A Basic
Overview of Commonly Encountered types of Random Access Memory (RAM)

ECE332
By:
Peter Haugen
Ian Myers
Bret Sadler
John Whidden

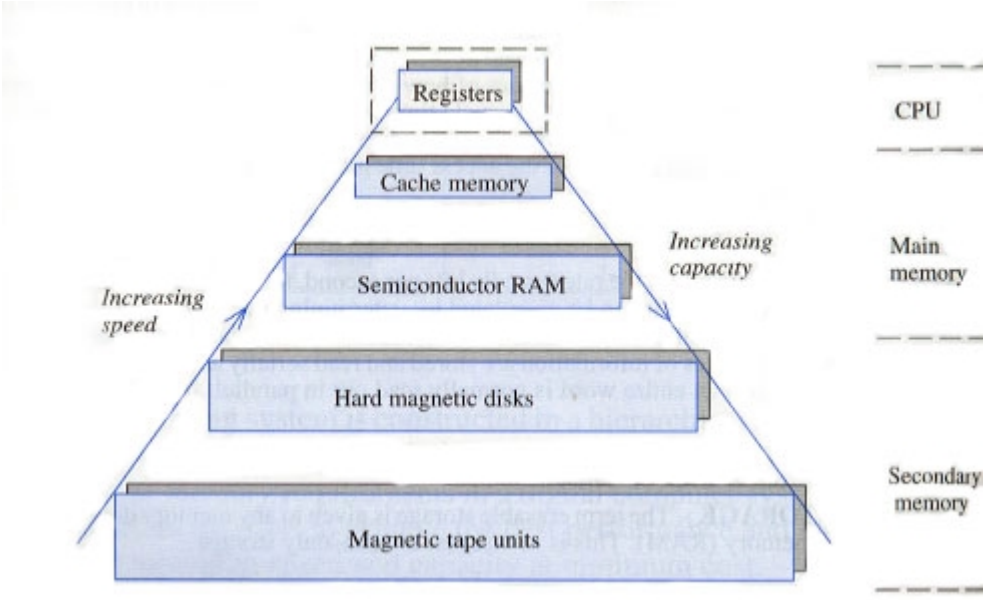

Figure 1: Memory Hierarchy (Simoncelli 2000)

Basic RAM Overview:

RAM (Random Access Memory) is the hardware location in a computer where the operating system, application programs, and data in current use are kept so that they can be quickly reached by the computer's processor. RAM is much faster to read from and write to than most other kinds of storage in a computer (the hard disk, floppy disk, and CD-ROM). However, the data in RAM stays there only as long as it has power. When you turn the computer off, RAM loses its data. When you turn your computer on again, your operating system and other files are once again loaded into RAM, usually from your hard disk.
RAM can be compared to a person's short-term memory and the hard disk to the long-term memory. The short-term memory focuses on work at hand, but can only keep so many facts in view at one time. If short-term memory fills up, your brain sometimes is able to refresh it from facts stored in long-term memory. A computer also works this way. If RAM fills up, the processor needs to continually go to the hard disk to overlay old data in RAM with new, slowing down the computer's operation. (Giakamozis 1999)

Why Random Access?

RAM is called "random access" because any storage location can be accessed directly. Originally, the term distinguished regular core memory from offline memory, usually on magnetic tape in which an item of data could only be accessed by starting from the beginning of the tape and finding an address sequentially. Perhaps it should have been called "non-sequential memory" because RAM access is hardly random. (Giakamozis 1999) RAM is organized and controlled in a way that enables data to be stored and retrieved directly to specific locations. A term IBM has preferred is *direct access* storage or memory. (Giakamozis 1999) Note that other forms of storage such as the hard disk and CD-ROM are also accessed directly (or "randomly") but the term *random access* is not applied to these forms of storage.
In addition to hard disk, floppy disk, and CD-ROM storage, another important form of storage is read-only memory (ROM), a more expensive kind of memory that retains data even when the computer is turned off. Every computer comes with a small amount of ROM that holds just enough programming (BIOS) so that the operating system can be loaded into RAM each time the computer is turned on.

What RAM architecture Looks Like

In general, RAM is much like an arrangement of cells in which each cell can hold a 0 or a 1. Each cell has a unique address that can be found by counting across columns and then counting down by row. To find the contents of a cell, the RAM controller sends the column/row address down a very thin electrical line etched into the chip. There is an *address line* for each row and each column in the set of cells. If data is being read, the bits that are read flow back on a separate *data line*. In describing a RAM chip or module, a notation such as 256Kx16 means 256 thousand columns of cells standing 16 rows deep.

An 8-megabyte module of dynamic RAM contains 8 million capacitors and 8 million transistors and the paths that connect them.
In the most common form of RAM, dynamic RAM, each cell has a charge or lack of charge held in something similar to an electrical capacitor. A transistor acts as a gate in determining whether the value in the capacitor can be read or written. In static RAM, instead of a capacitor-held charge, the transistor itself is a positional *flip/flop* switch, with one position meaning 1 and the other position meaning 0.
Externally, RAM is a chip that comes embedded in a personal computer motherboard with a variable amount of additional modules plugged into motherboard sockets. To add memory to your computer, you simply add more RAM modules in a prescribed configuration. These are single in-line memory modules (SIMMs) or dual in-line memory modules (DIMMs). Since DIMMs have a 64-bit pin connection, they can replace two 36-bit (32-bits plus 4 parity bits) SIMMs when synchronous DRAM is used. Laptop and notebook computers contain smaller 32-bit DIMMs known as small outline DIMMs (SO DIMMs).

How The Data In RAM Is Accessed

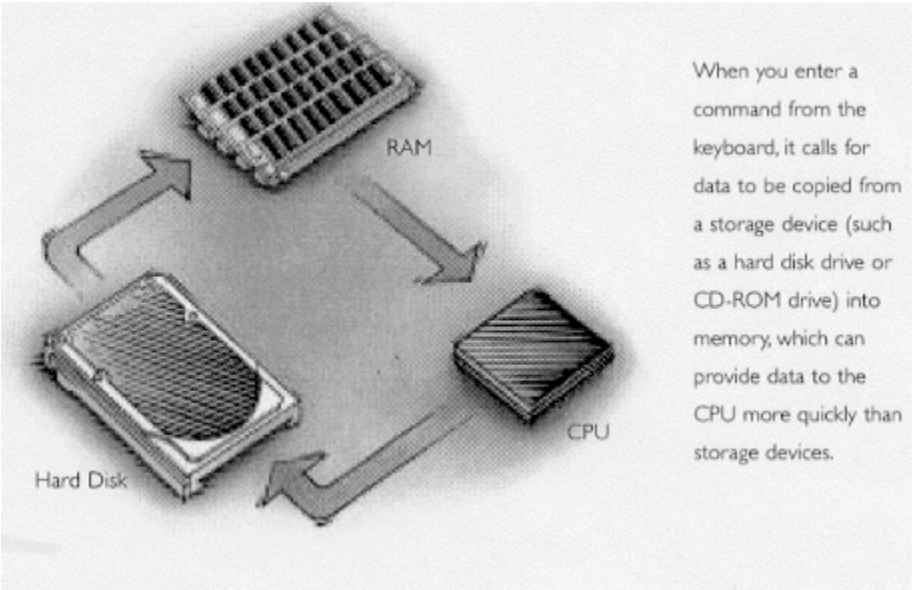

Figure 2: RAM Interaction (Simoncelli 2000)

When the processor or CPU gets the next instruction it is to perform, the instruction may contain the address of some memory or RAM location from which data is to be read (brought to the processor for further processing). This address is sent to the RAM controller. The RAM controller organizes the request and sends it down the appropriate address lines so that transistors along the lines open up the cells so that each capacitor value can be read. In DRAM a capacitor with a charge over a certain voltage level represents the binary value of 1 and a capacitor with less than that charge represents a 0. For dynamic RAM, before a capacitor is read, it must be power-refreshed to ensure that the value read is valid. Depending on the type of RAM, the entire line of data may be

read that the specific address happens to be located at or, in some RAM types, a unit of data called a *page* is read. The data that is read is transmitted along the data lines to the processor's nearby data buffer known as level-1 cache and another copy may be held in level-2 cache.
For video RAM, the process is similar to DRAM except that, in some forms of video RAM, while data is being written to video RAM by the processor, data can simultaneously be read from RAM by the video controller (for example, for refreshing the display image).

