UNIVERSITY OF THESSALY

MOSTEK DRAM 4116 and Z80 CPU context

11/11/2025

CPU DRAM Refresh Interface

zilog
Embedded in Life
An IXYS Company

Z80 Microprocessors

Z80 CPU

User Manual

UM008011-0816

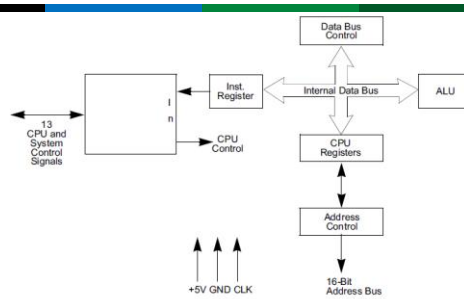


Figure 1. Z80 CPU Block Diagram

Memory Refresh (R) Register. The Z80 CPU contains a memory refresh counter, enabling dynamic memories to be used with the same ease as static memories. Seven bits of this 8-bit register are automatically incremented after each instruction fetch. The eighth bit remains as programmed, resulting from an LD *R*, *A* instruction. The data in the refresh counter is sent out on the lower portion of the address bus along with a refresh control signal while the CPU is decoding and executing the fetched instruction. This mode of refresh is transparent to the programmer and does not slow the CPU operation. The programmer can load the R register for testing purposes, but this register is normally not used by the programmer. During refresh, the contents of the I Register are placed on the upper eight bits of the address bus.

CPU DRAM Refresh Interface

CPU Register

The Z80 CPU contains 208 bits of read/write memory that are available to the programmer. Figure 2 shows how this memory is configured to eighteen 8-bit registers and four 16-bit registers. All Z80 CPU's registers are implemented using static RAM. The registers include two sets of six general-purpose registers that can be used individually as 8-bit registers or in pairs as 16-bit registers. There are also two sets of Accumulator and Flag registers and six special-purpose registers.

Main Register Set		Alternate Register Set		
Accumulator	Flags	Accumulator	Flags	
A	F	A '	F '	
B	C	B '	B '	General Purpose Registers
D	E	D '	E '	
H	L	H '	L '	

Interrupt Vector I	Memory Refresh R	Special Purpose Registers
Index Register	IX	
Index Register	IY	
Stack Pointer	SP	
Program Counter	PC	

Figure 2. CPU Register Configuration

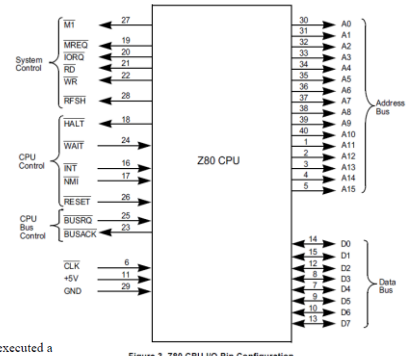


Figure 3. Z80 CPU I/O Pin Configuration

HALT. *HALT State (output, active Low).* HALT indicates that the CPU has executed a HALT instruction and is waiting for either a nonmaskable or a maskable interrupt (with the mask enabled) before operation can resume. During HALT, the CPU executes NOPs to maintain memory refreshes.

CPU DRAM Refresh Interface

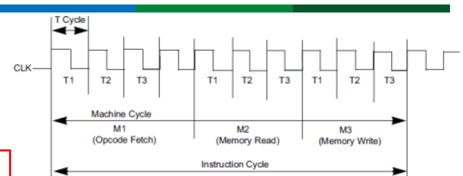


Figure 4. Basic CPU Timing Example

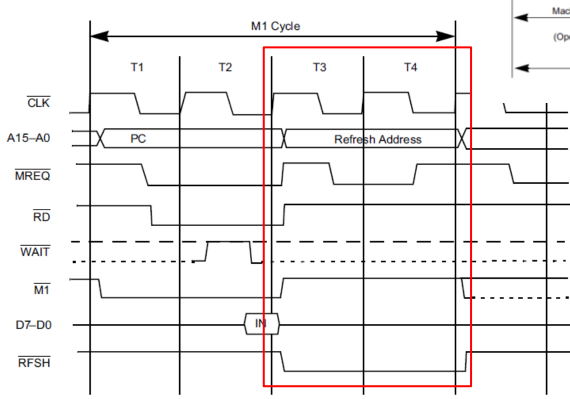


Figure 5. Instruction Op Code Fetch

CPU DRAM Refresh Interface

- During T3 and T4,
  - the lower seven bits of the address bus contain a memory refresh address
  - and the RFSH signal becomes active, indicating that a refresh read of all dynamic memories must be performed

