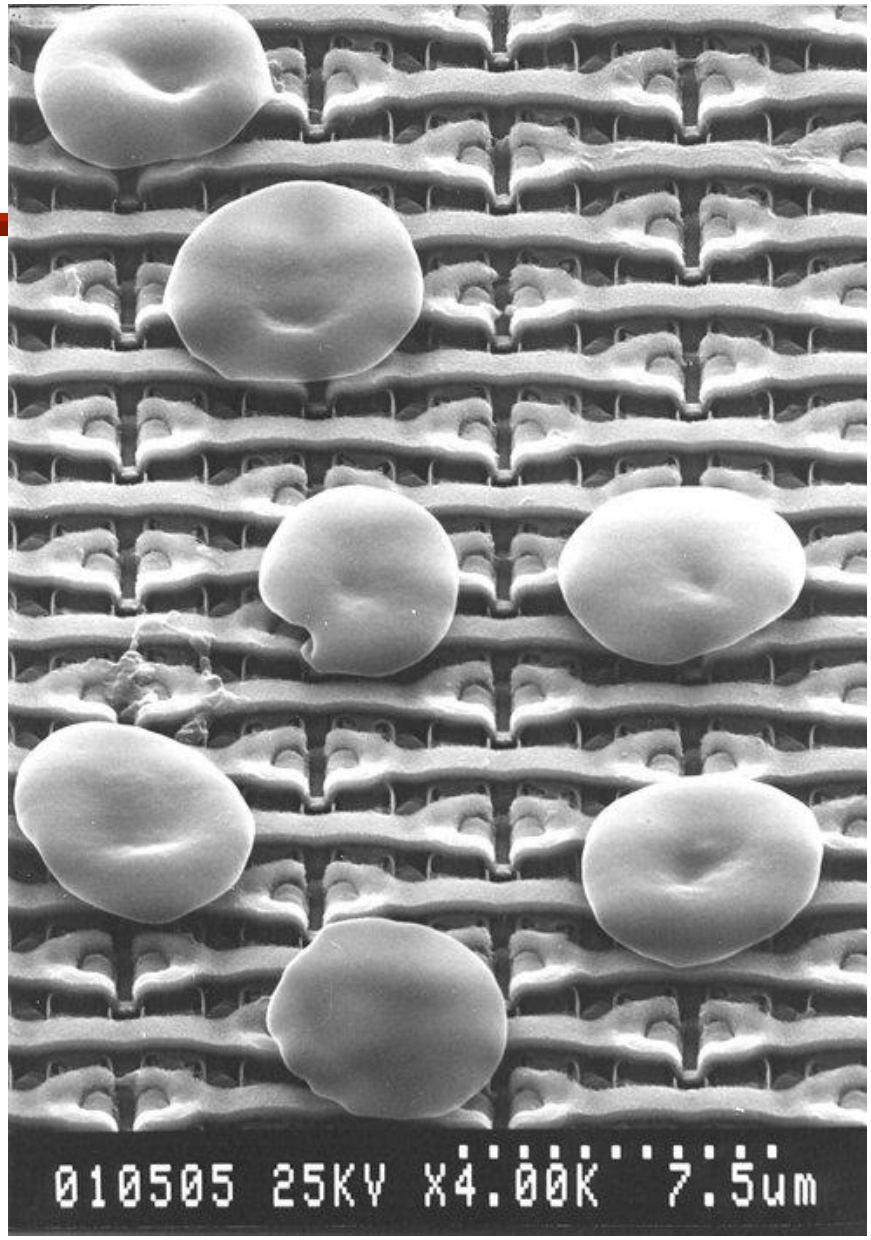
- Human hair on a 4 Mbit memory chip
- Note DRAM trench capacitors



Source: Helmut Föll

Memory Array

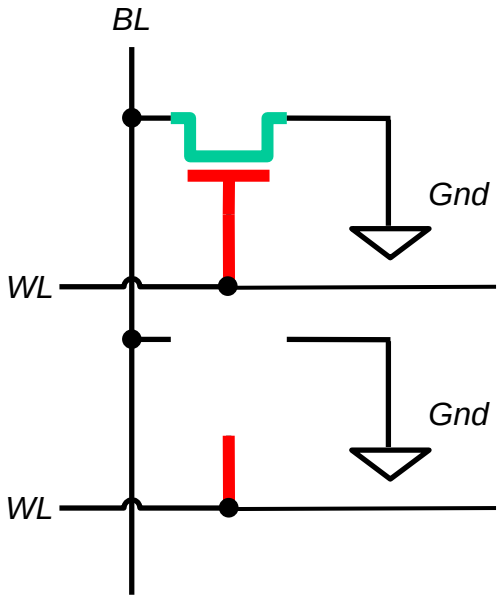
- Red blood cells on a 1 Mbit memory chip