Common Types of RAM

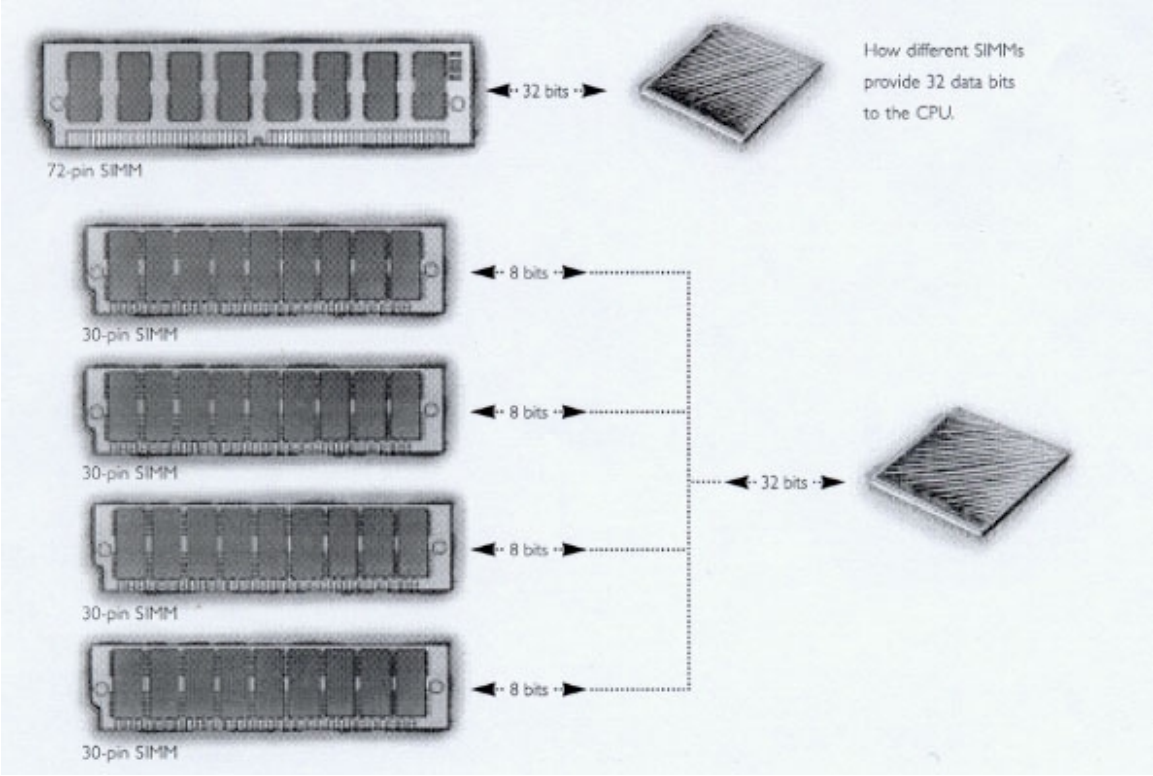

Figure 3: Common Types of RAM (Simoncelli 2000)

RAM is often divided into two main categories (1) main RAM, which stores every kind of data and makes it quickly accessible to a microprocessor and (2) video RAM, which stores data intended for your display screen, enabling images to get to your display faster.

Main RAM

Main RAM can be divided into static RAM (SRAM) and dynamic RAM (DRAM).

Static RAM (SRAM)

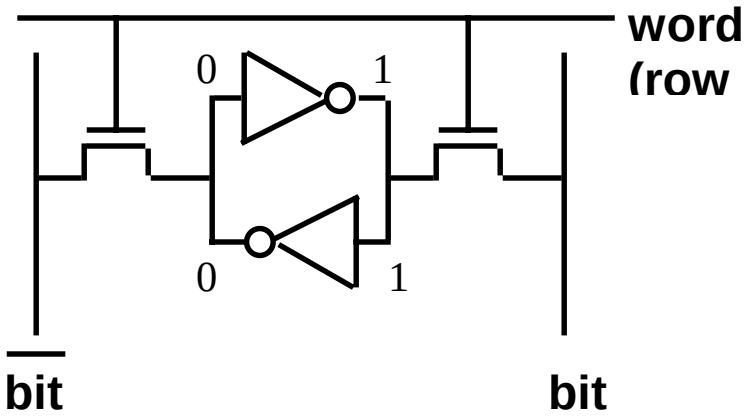

Figure 4: SRAM Cell (Kubiatowicz, 2001)

Static random access memory uses multiple transistors, typically four to six, for each memory cell but doesn't have a capacitor in each cell. It is used primarily for cache. Static RAM is more expensive, requires four times the amount of space for a given amount of data than dynamic RAM, but, unlike dynamic RAM, does not need to be power-refreshed and is therefore faster to access. One source gives a typical access time as 25 nanoseconds in contrast to a typical access time of 60 nanoseconds for dynamic RAM. (AMT International 2001) (More recent advances in dynamic RAM have improved access time.) Static RAM is used mainly for the level-1 and level-2 caches that the microprocessor looks in first before looking in dynamic RAM. (AMT International 2001)

Async Static RAM (ASRAM)

Async SRAM has been with us since the days of the 386, and is still in place in the L2 cache of many PCs. It's called *asynchronous* because it's not in sync with the system clock, and therefore the CPU must wait for data requested from the L2 cache. However, the wait isn't as long as it is with DRAM. (AMT International 2001)

Burst (or SynchBurst) Static RAM (BSRAM)

Burst SRAM (also known as SynchBurst SRAM) is synchronized with the system clock or, in some cases, the cache bus clock. This allows it be more easily synchronized with any device that accesses it and reduces access waiting time. It is used as the external level-2 cache memory for the Pentium II microprocessor chipset. (AMT International 2001) Async SRAM commonly used for L2 caches, with speeds of about 8.5 ns. Unfortunately, Sync SRAM isn't being produced in sufficient quantities to drive its cost down, so it seems destined for a relatively short life. That's especially true because it loses the ability to synchronize at bus speeds higher than 66 MHz. (AMT International 2001)

Pipeline Burst Static RAM (PB SRAM)

Using burst technology, SRAM requests can be *pipelined*, or collected so that requests within the burst are executed on a nearly instantaneous basis. PB SRAM uses pipelining, and while it's slightly behind system synchronization speeds, it's a possible improvement over Sync SRAM because it's designed to work well with bus speeds of 75 MHz and higher. (x2xtreme 2001)

SRAM Comparison

The following table shows the capabilities of various SRAM technologies at a number of external bus frequencies. The numbers in the table indicate the typical memory access achieved in terms of the number of wait states for initial and subsequent accesses. The shaded areas indicate the best performance/price points for the various bus speeds and technology. Note issue to keep in mind is that the most cost effective SRAM technology moves from asynchronous to flow-through synchronous to pipelined synchronous as bus frequencies increase. However, presently there are fewer suppliers with the underlying SRAM technology required for very fast flow-through synchronous SRAMs. As a result, in systems where performance is less critical designers are choosing pipelined synchronous for the 50 MHz to 66 MHz bus frequencies to increase the number of possible suppliers and hopefully, product availability. (x2xtreme 2001)