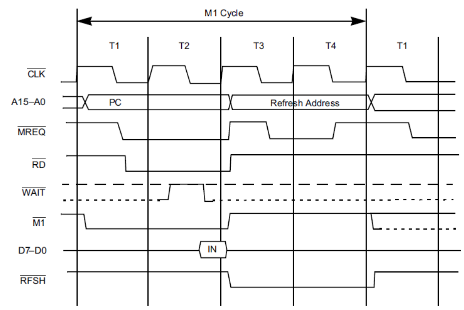


Figure 5. Instruction Op Code Fetch

5

DRAM Interface

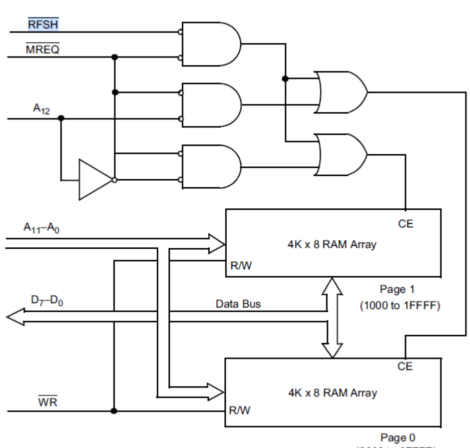


Figure 23. Interfacing Dynamic RAM Memory Spaces

Interfacing Dynamic Memories

Each individual dynamic RAM space includes its own specifications that require minor modifications to the examples provided here.

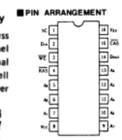
Figure 23 shows the logic necessary to interface 8KB of dynamic RAM using 18-pin 4K dynamic memories. This logic assumes that the RAMs are the only memory in the system so that A12 is used to select between the two pages of memory. During refresh time, all memories in the system must be read. The CPU provides the correct refresh address on lines A0 through A6. When adding more memory to the system, it is necessary to replace only the two gates that operate on A12 with a decoder that operates on all required address bits. Address buffers and data bus buffers are generally required for larger systems.

HM4864-2, HM4864-3
HM4864P-2, HM4864P-3

65536-word × 1-bit Dynamic Random Access Memory

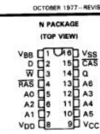
65536-word × 1-bit Dynamic Random Access Memory

The HM4864 is a 65,536-words by 1-bit, MOS random access memory circuit fabricated with HITACHI's double-poly N-channel silicon gate process for high performance and high functional density. The HM4864 uses a single transistor dynamic storage cell and dynamic control circuitry to achieve high speed and low power dissipation.



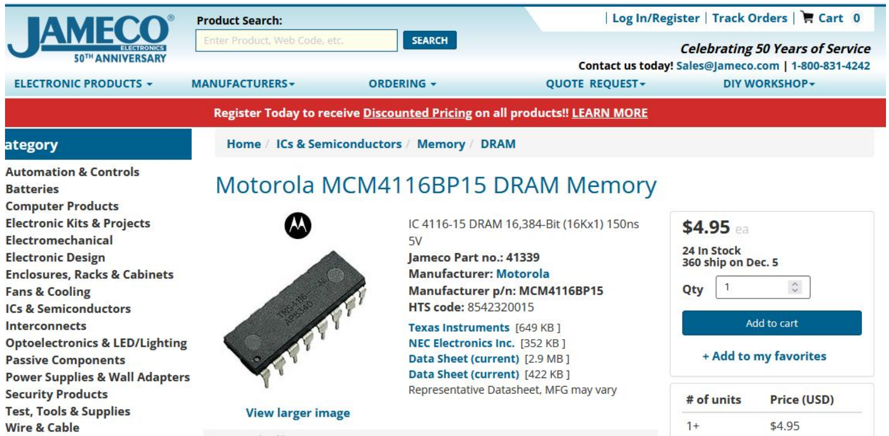
TMS4116
16,384-BIT DYNAMIC RANDOM-ACCESS MEMORY