Source: Helmut Föll

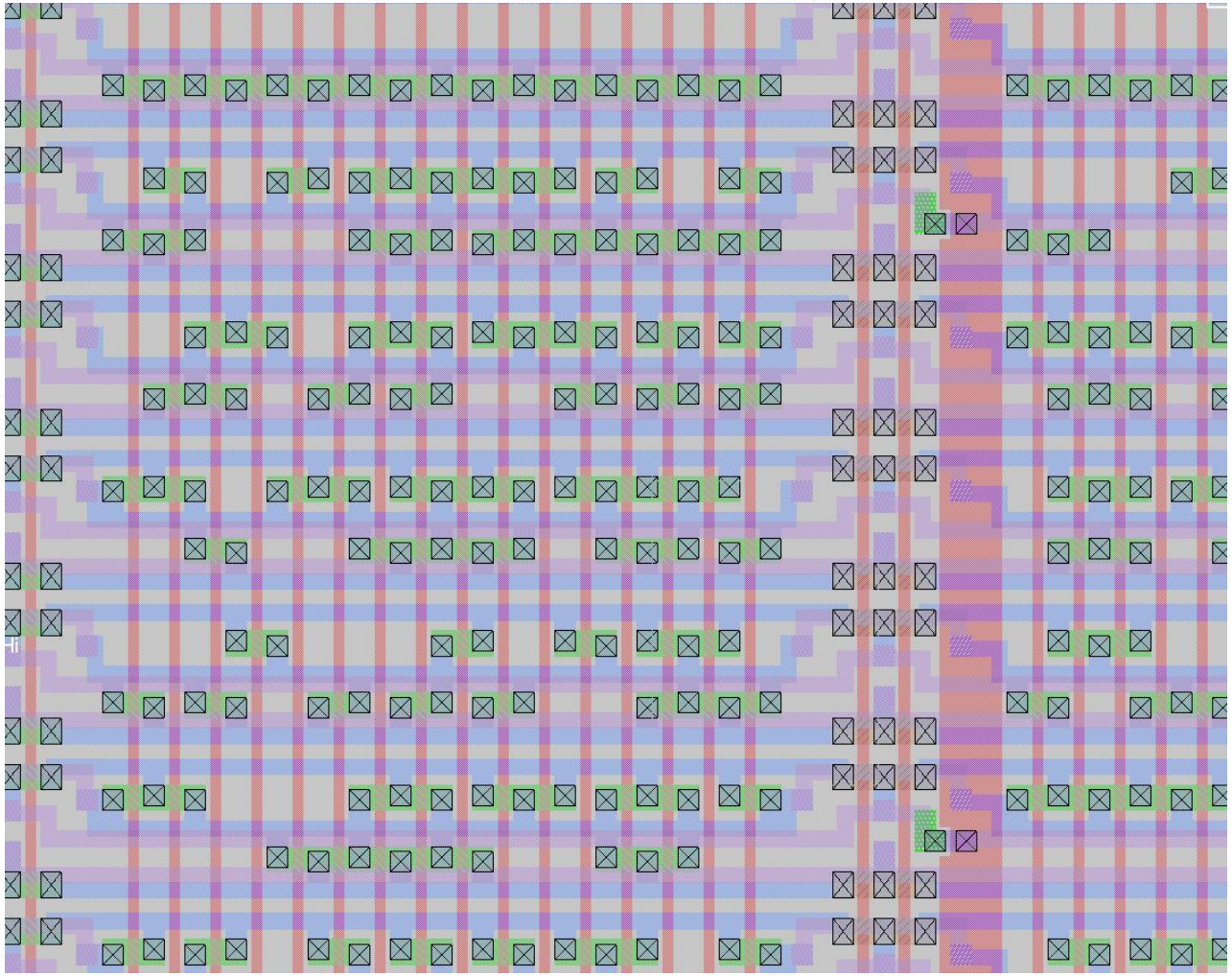
ROM

- Similar to DRAM except the source of the access transistor is tied to *Gnd* directly
- For the other value ($BL \neq 0$), the circuit is modified so that nothing happens when *WL* goes high. Possibilities include:
  - Omit cell-*Gnd* to global-*Gnd* connection
  - Omit source-to-*Gnd* or drain-to-*bitline* contact
  - Omit poly to *WL* contact
  - Omit MOS transistor diffusion
  - Omit transistor entirely
- In this particular example circuit:
  - A circuit to precharge bitlines is needed but not shown
  - *bitline* is driven low very strongly when a transistor is present
- ROMs are typically among the densest of all circuits
- Keep in mind that sometimes it is more efficient to use synthesized random logic instead of a ROM, especially if the stored data contains patterns and is not highly random