Bus Speed [MHz]	33	50	60	66	75	83	100	125
Async SRAM	2-1-1-1	3-2-2-2	3-2-2-2	3-2-2-2	3-2-2-2	3-2-2-2	3-2-2-2	3-2-2-2
Sync Burst SRAM	2-1-1-1	2-1-1-1	2-1-1-1	2-1-1-1	3-2-2-2	3-2-2-2	3-2-2-2	3-2-2-2
Pipelined Burst SRAM	3-1-1-1	3-1-1-1	3-1-1-1	3-1-1-1	3-1-1-1	3-1-1-1	3-1-1-1	3-1-1-1

Figure 5: SRAM Comparison (x2xtreme 2001)

Dynamic RAM (DRAM)

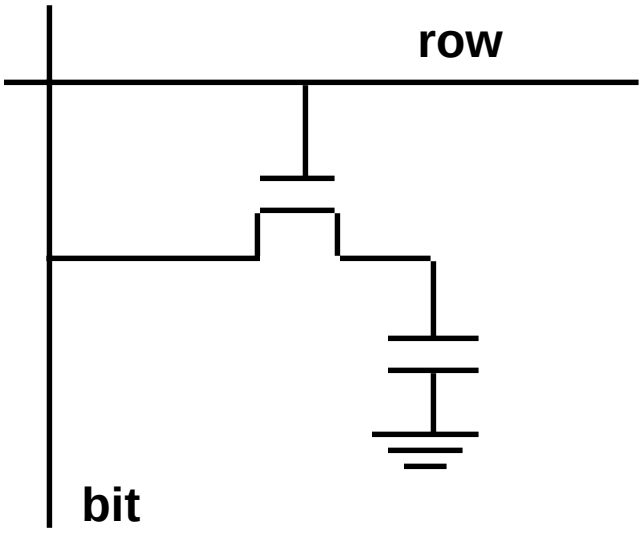

Figure 6: DRAM Cell (Kubiatowicz, 2001)

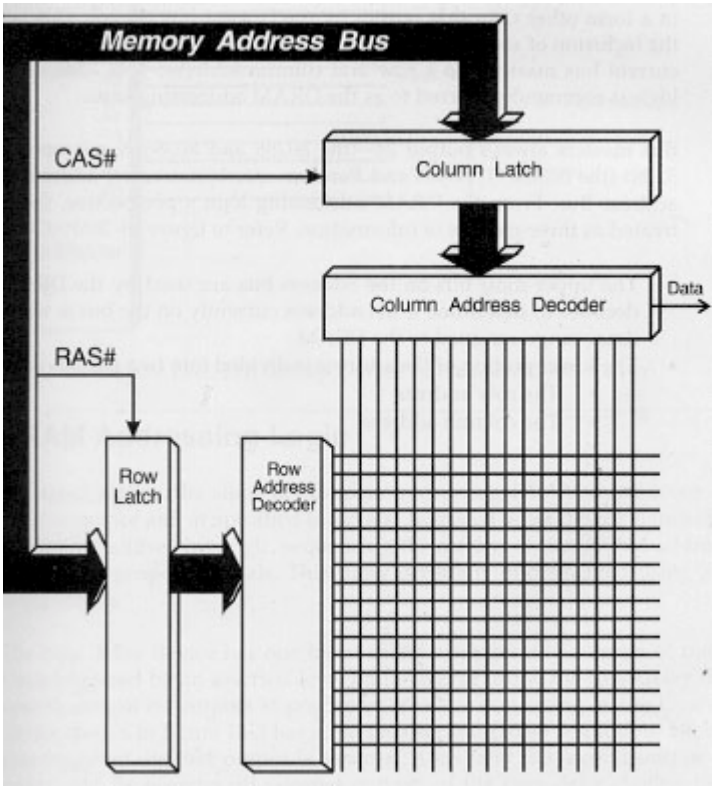

Figure 7: DRAM Architecture (Simoncelli 2000)

Dynamic random access memory has memory cells with a paired transistor and capacitor requiring constant refreshing. Because reading a DRAM discharges its contents, a power

refresh is required after each read. Apart from reading, just to maintain the charge that holds its content in place, DRAM must be refreshed about every 15 microseconds. This simple architecture makes DRAM the least expensive kind of RAM. (AMT International 2001)

Parity DRAM

Normally you assume 8 bits to one byte in memory. However, for many years, a ninth bit has been added as *parity* bit in the RAM blocks to verify correct transmission. That way you have to transmit 9 bits, to store 8 bits in the old 30 pin RAM chips. And it takes 36 bits to store 32 bits in the larger 72 pin chips, which increases the cost of the RAM chip by about 12%. If your motherboard required 36 bit modules, you had to use the larger parity modules. Fortunately, most system boards accepted 32 bit modules, so this was rarely a major concern. (AMT International 2001)

Fast Page Mode DRAM (FPM DRAM)

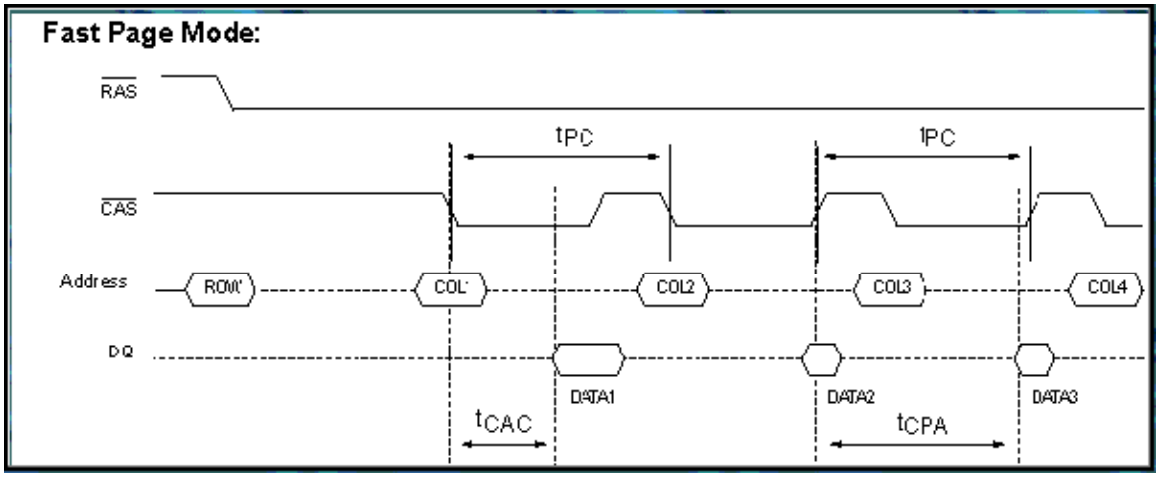

Figure 8: Fast Page Mode Timing Graph (Tang 1996)

A series of FPM read access's starts with the activation of a row in the DRAM array by providing a row address and bringing RAS LOW. Then, multiple column accesses may be executed by cycling CAS. Each CAS cycle includes applying a column address, bringing CAS LOW, waiting for valid data-out, latching data in the system and bringing CAS HIGH to prepare for the next cycle, in that order. This sequence of events is shown in the Figure. Note that CAS going HIGH disables the data outputs, and therefore must occur only after the valid data is latched by the system.