OCTOBER 1977—REVISED NOVEMBER 1985



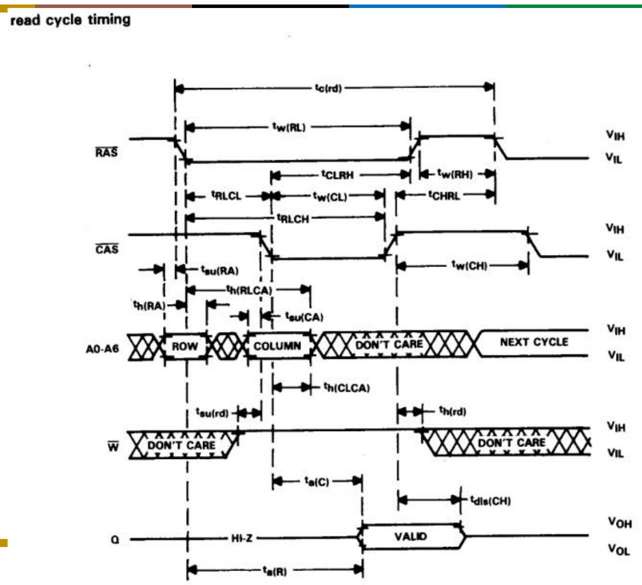
6

4116 DRAM



7

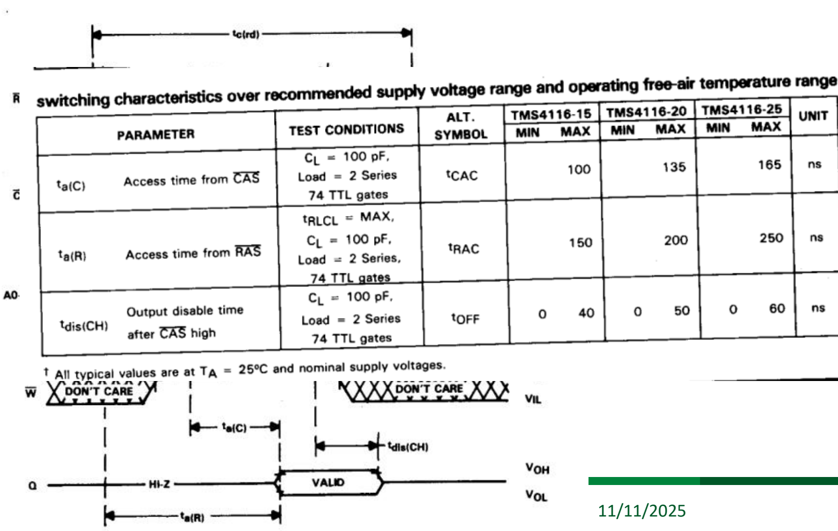
4116 DRAM - read



8

4116 DRAM - read

read cycle timing



switching characteristics over recommended supply voltage range and operating free-air temperature range

PARAMETER		TEST CONDITIONS	ALT. SYMBOL	TMS4116-15 MIN	TMS4116-15 MAX	TMS4116-20 MIN	TMS4116-20 MAX	TMS4116-25 MIN	TMS4116-25 MAX	UNIT
$t_{a(C)}$	Access time from $\overline{CAS}$	C_L = 100 pF, Load = 2 Series 74 TTL gates	t_{CAC}		100		135		165	ns
$t_{a(R)}$	Access time from $\overline{RAS}$	t_{RLCL} = MAX, C_L = 100 pF, Load = 2 Series, 74 TTL gates	t_{RAC}		150		200		250	ns
$t_{dis(CH)}$	Output disable time after $\overline{CAS}$ high	C_L = 100 pF, Load = 2 Series 74 TTL gates	t_{OFF}	0	40	0	50	0	60	ns

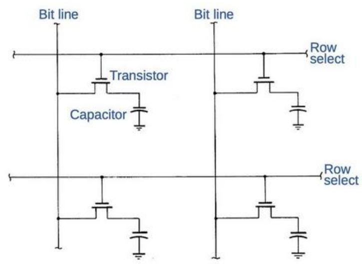
† All typical values are at T_A = 25°C and nominal supply voltages.

4116 DRAM

- Back in the late 1970s, the most popular memory chip was Mostek's MK4116, holding a whopping (for the time) 16 kilobits. It provided storage for computers such as the Apple II, TRS-80, ZX Spectrum, Commodore PET, IBM PC, and Xerox Alto as well as video games such as Defender and Missile Command

4116 DRAM

- In dynamic RAM, each bit is stored in a capacitor with the bit's value, 0 or 1, represented by the voltage on the capacitor.3
  - The advantage of dynamic RAM is that each memory cell is very small, so a lot of data can be stored on one chip.4
  - The downside of dynamic RAM is that the charge on a capacitor leaks away after a few milliseconds.
  - To avoid losing data, dynamic RAM must be constantly refreshed: bits are read from the capacitors, amplified, and then written back to the capacitors.
  - For the MK4116, all the data must be refreshed every two milliseconds.