Layout: ROM cell array

- 1 transistor cells
- The presence or absence of a transistor's connection determines the stored value



Layout: ROM array

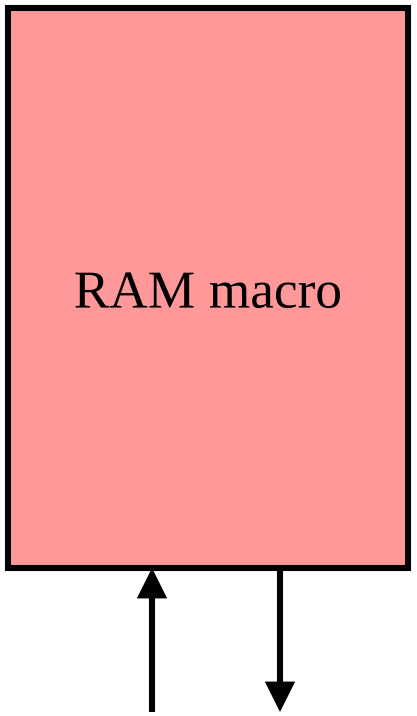
- 256 words x 40 bits
- Main issue: pitch matching decoders with cell array
  - ROM cells have a very fine pitch!

Memory View 4: Memory Types for Custom-Designed Chips (Also known as "ASICs")

- Memories for custom processors can be built in a number of ways:
  1) On-chip "macro" memory arrays
     A. Think of as a single giant standard cell
     B. FPGAs include them ("block RAMs" or "block memory")
  2) On-chip memory synthesized from verilog
     A. standard cells (e.g., NANDs, NORs, FFs, etc.)
     B. FPGA combinational logic blocks, LUTs, etc.
  3) Off-chip memories (often for > approx. 10 MB)
     - Very large DRAM
     - Non-volatile memory such as flash memory
     - (We could also include disks, NAS, cloud, etc.)

