Random access memory (RAM or PC memory)

Types of random access memory

There are generally two broad categories of random access memory:

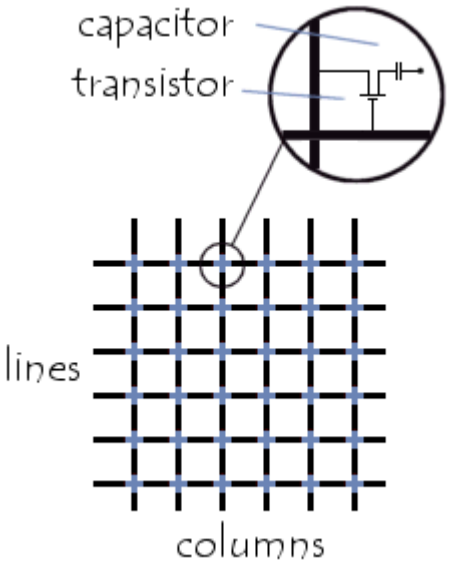
- **DRAM** memories (*Dynamic Random Access Module*), which are inexpensive. They are used essentially for the computer's main memory
- **SRAM** memories (*Static Random Access Module*), which are fast and costly. SRAM memories are used in particular for the processor's cache memory

Operation of the random access memory

The random access memory comprises hundreds of thousands of small capacitors that store loads. When loaded, the logical state of the capacitor is equal to 1, otherwise it is 0, meaning that each capacitor represents one memorybit.

Given that the capacitors become discharged they must be constantly recharged (the exact term is *refresh*) at regular intervals, known as the **refresh cycle**. DRAM memories for example require refresh cycles of around 15 nanoseconds (ns).

Each capacitor is coupled with a transistor (*MOS-type*) enabling "recovery" or amendment of the status of the capacitor. These transistors are arranged in the form of a table (matrix) thus we access a *memory box* (also called *memory point*) via a line and a column.



Each memory point is thus characterised by an address which corresponds to a row number and a column number. This access is not instant and the access time period is known as **latency time**. Consequently, time required for access to data in the memory is equal to cycle time plus latency time.

Thus, for a DRAM memory, access time is 60 nanoseconds (35ns cycle time and 25ns latency time). On a computer, the cycle time corresponds to the opposite of the clock frequency; for example, for a computer with frequency of 200 MHz, cycle time is 5 ns ($1/200*10^6$)).

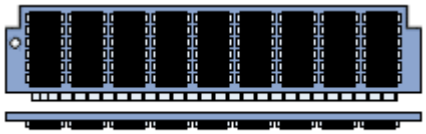
Consequently a computer with high frequency using memories with access time much longer than the processor cycle time must perform **wait states** to access the memory. For a computer with frequency of 200 MHz using DRAM memories (and access time of 60ns), there are 11 wait states for a transfer cycle. The computer's performance decreases as the number of wait states increases, therefore we recommend the use of faster memories.

RAM module formats

There are many type of random access memory. They exist in the form of memory modules that can be plugged into the mother board.

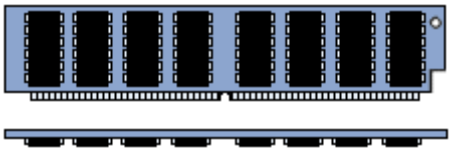
Early memories existed in the form of chips called *DIP* (*Dual Inline Package*). Nowadays, memories generally exist in the form of modules, which are cards that can be plugged into connectors for this purpose. There are generally three types of RAM module:

- modules in ***SIMM*** format (*Single Inline Memory Module*): these are printed circuit boards with one side equipped with memory chips. There are two types of SIMM modules, according to the number of connectors:
  - SIMM modules with 30 connectors (dimensions are 89x13mm) are 8-bit memories with which first-generation PCs were equipped (286, 386).



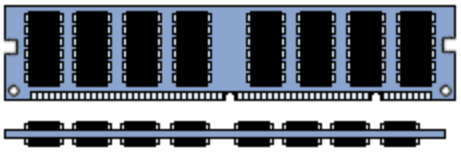
SIMM memory - 30 pin

  - SIMM modules with 72 connectors (dimensions are 108x25mm) are memories able to store 32 bits of data simultaneously. These memories are found on PCs from the 386DX to the first Pentiums. On the latter, the processor works with a 64-bit data bus; this is why these computers must be equipped with two SIMM modules. 30-pin modules cannot be installed on 72-connector positions because a notch (at the centre of the connectors) would prevent it from being plugged in.



SIMM memory - 72 pin

- modules in **DIMM** format (*Dual Inline Memory Module*) are 64-bit memories, which explains why they do not need pairing. DIMM modules have memory chips on both sides of the printed circuit board and also have 84 connectors on each side, giving them a total of 168 pins. In addition to having larger dimensions than SIMM modules (130x25mm), these modules have a second notch to avoid confusion.



DIMM memory

It may be interesting to note that the DIMM connectors have been enhanced to make insertion easier, thanks to levers located either side of the connector.

Smaller modules also exist; they are known as **SO DIMM** (*Small Outline DIMM*), designed for portable computers. *SO DIMM* modules have only 144 pins for 64-bit memories and 77 pins for 32-bit memories.