Structure of the memory cells, based on the patent

11

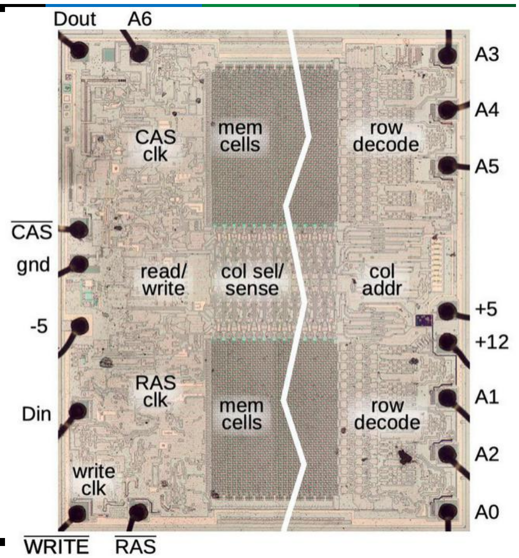
4116 DRAM

- Mostek cut the number of address pins in half by using each address pin twice,
  - first for a "row" address, and then a "column" address.
  - This approach became the industry standard, allowing memory chips to fit into inexpensive 16-pin packages
- Externally, the chip stores a single bit for 16,384 different addresses.
  - (Typically, eight of these chips were used in parallel to store bytes.)
  - Internally, however, the chip is implemented as a 128×128 matrix of storage cells

12

4116 DRAM

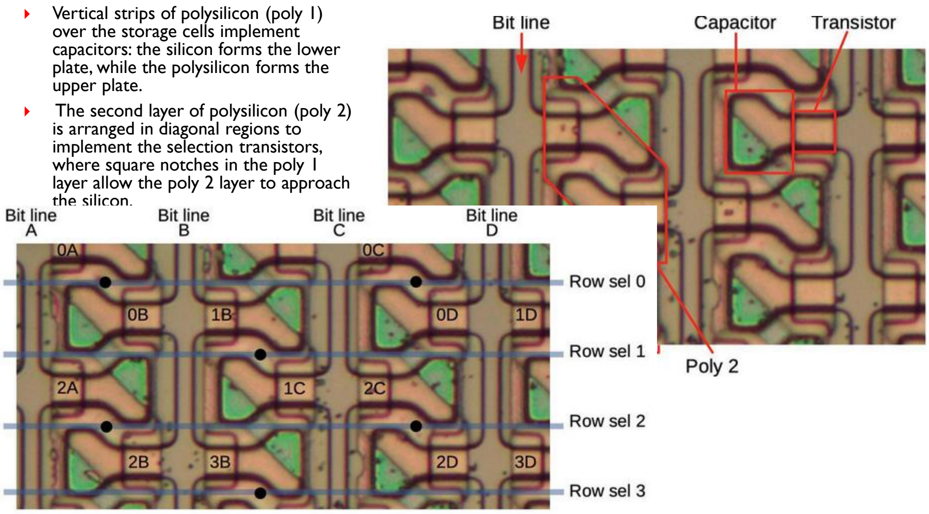
- seven address pins (A0-A6). The Row Address Strobe pin (RAS) is used to indicate the row address is ready, while the Column Address Strobe pin (CAS) indicates that the column address is ready
- The two memory arrays are in the center
- At the right, the row decode circuitry selects a row based on the address pins, while the column address circuitry buffers the address for the column select circuitry.
- At the left, the clock circuits generate the chip's timing pulses, triggered by the RAS, CAS, and WRITE pins.
- Finally, the Data Out and Data In pins provide access to the selected data bit.



13

4116 DRAM

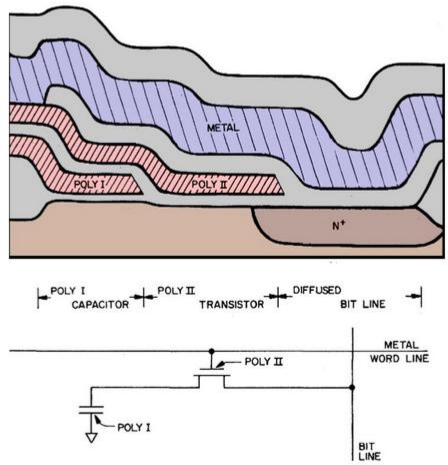
- Vertical strips of polysilicon (poly 1) over the storage cells implement capacitors: the silicon forms the lower plate, while the polysilicon forms the upper plate.
- The second layer of polysilicon (poly 2) is arranged in diagonal regions to implement the selection transistors, where square notches in the poly 1 layer allow the poly 2 layer to approach the silicon.



14

4116 DRAM

- The cross-section diagram below shows the three-dimensional, layered structure of a memory cell. At the bottom is the silicon (brown); the bit line (dark brown) is made from doped silicon. Above the silicon are the two polysilicon layers (red) and the metal layer (purple), separated by insulating silicon dioxide (gray).
- At the far left, the poly I layer and underlying silicon form a capacitor. In between the capacitor and the bit line, the poly 2 layer forms the gate of the transistor.
- At the left, the poly 2 layer is connected to the metal of the word line, which turns the transistor on, connecting the capacitor to the bit line.