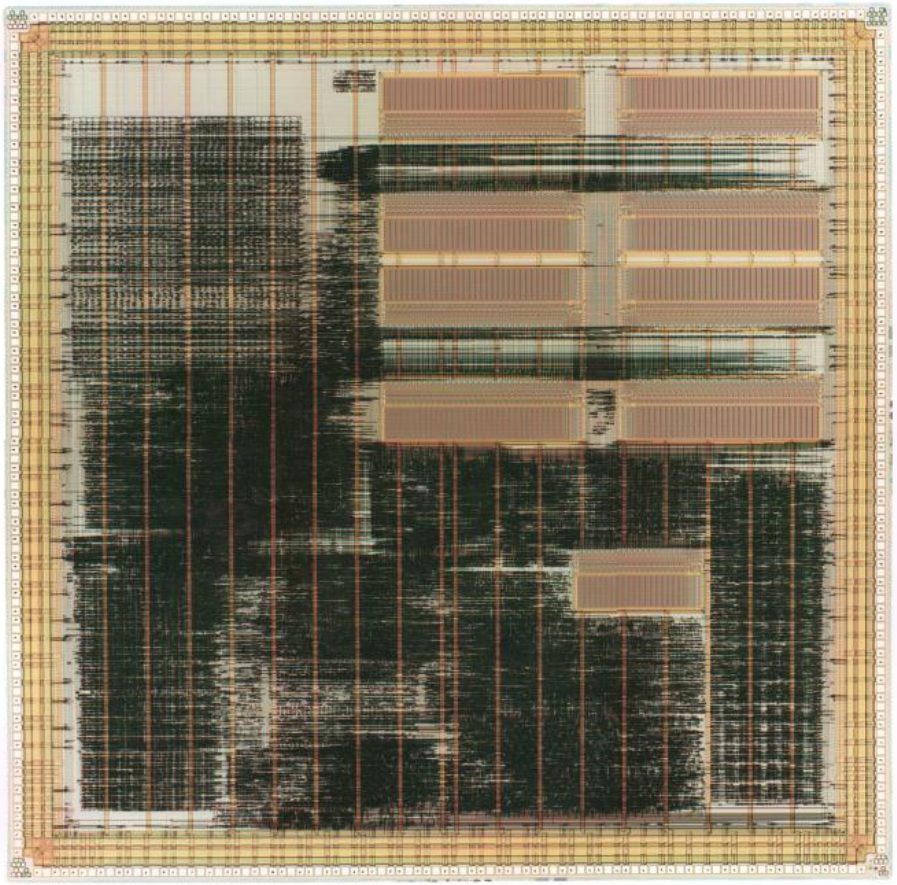
1A) On-Chip Memory "Macro" Arrays

- Memory macro-cell generators are available for larger memories
- Typically a software tool generates a large variety of possible memories where a user may select options such as:
  - Number of words
  - Word-width (in bits)
  - Number of read ports
  - Number of write ports
  - Rd/wr or ROM
  - Built-in test circuits
  - Registered inputs and/or outputs
- Tool produces models for verilog, place & route, and other CAD views



1A) On-Chip Memory "Macro" Arrays

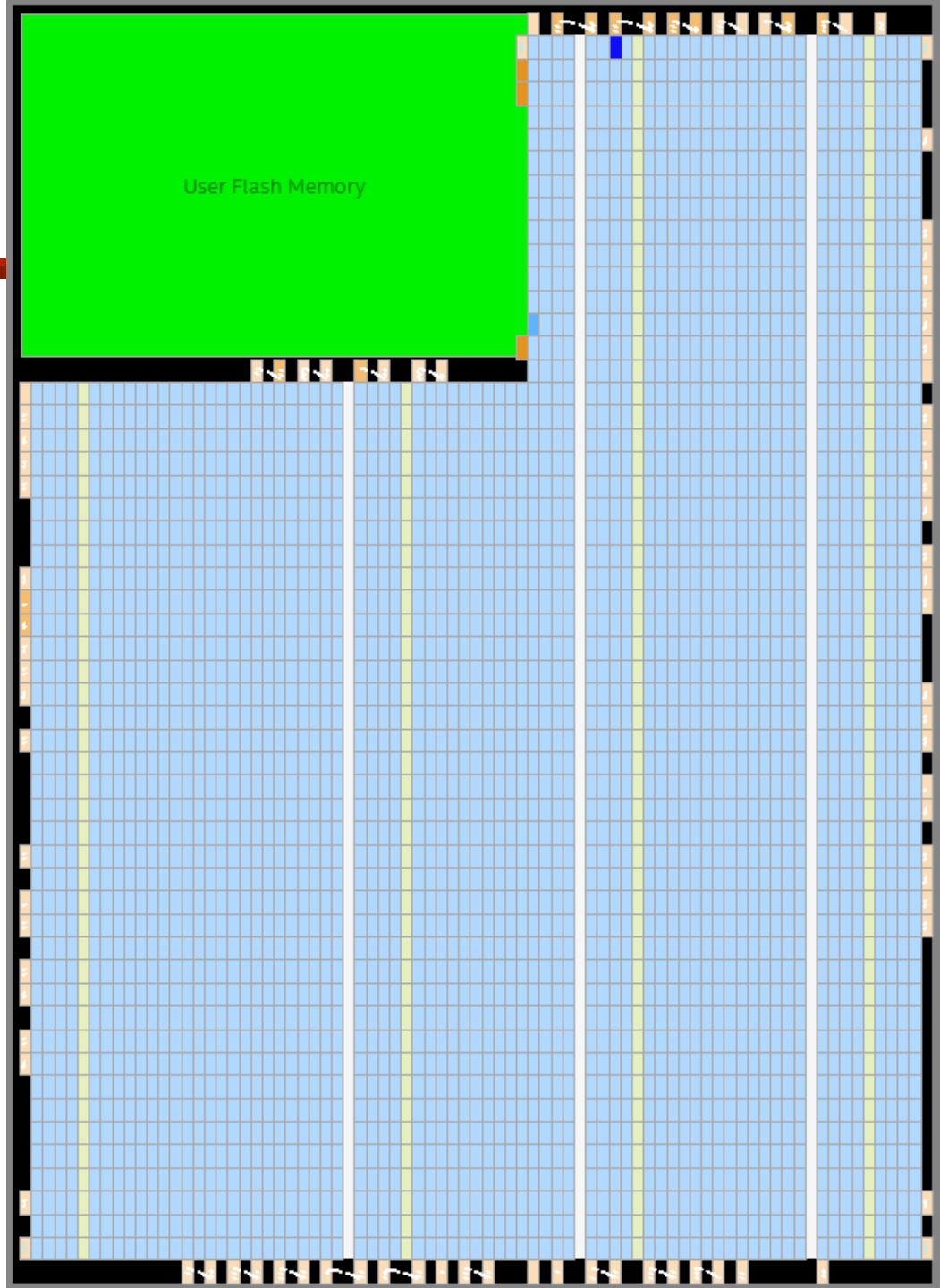
- Generally very area efficient due to dense memory cells (single-ported memories likely use 6-transistor (6T) memory cells)
- Generally good energy efficiency due to low-activity memory array architecture
- Example: CMOS chip