Fast page mode dynamic random access memory was the original form of DRAM. A timing graph of a typical FPM DRAM is shown above. It waits through the entire process of locating a bit of data by column and row and then reading the bit before it starts on the next bit. Maximum transfer rate to L2 cache is approximately 176 megabytes per second. FPM was the traditional RAM for PCs long before the EDO was introduced. It was usually mounted in SIMM modules of 2, 4, 8, 16, or 32 MB. Typically, it is found in 60 ns or 70 ns versions. Clock timings for FPM DRAM are typically 6-3-3-3 (meaning 3 clock cycles for access setup, and 3 clock cycles for the first and each of three successive accesses based on the initial setup). (Tang 1996)

Enhanced DRAM (EDRAM)

Enhanced DRAM (EDRAM) is the combination of SRAM and DRAM in a single package that is usually used for a level-2 cache. Typically, 256 bytes of static RAM is included along with the dynamic RAM. Data is read first from the faster (typically 15 nanoseconds) SRAM and if it is not found there, it is read from the DRAM, typically at 35 nanoseconds. By switching from FPM to EDO, one can expect a performance improvement of 2 to 5 percent. (Tang 1996)

ECC RAM

ECC RAM is a special error correcting RAM type. It is especially used in servers where data verification is essential. This type of RAM was never widely adopted by PC users due to its high cost and lower performance.

Extended Data Output RAM or DRAM (EDO RAM or EDO DRAM)

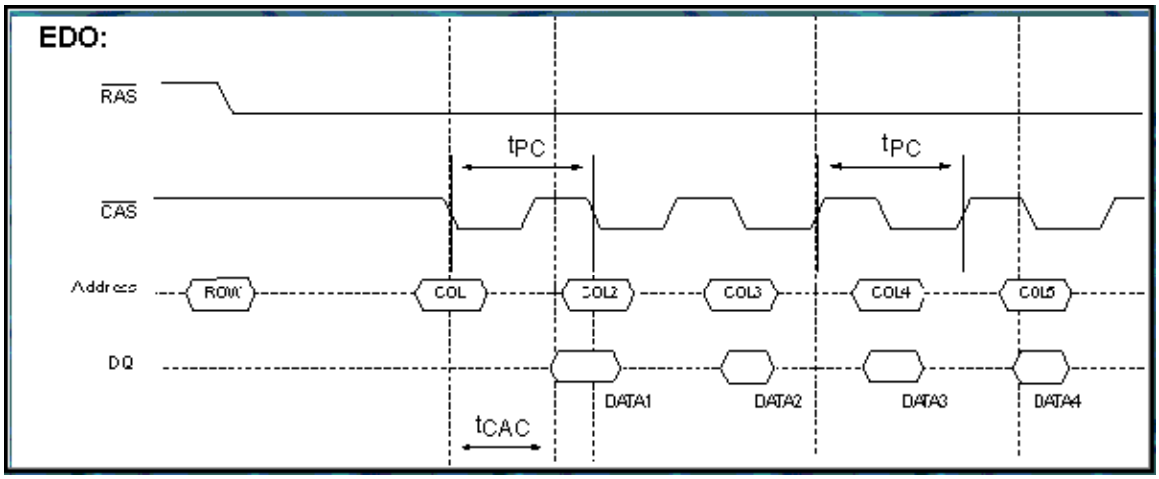

Figure 9: Extended Data Output RAM Timing Graph (Tang 1996)

Extended Data-Out page mode read access's are similar to those of FPM, with the exception that CAS going HIGH does not disable the data outputs, and the data latch is used to guarantee that valid data is held until CAS goes LOW again. In the case of EDO, the data latch (which is already available) is now controlled during page-mode accesses by CAS. Data is effectively captured in the latch as a result of CAS going HIGH. A new address can then be applied, and new data accessed in the array, without corrupting the output data from the previous access.

Despite the hype originally surrounding it, EDO RAM is no more than another type of FPM RAM. Essentially, it recognizes that most of the time when the CPU requests memory for a particular address, it's going to want some more addresses nearby. Instead of forcing each memory access to start afresh, EDO RAM hangs onto the location of the previous access, thereby speeding access to nearby addresses. EDO RAM speeds up the memory cycle, with improvements in memory performance of as much as 40 percent. (Tang 1996) On the down side, EDO RAM is originally effective only up to a bus speed of 66 MHz, and that’s quickly being bypassed by the most recent crop of AMD, Cyrix, and Intel processors.

Burst Extended Data Output DRAM (BEDO DRAM)

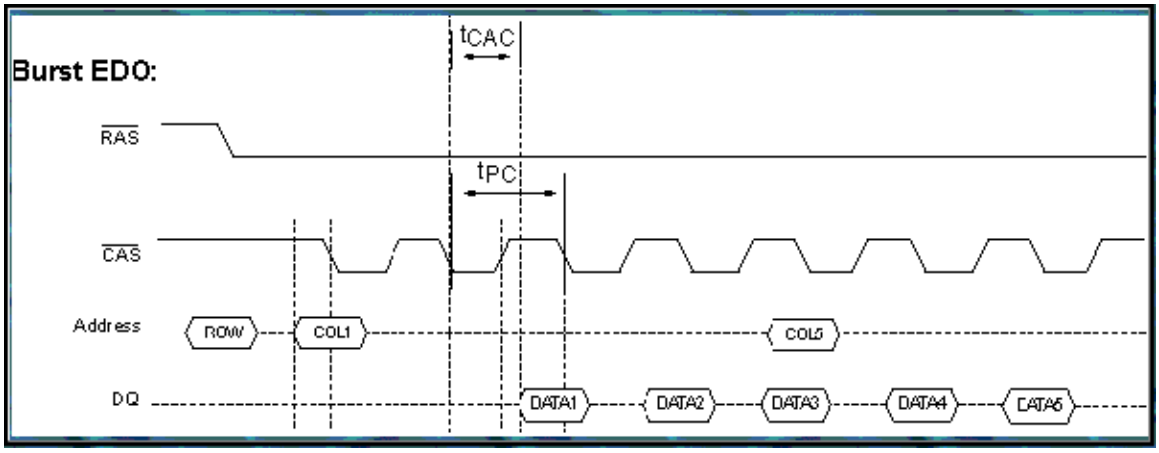

Figure 10: Burst EDO Timing Graph (Tang 1996)

BEDO read accesses differ from those of EDO in two ways. First, because the data latch is replaced by a register (i.e. an additional latch stage is added) data will not reach the outputs as a result of the first CAS cycle. The benefit of this internal pipeline stage is that data will appear in a shorter time from the activating CAS edge in the second cycle (i.e. tCAC is shorter). The second difference is that BEDO devices include an internal address counter so that only the initial address in a burst of four needs to be provided externally. The simplified functional representation for BEDO and the sequence of events is shown in the Figure.