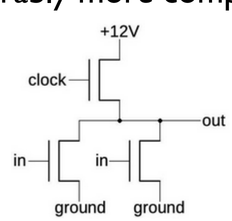


Cross-section structure of a storage cell. Based on 16K—The new generation dynamic

15

4116 DRAM

- the 4116 used earlier NMOS transistors. Most NMOS integrated circuits constructed logic gates with load transistors,
- a simple technique with the disadvantage of wasting power. Instead, the MK4116 memory chip uses dynamic logic, which is considerably more complex but saves power while idle.



A NOR gate using dynamic logic.

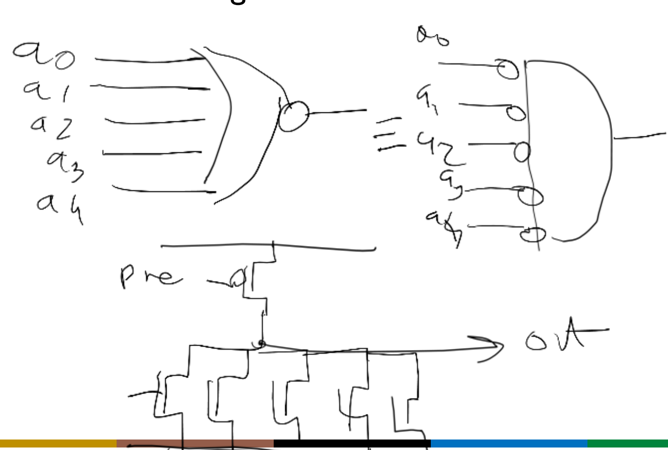
16

4116 DRAM – Row Select

- The purpose of the row-select circuitry is to
  - decode the 7 address bits and energize the corresponding row select line (out of 128) to read one row of memory
- In the first step, 32 5-input NOR gates decode address bits A0 through A4
- Each NOR gate takes a different combination of non-inverted and inverted address bits and matches a particular 5-bit address.
- These NOR gates use dynamic logic, first pulled high and then discharged to ground, except for the selected address which remains high
- Next, each NOR output is split into four, based on A5 and A6.
- The result is that one of 128 row select lines is activated, turning on the transistors for that row in the matrix.

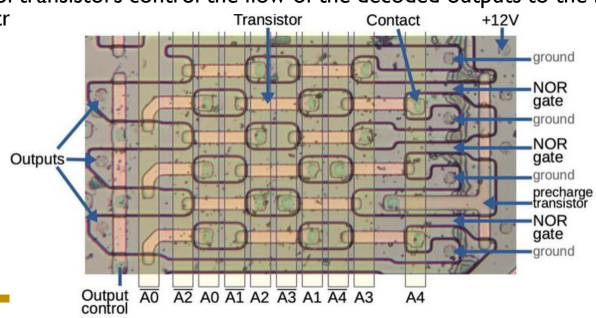
4116 DRAM – Row Select

- In the first step, 32 5-input NOR gates decode address bits A0 through A4



4116 DRAM – Row Select

- The NOR gates are implemented in several compact blocks; one block of three NOR gates is shown below.
- Each NOR gate is a horizontal stripe of doped silicon, with ground above and below it.
- Each NOR gate has transistors (pink stripes) connected to ground alternating above and below it.
- A transistor will pull the NOR gate low if the connected address line is high. The precharge transistors at the left pull the NOR gates to +12 volts, while the output control transistors control the flow of the decoded outputs to the rest of the circuitr

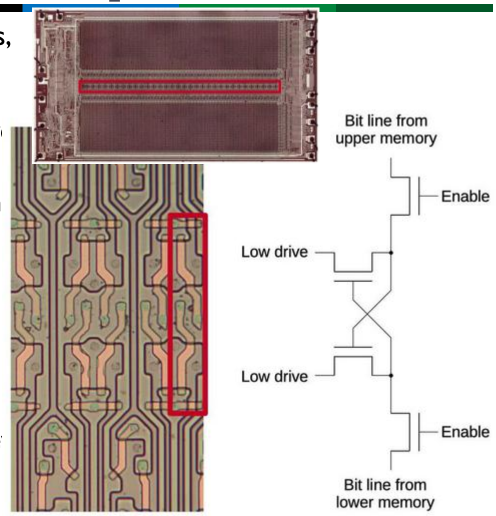


4116 DRAM – Sense Amplifiers

- The sense amplifiers are one of the most challenging parts of designing a memory chip.
- The job of the sense amplifier is to take the tiny voltage from a capacitor and amplify it into a binary 0 or 1.11
- The challenge is that even though 12 volts is stored in a capacitor, the signal from the capacitor is very small, is only 100 millivolts or so.
  - (Because the bit line is much larger than the tiny memory cell capacitor, the capacitor causes a very small voltage swing.)12
  - It is critically important for the sense amplifier to operate accurately, even in the presence of noise or voltage fluctuations, because any error will corrupt the data.
  - The sense amplifier circuit must also be compact and low power since there are 128 sense amplifiers.