[T. Nanya, et al., TITAC-2 0.5 um CMOS, 496K transistors, 12.15 mm × 12.15 mm processor]

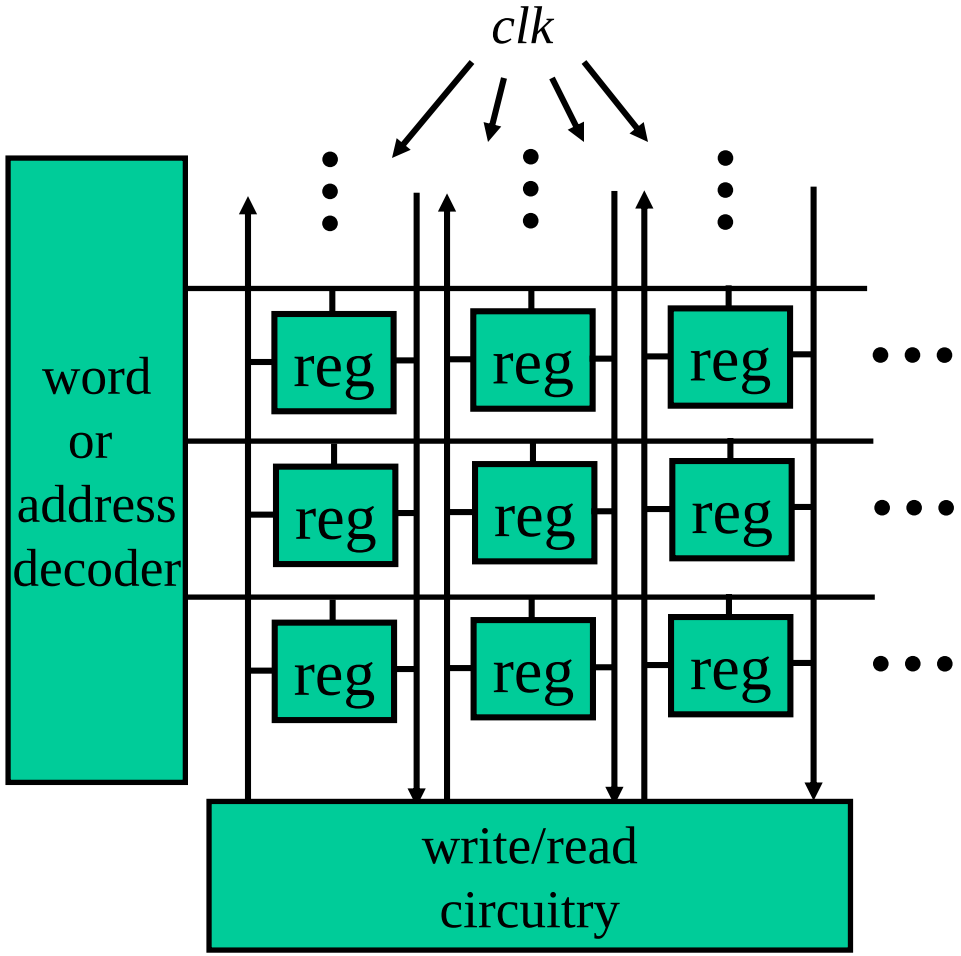
1B) On-chip Memory "macro" arrays: FPGAs

- Example: FPGA
- Altera Max 10 10M50DAF484C7G chip
- Yellow rectangles are M9K memory blocks
  - Each block contains 8192 bits (9216 including parity)
  - 182 on each chip
  - Total of 182 KBytes (204 KB)
- Light-blue rectangles: Logic Array Blocks (LAB)
- White rectangles: hardware 18x18 multipliers (144 on chip)



2) Synthesized Memory

- Can synthesize memory from standard cells
  - Memory cells are now flip-flops
  - *clk* likely routed to all cells
  - Probably best for small memories only
  - Read bitline logic may be muxes



2A) Synthesized Memory: Standard Cells

- Standard cell layout is typically irregular
  - Wires not shown
  - Clocks routed to each "reg" (flip-flop)