As the need for faster access to DRAM has increased, technologies have been developed to provide it. One such technology is known as *bursting*, in which large blocks of data are sent and processed in the form of an uninterrupted "burst" of smaller units. What this means to DRAM is that the burst carries details not only about the address of the first page, but also of the next few. BEDO DRAM improves page mode DRAM by "building in" three successive column address shifts after the first column address is specified so that four bits are read as a burst. Together with a dual-bank architecture, BEDO DRAM promised to offer 4-1-1-1 access times. BEDO RAM can handle four data elements in one burst, and this allows the final three elements to avoid experiencing the delays of the first--all the addresses are ready to be processed. The DRAM is given the first address, and then can process the rest at a rate of 10 ns each. BEDO RAM, however, despite its substantial speed increase, still has difficulty moving past the 66-MHz bus barrier. BEDO RAM exists because SDRAM manufacturers were uninterested in pricing SDRAM to be competitive with EDO RAM; as a result, more work was done with EDO to add bursting technologies for speed rivaling that of SDRAM. (Tang 1996) However, because Intel and other manufacturers preferred SDRAM, BEDO DRAM was not widely used. (Tang 1996)

Multilevel DRAM (MLDRAM)

In MLDRAM, data is stored in cells that contain multiple bits. This is still an experimental architecture. Currently several two-bit-per-cell methods have been proposed. The figure on the left shows one of these methods. To extract the most significant bit (MSB), the cell voltage is compared to a reference voltage of $\frac{1}{2}V_{DD}$. If the cell voltage is larger, then the MSB is 1, and the cell voltage is then compared to a reference voltage of $5/6V_{DD}$ to find the least significant bit (LSB). If the cell voltage is smaller than $\frac{1}{2}V_{DD}$, the MSB is 0 and the cell voltage is compared to $1/6V_{DD}$ to find the LSB. The difficulty of building this four level MLDRAM, is the margins between each level is 1/3 of conventional DRAM. The challenge with MLDRAM will be to make it reliable. (Birk, 2001)

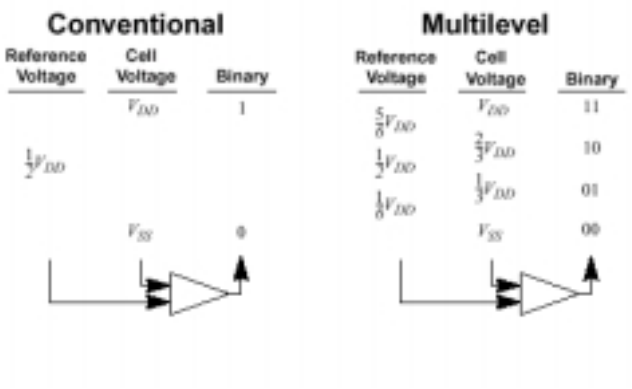

Figure 11: Multilevel DRAM (Birk, 2001)

Synchronous DRAM (SDRAM)

Synchronous dynamic random access memory takes advantage of the burst mode concept to greatly improve performance. It does this by staying on the row containing the requested bit and moving rapidly through the columns, reading each bit as it goes. The idea is that most of the time the data needed by the CPU will be in sequence. By keeping the RAM and the clock synchronized it increases the number of instructions that the processor can perform in a given time. SDRAM is about five percent faster than EDO RAM and is the most common form in desktops today. Maximum transfer rate to L2 cache is approximately 528 megabytes per second. (Spock 2001)

JEDEC SDRAM

JEDEC (Joint Electron Device Engineering Council) SDRAM is an industry standard synchronous DRAM. It has a dual-bank architecture and several burst mode accesses that can be preset. JEDEC SDRAM chips operate at either 83 MHz or 100 MHz. JEDEC SDRAM is also known as PC66 SDRAM because it was originally rated for 66 MHz bus operation and to distinguish it from Intel's PC100 architecture. (Kent 1998)

PC100 SDRAM

PC100 SDRAM is SDRAM that states that it meets the PC100 specification from Intel. (Kardo 1999) Intel created the specification to enable RAM manufacturers to make chips that would work with Intel's i440BX processor chipset. The i440BX was designed to achieve a 100 MHz system bus speed. Ideally, PC100 SDRAM would work at the 100 MHz speed, using a 4-1-1-1 access cycle. It's reported that PC100 SDRAM improves performance by 10-15% in an Intel Socket 7 system (but not in a Pentium II because its L2 cache speed runs at only half of processor speed). (Kardo 1999)

PC133 SDRAM

PC133 SDRAM has the same basic architecture as the PC100 SDRAM with the added ability to support a 133MHz bus speed. (Gary 2001)

PCxxx Comparison

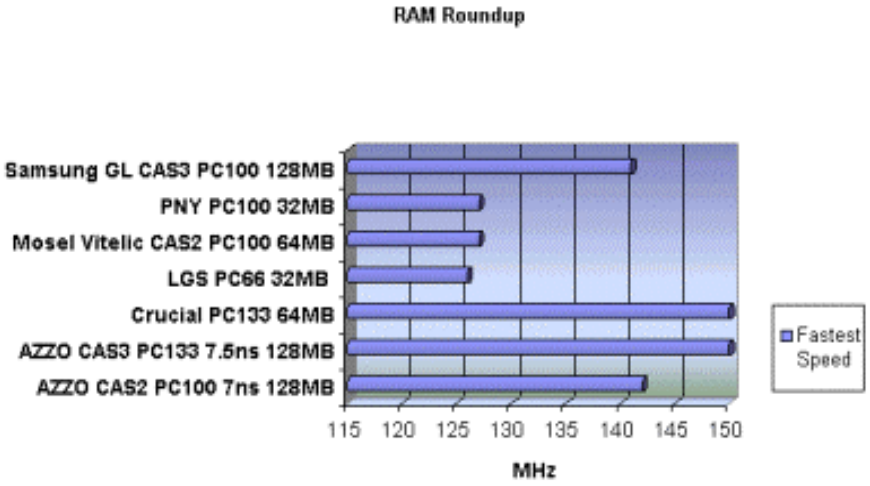

Figure 12: PCxx Series Comparison (Gary 2001)

Double Data Rate SDRAM (DDR SDRAM)

Double Data Rate SDRAM can theoretically improve RAM speed to at least 200 MHz. It activates output on both the rising and falling edge of the system clock rather than on just the rising edge, potentially doubling output. It's expected that a number of Socket 7 chipset makers will support this form of SDRAM. (Kent 1998)

Enhanced SDRAM (ESDRAM)

Enhanced SDRAM (ESDRAM), made by Enhanced Memory Systems, is a new JEDEC standard type of DRAM, which attempts to improve performance not by decreasing latency or increasing bus speed, but by improving the internal function of the RAM. The goal of ESDRAM is to obtain greater throughput by removing some of the internal delays of the standard SDRAM architecture. This is done by adding a row register cache

(RRC) which immediately absorbs the data from the Sense amps, freeing them to refresh the DRAM cell in parallel with the data release to the cache lines as opposed to in serial as it is usually done.

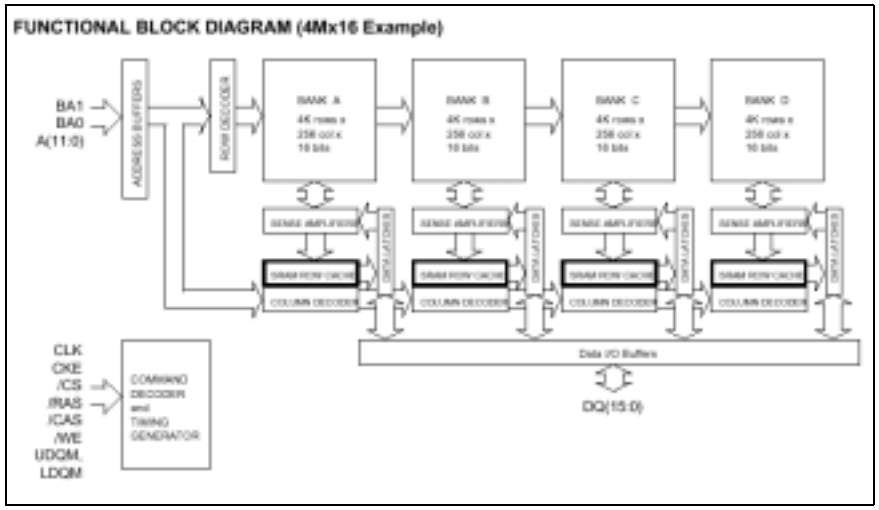

Figure 13: Enhanced SDRAM (Enhanced Memory Systems 2000)