4116 DRAM – Sense Amplifiers Refresh

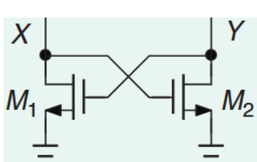
- The chip's 128 sense amplifiers, one for each column, are located between the two memory arrays as shown abov
- During a read, 128 values in a row are accessed in parallel an amplified by the sense amplifiers.
- These 128 values are then written back to refresh the values in the capacitor.
- For a write operation, one of the bits is updated with the ne value before they are written back.

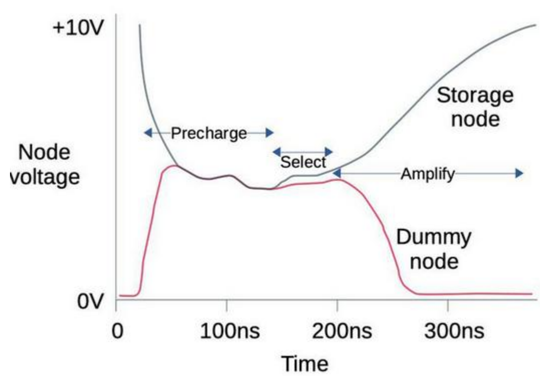


21

Sense Amplifier Operation

- Sense Amp precharged at Voltage midpoint

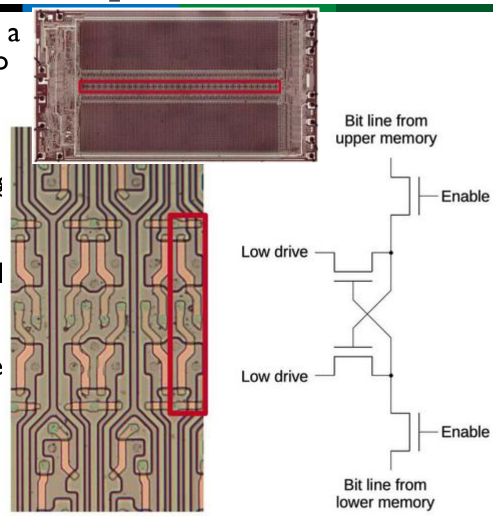




22

4116 DRAM – Sense Amplifiers Refresh

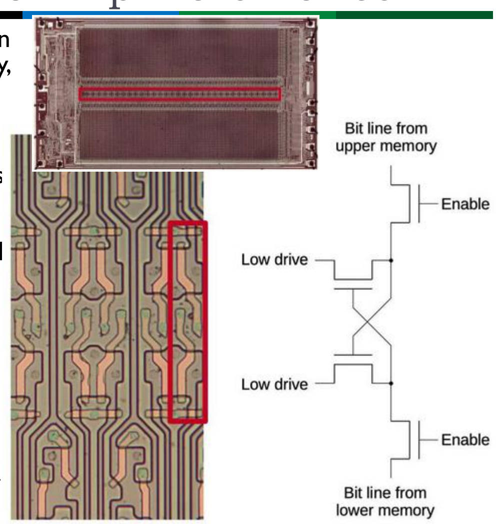
- Each sense amplifier (above) is a very simple circuit. It takes two inputs and compares them, pulling the lower one to 0.13
- It is built from two cross-coupled transistors, each tryin to pull the other one low.
- Whichever transistor has the higher voltage to start with wil "win", forcing the other side low.14
- The sense amplifier is sensitive to very small voltage differentials, allowing it to distinguish the small signals from a storage cell.



23

4116 DRAM – Sense Amplifiers Refresh

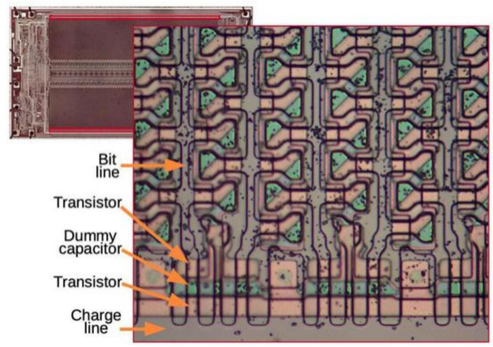
- Locating the sense amplifiers between the two memory arrays isn't arbitrary, but the key to their operation: this is the "divided bit line" architecture introduced in 1972.
- The idea is that one input to the sens amp is the voltage from the desired memory cell, while the other input is threshold voltage from a "dummy cell in the opposite memory array.
- Dummy cells are constructed and precharged like real memory cells except the capacitor is half-sized, so they provide a voltage midway between a 0 bit and a 1 bit.3
- If the voltage from the real memory cell is lower, the sense amp outputs a 0, and if higher, it outputs a 1.