- modules in **RIMM** format (*Rambus Inline Memory Module*, also called *RD-RAM* or *DRD-RAM*) are 64-bit memories developed by Rambus. They have 184 pins. These modules have two locating notches to avoid risk of confusion with the previous modules.

Given their high transfer speed, RIMM modules have a thermal film which is supposed to improve heat transfer.

As for DIMMs, smaller modules also exist; they are known as **SO RIMM** (*Small Outline RIMM*), designed for portable computers. *SO RIMM* modules have only 160 pins.

DRAM PM

The **DRAM** (*Dynamic RAM*) is the most common type of memory at the start of this millennium. This is a memory whose transistors are arranged in a matrix in rows and columns. A transistor, coupled with a capacitor, gives information on a bit. Since 1 octet contains 8 bits, a DRAM memory module of 256 Mo will thus contain 256 * 2^10 * 2^10 = 256 * 1024 * 1024 = 268,435,456 octets = 268,435,456 * 8 = 2,147,483,648 bits = 2,147,483,648 transistors. A module of 256 Mo thus has a capacity of 268,435,456 octets, or 268 Mo! These memories have access times of 60 ns.

Furthermore, access to memory generally concerns data stored consecutively in the memory. Thus **burst mode** allows access to the three pieces of data following the first piece with no additional latency time. In this burst mode, time required to access the first piece of data is equal to cycle time plus latency time, and the time required to access the other three pieces of data is equal to just the cycle time; the four access times are thus written in the form X-Y-Y-Y, for example *5-3-3-3* indicates a memory for which 5 clock cycles are needed to access the first piece of data and 3 for the subsequent ones.

DRAM FPM

To speed up access to the DRAM, there is a technique, known as **paging**, which involves accessing data located in the same column by changing only the address of the row, thus avoiding repetition of the column number between reading of each row. This is known as **DRAM FPM** (*Fast Page Mode*). FPM achieves access times of around 70 to 80 nanoseconds for operating frequency between 25 and 33 Mhz.

DRAM EDO

DRAM EDO (*Extended Data Out*, sometimes also called *hyper-page*") was introduced in 1995. The technique used with this type of memory involves addressing the next column while reading the data in a column. This creates an overlap of access thus saving time on each cycle. EDO memory access time is thus around 50 to 60 nanoseconds for operating frequency between 33 and 66 Mhz.

Thus the RAM EDO, when used in burst mode, achieves 5-2-2-2 cycles, representing a gain of 4 cycles on access to 4 pieces of data. Since the EDO memory did not work with frequencies higher than 66 Mhz, it was abandoned in favor of the SDRAM.

SDRAM

The **SDRAM** (*Synchronous DRAM*), introduced in 1997, allows synchronized reading of data with the mother-board bus, unlike the EDO and FPM memories (known as *asynchronous*) which have their own clock. The SDRAM thus eliminates waiting times due to synchronization with the mother-board. This achieves a 5-1-1-1 burst mode cycle, with a gain of 3 cycles in comparison with the RAM EDO. The SDRAM is thus able to operate with frequency up to 150 Mhz, allowing it to achieve access times of around 10 ns.

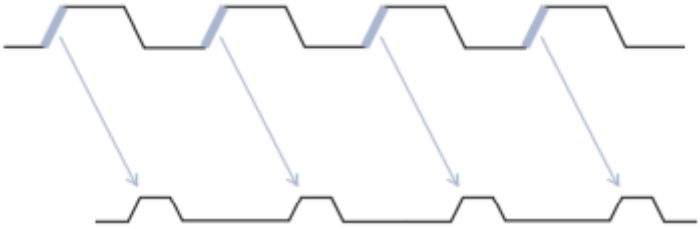
DR-SDRAM (Rambus DRAM)

The **DR-SDRAM** (*Direct Rambus DRAM*) is a type of memory that lets you transfer data to a 16-bit bus at frequency of 800Mhz, giving it a bandwidth of 1.6 GB/s. As with the SDRAM, this type of memory is synchronized with the bus clock to enhance data exchange. However, the RAMBUS memory is a proprietary technology, meaning that any company wishing to produce RAM modules using this technology must pay royalties to both RAMBUS and Intel.

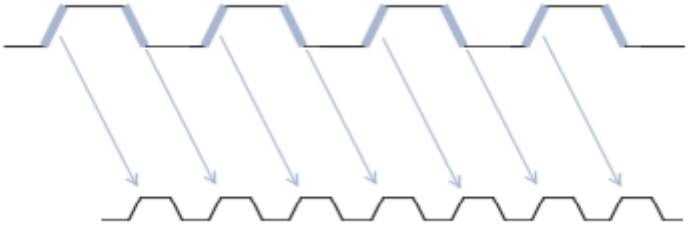
DDR-SDRAM

The **DDR-SDRAM** (*Double Data Rate SDRAM*) is a memory, based on the SDRAM technology, which doubles the transfer rate of the SDRAM using the same frequency.