The RRC is really nothing more than an extra bank of SRAM, which is used as a buffer to free up the sense amps so that the DRAM cell can be restored to its original state. In addition, a bit of primitive pipelining is used to increase throughput even further. At the release of the third word, a bank activate command can be issued and the next read command can occur during the release of the fourth word. This translates into, at most, a one-cycle delay when a page miss occurs. Random accesses to other banks don't even have any delay. What this means is that ESDRAM is able to utilize around 80% of the maximum theoretical bandwidth while standard SDRAM is around 40%-50%.(Enhanced Memory Systems 2000) ESDRAM is apparently competing with DDR SDRAM as a faster SDRAM chip for Socket 7 processors. (Tang 1996)

Enhanced DDRRAM

Enhanced DRRRAM, or EDDR uses a concept similar to that of ESDRAM to improve latency and power consumption when compared to a standard DDRRAM module. EDDR takes the approach of replacing the secondary sense amps used to transmit data to global output lines with a small SRAM cache. (Enhanced Memory Systems 2000)

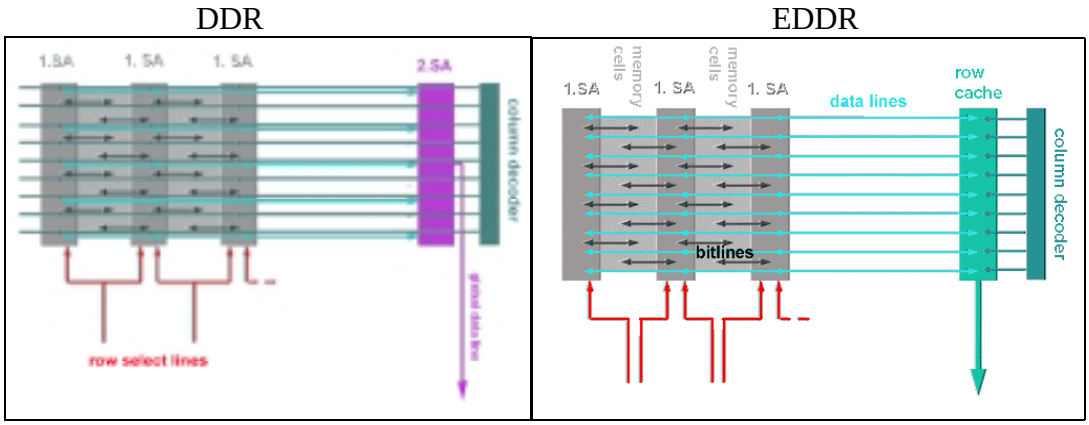

Figure 14: Enhanced DDRRAM (LostCircuits 2000)

When a row page is opened, the entire contents of the row are sent by the primary sense amps to the SRAM cache. Then, when a read command is serviced, the column decoder accesses the SRAM instead of activating the required column and waiting for the secondary sense amps to place the data on the global output lines (LostCircuits 2000).This approach eliminates one set of trace lines into the DRAM cells by attaching the column decoder to the SRAM cache instead. This lack of a 2-way trace delay will shave about 2ns off the CAS delay. By decreasing the latency, of the read process, EDDR decreases the time wasted on page misses by about 33%. (Davis 2000)

Rambus Dynamic Random Access Memory (RDRAM, DRDRAM)

Figure 15: RDRAM (Simoncelli 2000)

Rambus dynamic random access memory is a radical departure from the previous DRAM architecture. Designed by Rambus in partnership with Intel, RDRAM uses a Rambus in-

line memory module (RIMM), which is similar in size and pin configuration to a standard DIMM. (Rambus 2001) What makes RDRAM so different is its use of a special high-speed data bus called the Rambus channel. Direct Rambus (DRDRAM) provides a two-byte (16-bit) bus rather than DRAM's 8-bit bus or SDRAM's 64-bit bus. At a RAM speed of 800 megahertz (800 million cycles per second), the peak data transfer rate is 1.6 billion bytes per second. Direct Rambus uses pipelining to move data from RAM to cache memory levels that are closer to the microprocessor or display. Up to eight operations may be underway at the same time. Rambus is designed to fit into existing motherboard standards. (Kardo 1999)

SyncLink DRAM (SLDRAM)

A new standard for SDRAM is being developed by the SCIzzL Association at Santa Clara University (California) along with many industry leaders. (Randall 1997) Called SLDRAM, this technology improves on SDRAM by offering a higher bus speed and by using packets (small packs of data) to take care of address requests, timing, and commands to the DRAM. The result is less reliance on improvements in DRAM chip design, and ideally a lower-cost solution for high-performance memory. (Randall 1997) SyncLink DRAM is, along with Direct Rambus DRAM (DRDRAM), a protocol-based approach. In this approach, all signals to RAM are on the same line (rather than having separate CAS, RAS, address, and data lines). Since access time does not depend on synchronizing operations on multiple lines, SLDRAM promises RAM speed of up to 800 MHz. Like Double Data Rate SDRAM, SLDRAM can operate at twice the system clock rate. SyncLink is an open industry standard that is expected to compete and perhaps prevail over Direct Rambus DRAM.

Virtual Channel RAM

Most all modern operating systems are capable of running multiple tasks or threads at one time. When running multiple threads at once, each thread is a memory master, meaning it has its own address locality and tends to access memory in a fairly contiguous manner. However, since many threads are running at once, the RAM bank ends up opening and closing pages frequently even though the same basic address spaces are being accessed. This is known as thrashing. Virtual Channel Memory architecture, however, is designed for multithreading. VCRAM provides each thread with its own Virtual Channel: basically a dedicated port to RAM which allows each thread to access its own virtual memory space as if it were the only thread accessing the RAM bank, with up to 16 channels available. (NEC 1998)

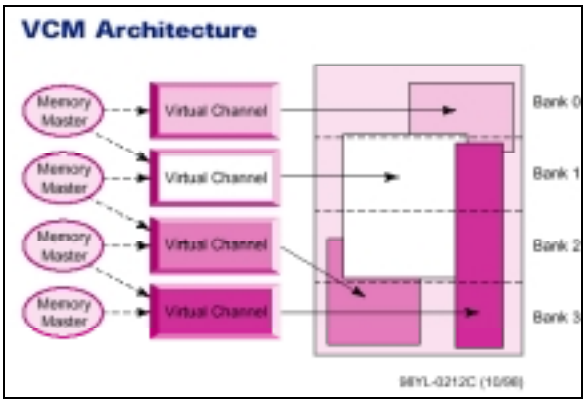

Figure 16: Virtual Channel RAM (NEC Electronics 1998)

The Virtual Channels minimize the latency resulting from other thread accesses and facilitate automatic data sharing without added external memory management. Each channel is equipped with a data row buffer and its own independent operating modes. The system memory controller associates these channels to their respective threads, thereby tracking the accesses of these threads. The system memory controller has complete controls over the operations of the channels. The Read/Write operations of the channels can operate independently of the DRAM bank operations of loading row data to channel and writing channel data to row. The system memory controller can schedule these memory events dynamically to avoid thrashing and unnecessary row accesses. It can also recognize when needed data for one channel is already loaded into another and share it appropriately. (NEC 1998)

Cached DRAM (CDRAM)

The CDRAM architecture incorporates an SRAM cache with DRAM. The separate terminals of the two RAM types allow independent control of each. This type of DRAM has been useful in graphics applications and PDAs. It also can be used as cache and main memory. (Kumanoya, 1995)