24

4116 DRAM – Dummy Cells

- The dummy cells are located on the edges of the memory arrays, as shown above.
- They consist of capacitors and transistors (similar to real memory cells), but with a separate line to charge them. The advantage of the dummy cell approach is that manufacturing differences or fluctuations during operation will (hopefully) affect the real cells and dummy cells equally, so the voltage from the dummy cell will remain at the correct level to distinguish beween a 0 and a 1.
- Address bit A0 controls which half of the array provides real data to the bit lines and which half connects dummy cells to the bit lines.

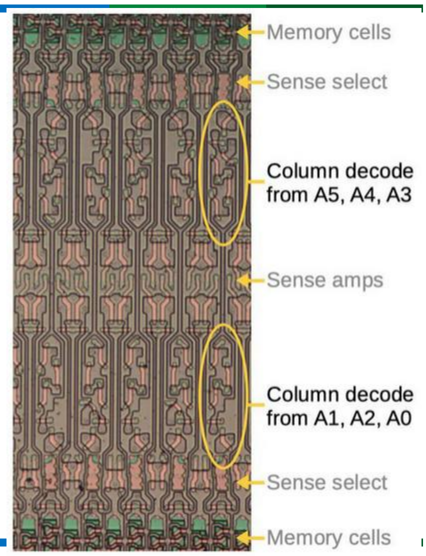


Dummy cells provide the threshold voltage for deciding if a bit is 0 or 1. The dummy cells are

25

4116 DRAM – Column Select

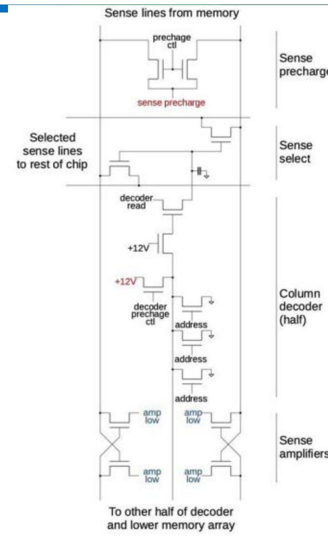
- the column select circuitry selects one column out of the 128-bit row; this is the bit that is read or written.
- Each column select circuit is twice as wide as a memory cell, so they only decode one of 64 columns.
- The result is that two bits are selected at a time, and circuitry elsewhere selects one of the two bits.
- Like the row select circuitry, the column select circuitry is implemented by numerous NOR gates, each matching one address.
- For column select address bits A0 through A5 select one of 64 lines, selecting two columns at a time.
- These two bit lines are connected to data lines transmitting the signals to the I/O circuitry. (Since the bit lines for the upper and lower halves of the matrix are separate, there are actually four bit lines selected by the column select circuit.)
- As with the row select circuitry, dynamic logic is used, controlled by various timing signals. Note that each NOR gate is physically split into two parts with the sense amp in the middle.



26

4116 DRAM – Column Select

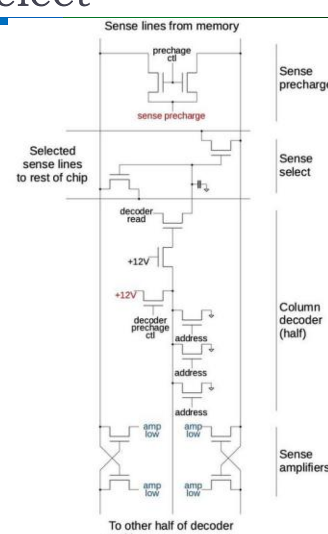
- The schematic shows how the column decoder works with the sense amplifier.
- The diagram shows two bit lines and the top half of the column decoder and sense circuitry; it is mirrored for the lower array.
- At the top, the sense precharge circuit pulls all the bit lines high.
- At the bottom, the sense amplifiers amplify and refresh the signals as explained above.
- The column decoder matches a particular 6-bit address, so one of the 64 decoders will activate the associated sense select circuit, connecting the chip's I/O circuitry to four bit lines (two from the upper memory array as shown here and two from the lower memory array).