Data are read or written into memory based on a clock. Standard DRAM memories use a method known as **SDR** (*Single Data Rate*) involving reading or writing a piece of data at each leading edge.



The DDR doubles the frequency of reading/writing, with a clock at the same frequency, by sending data to each leading edge and to each trailing edge.

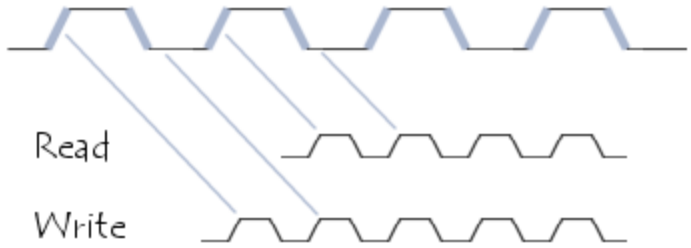


DDR memories generally have a product name such as PC*XXXX* where "*XXXX*" represents the speed in Mo/s.

DDR2-SDRAM

DDR2 (or DDR-II) memory achieves speeds that are twice as high as those of the DDR with the same external frequency.

QDR (*Quadruple Data Rate* or *quad-pumped*) designates the reading and writing method used. DDR2 memory in fact uses two separate channels for reading and writing, so that it is able to send or receive twice as much data as the DDR.



DDR2 also has more connectors than the classic DDR (240 for DDR2 compared with 184 for DDR).

summary table

The table below gives the equivalence between the mother-board frequency (FSB), the memory (RAM) frequency and its speed:

Memory	Name	Frequency (RAM) [!Frequency (FSB)	Speed	
DDR200	PC1600	200 MHz	100 MHz	1.6 GB/s
DDR266	PC2100	266 MHz	133 MHz	2.1 GB/s
DDR333	PC2700	333 MHz	166 MHz	2.7 GB/s
DDR400	PC3200	400 MHz	200 MHz	3.2 GB/s
DDR433	PC3500	433 MHz	217 MHz	3.5 GB/s
DDR466	PC3700	466 MHz	233 MHz	3.7 GB/s
DDR500	PC4000	500 MHz	250 MHz	4 GB/s
DDR533	PC4200	533 MHz	266 MHz	4.2 GB/s
DDR538	PC4300	538 MHz	269 MHz	4.3 GB/s
DDR550	PC4400	550 MHz	275 MHz	4.4 GB/s

DDR2-400	PC2-3200	400 MHz	100 MHz	3.2 GB/s
DDR2-533	PC2-4300	533 MHz	133 MHz	4.3 GB/s
DDR2-667	PC2-5300	667 MHz	167 MHz	5.3 GB/s
DDR2-675	PC2-5400	675 MHz	172.5 MHz	5.4 GB/s
DDR2-800	PC2-6400	800 MHz	200 MHz	6.4 GB/s

Synchronization (timings)

It is not unusual to see scores such as 3-2-2-2 or 2-3-3-2 to describe the parameterization of the random access memory. This succession of four figures describes the synchronization of the memory (*timing*), i.e. the succession of clock cycles needed to access a piece of data stored in the RAM. These four figures generally correspond, in order, to the following values:

- **CAS delay** or **CAS latency** (CAS meaning *Column Address Strobe*): this is the number of clock cycles that elapse between the reading command being sent and the piece of data actually arriving. In other words, it is the time needed to access a column.
- **RAS Precharge Time** (known as *tRP*, RAS meaning *Row Address Strobe*): this is the number of clock cycles between two RAS instructions, i.e. between two accesses to a row. operation.
- **RAS to CAS delay** (sometimes called *tRCD*): this is the number of clock cycles corresponding to access time from a row to a column.
- **RAS active time** (sometimes called *tRAS*): this is the number of clock cycles corresponding to the time needed to access a row.

The memory cards are equipped with a device called **SPD** (*Serial Presence Detect*), allowing the BIOS to find out the nominal setting values defined by the manufacturer. It is an EEPROM whose data will be loaded by the BIOS if the user chooses "auto" setting.

Error correction

Some memories have mechanisms for correcting errors to ensure the integrity of the data they contain. This type of memory is generally used on systems working on critical data, which is why this type of memory is found in servers.

Parity bit

Modules with parity bit ensure that the data contained in the memory are the ones required. To achieve this, one of the bits from each octet stored in the memory is used to store the sum of the data bits. The parity bit is 1 when the sum of the data bits is an odd number and 0 in the opposite case.

Thus the modules with parity bit allow the integrity of data to be checked but do not provide for error correction. Moreover, for 9 Mo of memory, only 8 will be used to store data since the last mega octet is used to store the parity bits.

ECC modules

ECC (*Error Correction Coding*) memory modules are memories with several bits dedicated to error correction (they are known as *control bits*). These modules, used mainly in servers, allow detection and correction of errors.

Dual Channel

Some memory controllers offer a dual channel for the memory. The memory modules are used in pairs to achieve higher bandwidth and thus make the best use of the system's capacity. When using the Dual Channel, it is vital to use identical modules in a pair (same frequency and capacity and preferably the same brand).

Source: http://en.kioskea.net/contents/409-random-access-memory-ram-or-pc-memory