Video RAM

Video RAM as "video RAM" means in general all forms of RAM used to store image data for the video display monitor. Somewhat confusingly, the most common type of video RAM is called Video RAM (VRAM). All types of video RAM are special arrangements of dynamic RAM (DRAM). (Randall 1997) VRAM is used to store the pixel values of a graphical display, and the board's controller reads continuously from this memory to refresh the display. Its purpose is not only to give you faster video performance than you'd get with a standard video board, but also to reduce strain on the CPU. VRAM is dual-ported memory; there are two access ports to the memory cells, with one used to constantly refresh the display and the other used to change the data that will be displayed. Two ports means a doubling of bandwidth, and faster video performance as a result. By comparison, DRAM and SRAM have only one access port. Video RAM is really a buffer between the processor and the display monitor and is often

called the frame buffer. When images are to be sent to the display, they are first read by the processor as data from some form of main storage RAM and then written to video RAM. From video RAM (the frame buffer), the data is converted by a RAM digital-to-analog converter (RAMDAC) into analog signals that are sent to the display presentation mechanism such as a cathode ray tube (CRT). Usually, video RAM comes in a 1 or 2 megabyte package and is located on the video or graphics card in the computer. (Randall 1997)

RAMDAC

RAMDAC (random access memory digital-to-analog converter) is a microchip that converts digital image data into the analog data needed by a computer display. A RAMDAC microchip is built into the video adapter in a computer. It combines a small static RAM (SRAM) containing a color table with three digital-to-analog converters (DACS) that change digital image data into analog signals that are sent to the display's color generators, one for each primary color - red, green, and blue. (Randall 1997) In a cathode ray tube (CRT) display, an analog signal is sent to each of three electron guns. With displays using other technologies, the signals are sent to a corresponding mechanism.
The SRAM part of the RAMDAC contains a color palette table. A logical color number in the digital data input to SRAM is used to generate three separate values obtained from the table - one for each of red, green, and blue - that are output to one of three digital-to-analog converters. The analog signal output from the converter is input directly to the display electron guns or other image projecting mechanisms. For displays with true color, the digital color data is fed directly to the DACs, bypassing the SRAM table, which is not needed. (Randall 1997)

Multiport Dynamic Random access memory (MPDRAM)

MPDRAM, multiport dynamic random access memory (MPDRAM) is another subset of Video RAM. This type of RAM used specifically for video adapters or 3-D accelerators. The "multiport" part comes from the fact that MPDRAM normally has both random access memory and serial access memory. MPDRAM is located on the graphics card and comes in a variety of formats, many of which are proprietary. The amount of MPDRAM is a determining factor in the resolution and color depth of the display. MPDRAM is also used to hold graphics-specific information such as 3-D geometry data and texture maps. (Randall 1997)

Window RAM

Window RAM (WRAM), unrelated to Microsoft Windows, is very high-performance video RAM that is *dual-ported* and has about 25% more bandwidth than VRAM, in addition to several graphics features that applications developers can exploit, but costs less. Like VRAM, WRAM is a dual-ported type of RAM and it is used exclusively for graphics performance. Additional graphics features include a double-buffering data system several times faster than VRAM's buffer, resulting in considerably faster screen

refresh rates. It is often used for very high resolution (such as 1600 by 1200 pixels) projection using true color. (Randall 1997)

Synchronous Graphics RAM

Unlike VRAM and WRAM, and despite the fact that its primary use is on video accelerator cards, SGRAM is a single-ported RAM type. It speeds performance through a dual-bank feature, in which two memory pages can be opened simultaneously; it therefore approximates dual-porting. SGRAM is proving to be a significant player in 3-D video technology because of a block-write feature that speeds up screen fills and allows fast memory clearing. (AMT International 2001) Three-dimensional video requires extremely fast clearing, in the range of 30 to 40 times per second. Other features of this type of RAM include clock-synchronization and a technique called *masked write*, which enables selected data to be modified in a single operation rather as a sequence of read, update, and write operations. (AMT International 2001)

3D-RAM

3D-RAM is a memory chip that is optimized for high performance graphics systems. It stores 3 dimensional data and uses Z-compare and alpha blend units. The bandwidth for these chips are as high as 400 Mbytes/s externally, and 1.6 Gbytes/s internally. This type of RAM is not as effective for non-graphics intensive systems. (Kumanoya, 1995)

Multibank Dynamic RAM

Multibank Dynamic RAM (MDRAM) is a high-performance RAM, developed by MoSys, which divides memory into multiple 32 KB parts or "banks" that can be accessed individually. (Tyson 2000) Traditional video RAM is monolithic; the entire frame buffer is accessed at one time. Having individual memory banks allows accesses to be interleaved concurrently, increasing overall performance. It's also cheaper since, unlike other forms of video RAM, cards can be manufactured with just the right amount of RAM for a given resolution capability instead of requiring it to be in multiples of megabytes. (Tyson 2000)

Nonvolatile RAM (NVRAM)

Nonvolatile RAM (NVRAM) is a special kind of RAM that retains data when the computer is turned off or there is a power failure. Similar to the computer's read-only memory (ROM), NCRAM is powered by a battery within the computer. It can also work by writing its contents to and restoring them from an EEPROM.

Ferroelectric Random Access Memory

A type of non-volatile read/write random accesses semiconductor memory. FRAM combines the advantages of SRAM - writing is roughly as fast as reading, and EPROM non-volatility and in-circuit programmability. (Tyson 2000) Current disadvantages are

high cost and low density, but that may change in the future. A ferroelectric memory cell consists of a ferroelectric capacitor and a (metal oxide semiconductor) MOS transistor. Its construction is similar to the storage cell of a DRAM. The difference is in the dielectric properties of the material between the capacitor's electrodes. This material has a high dielectric constant and can be polarized by an electric field. The polarization remains until an opposite electrical field reverses it. This makes the memory non-volatile. FRAM has similar applications to EEPROM, but can be written much faster. The simplicity of the memory cell promises high-density devices, which can compete with DRAM.

How RAM Effectiveness is measured

The amount of time that RAM takes to write data or to read it once the request has been received from the processor is called the *access time*. Typical access times vary from 9 nanoseconds to 70 nanoseconds, depending on the kind of RAM. Although fewer nanoseconds access is better, user-perceived performance is based on coordinating access times with the computer's clock cycles. Access time consists of latency and *transfer time*. Latency is the time to coordinate signal timing and refresh data after reading it.

Peak Bandwidth:

Here you see the maximal peak bandwidth of the three well-known RAM types. The figure illustrates the absolutely maximal transfer from RAM to the L2-cache - in peaks, not as continuously transferred.

RAM type	Max. Peak bandwidth
FPM	176 MB/sec
EDO	264 MB/sec
SD	528 MB/sec

Figure 17: Peak Bandwidth (Kardo 1999)

DDR RAM:

RAM type	Bandwidth
SDRAM 100 MHz	100 MHz X 64 bit= 800 MB/sec
SDRAM 133 MHz	133 MHz X 64 bit= 1064 MB/sec
DDRAM 100 MHz	2 X 100 MHz X 64 bit= 1600 MB/sec
DDRAM 133 MHz	2 X 133 MHz X 64 bit= 2128 MB/sec
RDRAM 800 MHz	800 MHz X 16 bit= 1600 MB/sec

Figure 18: Peak Bandwidth of DDR RAM Types (Kardo 1999)