27

4116 DRAM – Column Select

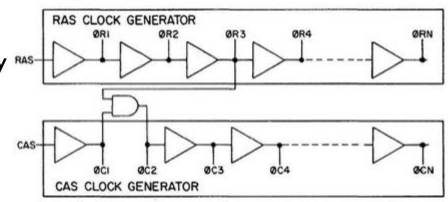
- At this point, four bit lines have been selected for use and their signals are passed to the input/output circuitry; the column select circuitry only decoded 1-of-64, while there are 128 columns, and each half of the array has separate bit lines.
- Column address bit A6 provides the final selection between the two columns. The selected bit is sent to the data-out pin for a read.
- For a write, the value on the data-in pin is sent back through the appropriate wire to overwrite the value in the sense amplifier.
- This circuitry is implemented using dynamic logic and latches, controlled by various timing signals.
- Much of the circuitry is duplicated, with one copy for the upper half of the memory array and one copy for the lower half.
- Row address bit A0 distinguishes which half of the matrix is active and which half is providing dummy data).
- (Note that row address bit A0 was already used to select a particular row, but the circuitry has "lost track" of which was the real row and which was the dummy row, so it must make the selection again.)



28

4116 DRAM – Clock Generation

- chip requires many timing signals for the various steps in a memory operations.
- The memory chip doesn't use an external clock, unlike a CPU, but generates its own timing signals internally.
- The diagram illustrates the clock generators, using buffers to create a delay between each successive clock output.
- The first set of timing signals is triggered by the row-access strobe (RAS), indicating that the computer has put the row address on the address pins.
- The next set of timing signals is triggered by the column-access strobe (CAS), indicating the column address is on the address pins.
- Other timing signals are triggered by the WRITE pin.



Conclusions

- The MK4116 is an engineering work of art!
- DRAM cells are essentially analog!!!
  - Analog circuits to sense and amplify signals
- TO Remember: REFRESH = READ
  - Due to sense amplifiers
- Intel produced 16,384-bit DRAM chips before Mostek, the 2116 and others,
  - but Mostek's chips beat Intel in the marketplace
- 16-gigabit DRAMs are now into production

16-gigabit DRAMs

SEOUL, South Korea--(BUSINESS WIRE)--Samsung Electronics Co., Ltd., the world leader in advanced memory technology, today announced that its second production line in Pyeongtaek, Korea, has commenced mass production of the industry's first 16-gigabit (Gb) LPDDR5 mobile DRAM, using extreme ultraviolet (EUV) technology. Built on Samsung's third-generation 10nm-class (1z) process, the new 16Gb LPDDR5 boasts the highest mobile memory performance and largest capacity to enable more consumers to enjoy the full benefits of 5G and AI features in next-generation smartphones.

"The 1z-based 16Gb LPDDR5 elevates the industry to a new threshold, overcoming a major developmental hurdle in DRAM scaling at advanced nodes," said Jung-bae Lee, executive vice president of DRAM Product & Technology at Samsung Electronics. "We will continue to expand our premium DRAM lineup and exceed customer demands, as we lead in growing the overall memory market."

Expanding Manufacturing Capacity in Pyeongtaek Complex

Spanning more than 128,900 square meters (over 1.3 million square feet) – equivalent to about 16 soccer fields – Samsung's Pyeongtaek Line 2 is the largest-scale semiconductor production line to date.

The new Pyeongtaek line will serve as the key manufacturing hub for the industry's most advanced semiconductor technologies, delivering cutting-edge DRAM followed by next-generation V-NAND and foundry solutions, while reinforcing the company's leadership in the Industry 4.0 era.

Fastest, Largest-capacity Mobile Memory

Based on today's most advanced (1z) process node, Samsung's new 16Gb LPDDR5 is the first memory to be mass produced using EUV technology, providing the highest speed and largest capacity available in mobile DRAM.

At 6,400 megabits per second (Mb/s), the new LPDDR5 is about 16 percent faster than the 12Gb LPDDR5 (5,500Mb/s) found in most of today's flagship mobile devices. When made into a 16GB package, the LPDDR5 can transfer about 10 5GB-sized full-HD movies, or 51.2GB of data, in one second.

Thanks to its use of the first commercial 1z process, the LPDDR5 package is 30 percent thinner than its predecessor, enabling 5G and multi-camera smartphones as well as foldable devices to pack more functionality into a slim design. The 16Gb LPDDR5 can build a 16GB package with only eight chips, whereas its 1y-based predecessor requires 12 chips (eight 12Gb chips and four 8Gb chips) to provide the same capacity.

By delivering the first 1z-based 16GB LPDDR5 to global smartphone makers, Samsung plans to further strengthen its presence in the flagship mobile device market throughout 2021. Samsung will also expand the use of its LPDDR5 offerings into automotive applications, offering an extended temperature range to meet strict safety and reliability standards in extreme environments.