The RAM Table

RAM Technology	Application and Computer Location	Access Speed Range	Ports	Characteristics
Static RAM (SRAM)	Level-1 and level-2 cache memory Also used in RAMDAC	Fast	One	RAM that is continually charged More expensive than DRAM
Burst SRAM (BSRAM)	Level-2 cache memory	Fast	One	SRAM in burst mode
DRAM	Main memory Low-cost video memory	Slow	One	A generic term for any kind of dynamic (constantly recharged) RAM
FPM (Fast Page Mode) DRAM	Main memory Low-cost video memory	Slow	One	Prior to EDO DRAM, the most common type of DRAM
EDO (Extended Data Out) DRAM	Main memory Low-cost video memory	5-20% faster than FPM DRAM	One	Uses overlapping reads (one can begin while another is finishing) Currently, the most common type of DRAM
BEDO (Burst Extended Data Out) DRAM	Main memory and low-cost video	Faster than EDO DRAM 4-1-1-1 at 66 MHz	One	Not widely used because not supported by processor chipset makers
EDRAM (Enhanced DRAM)	Level-2 cache memory	15 ns access to SRAM 35 ns access to DRAM	One	Contains a 256-byte SRAM inside a larger DRAM
Nonvolatile RAM (NVRAM)	Preset phone numbers and profiles in modems	Fast	One	Battery-powered RAM

Synchronous DRAM (SDRAM)	Main memory	See specific forms of SDRAM Rated in MHz rather than nanoseconds	One	Generic term for DRAMs with a synchronous interface
JEDEC Synchronous DRAM (JEDEC SDRAM)	Main memory	Fast	One	Dual-bank architecture Burst mode Most common form of SDRAM
PC100 Synchronous DRAM (PC100 SDRAM)	Main memory	Intended to run at 100 MHz with 4-1-1-1 timing	One	An Intel specification designed to work with their i440BX
Double Data Rate Synchronous DRAM (DDR DRAM)	Main memory	Up to 200 MHz	One	Activates output on both the up and the down part of the clock cycle, doubling the data rate of PC100 SDRAM
Enhanced Synchronous DRAM (ESDRAM)	Main memory	Fast (100 MHz +)	Two	Twice as fast as SDRAM See Enhanced Memory Systems (EMS)
SyncLink DRAM (SLDRAM)	Main memory	Fastest (200 MHz +)	One	Open protocol-based design Uses "packets" for address, data, and control signals
Direct Rambus DRAM (DRDRAM)	Main memory	Up to 800 MHz but with a 16-bit bus width	One	Backed by Intel and Rambus Inc.
RAMDAC	Video card	Fast	One	Uses a small SRAM to store the color palette table used to provide data for digital-to-analog conversion

Rambus DRAM (RDRAM)	Video memory for Nintendo	Up to 600 MHz	One	Intel and Rambus Inc. architecture
Synchronous Graphics RAM (SGRAM)	Moderate to high-end video memory	Closer to VRAM than DRAM	One	Has special performance-enhancing features Example: Matrox Mystique
VRAM (Video RAM)	Higher-cost video memory	Twice the speed of DRAM	Two	Dual-ported, meaning a new image can be stored in RAM while a previous image is being sent to the display
WRAM (Window RAM)	Less expensive video memory	25% faster than VRAM	Two	With RAMDAC, can handle true color at 1600 by 1200 pixel resolution
Multibank DRAM (MDRAM)	Low-cost high-end video memory applications	Faster	One	Interleaved memory accesses between banks Memory has multiple 32 kilobyte banks that can be accessed independently Can be manufactured to fit the amount of memory the card requires No size-related performance penalty
Enhanced Synchronous DRAM (ESDRAM)	Main memory	Fast (100 MHz +)	Two	Twice as fast as SDRAM See Enhanced Memory Systems (EMS)
Enhanced DDR	Main memory	Fastest (200/266 MHz)	One	Uses fast SRAM cache in place of Sense Amps
Virtual Channel RAM	Main memory	100/133 MHz	One	Optimized for multi-threaded operation with up to 16 independent channels, one per thread.

Figure 19: RAM comparison table (Giakamozis 1999)

GLOSSARY: (Simoncelli 2000)

ACCESS TIME — The access time is the delay between the time the memory device receives an address and the time when the data from that address is available at the output of the memory. This is sometimes also referred to as the read time. The access time is usually constant for a particular random access memory (RAM) device and may be regarded as a specification of the speed of the RAM device.
For memory other than RAM, the access time is often regarded as the time from when an instruction is decoded asking for a memory location until the desired information is found but not read. In this case the access time is a function of the location of the data on the medium with reference to the position of the read - write transducer(s). It can therefore vary substantially depending on the location of the information being sought and as a consequence the average access time over a short period depends on the pattern of memory references.
BANDWIDTH - Memory bandwidth is the number of bytes per second that the memory can deliver to the processor. The bandwidth of a memory is also the amount of information made available to the processor by a single memory access.
CAPACITY - The capacity of a memory module or system is simply the maximum number of bits, bytes, or words within the module or system.
Example a 2K x 4 memory can store 2K (K = 1024 = 2^{1o}) words each containing 4 bits of data or a total of 2 x 1024 x 4 bits = 8192 bits.
CYCLE TIME - In general, the cycle time is the time interval for which a set of operations is repeated regularly in the same sequence. In the area of computers, the cycle time is the total time for a program instruction to reference a memory location, cad from or write to it, and then return to the next instruction.
The cycle time is also used as a measure of how often a memory can be accessed per unit time. For a static random access semiconductor memory (SRAM) the cycle time is equal to the access time; however, for a dynamic random access memory (DRAM) the cycle time is greater than the access time since a DRAM requires a frequent restore period. In this latter case the cycle time then consists of the access time plus any additional time after the completion of a memory access until the memory is available again.
It can be noted that the cycle time for a magnetic disk drive to deliver the contents of its disk is much greater than that of silicon random access memory.
DATA RATE - The data rate is the rate (usually bits per second, bytes per second or words per second), at which data can be read out of a storage device. It can be calculated by determining the product of the reciprocal of the access time and the number of bits in the unit of data (data word) being read. The term data rate is usually associated with nonrandom access memory, where large pieces of information are stored and read serially and has little significance in random access memory since in such memories an entire word is normally read out in parallel. In most cases the data rate is constant for a given memory module.
DENSITY - In the area of magnetic storage, density is the number of characters or bits that can be stored on either one inch of magnetic tape, or per track on a magnetic disk.
ERASABLE STORAGE - The term erasable storage is given to any memory device whose contents can be modified, e.g. random access memory (RAM). This is in contrast to read - only storage (ROM).

Bibliography:

AMT International. (January 2001) Your Complete Memory and Hardware Store. <http://www.amtiinc.com/>

Birk, G., D. G. Elliot, and B. F. Cockburn. (1999) A comparative simulation study of four multilevel DRAMs. <http://www.ee.ualberta.ca/~elliott/memory/conferences/1999mtdt_birk_mldram.pdf> (Retrieved 20 January 2001).

Davis, Brian. et. al. (June 2000) DDR2 and low latency variants <http://www.eecs.umich.edu/~btdavis/papers/>

Enhanced Memory Systems. (November 2000) ESDRAM Datasheet. <http://www.edram.com/Library/datasheets/SM2603,2604pb_r1.8.pdf>

Gary. (January 2001) RAM Roundup. *Overclockin.* <http://www.overclockin.com/reviews/RamRoundup/RamRoundup-3.html>

Geer, Will. (July 2000) Memory Guide. <http://www.3drage.com/guides/memory/>

Giakamozis, Con (July 1999) A RAM Guide. <http://cie.changwon.ac.kr/~cwjo/microlect/ramguide.htm>

Kardo, Michael. (July 1999) About RAM. <http://www.mkdata.dk/click/module2e.htm>

Kent, Dean. (October 1998) RAM Guide. *Tom's Hardware* <http://www.tomshardware.com/mainboard/98q4/981024/index.html>

Kubiatowicz, John. (Spring 2001) Lecture Notes, Copyright © 2001 UCR. <http://www-inst.eecs.berkeley.edu/~cs152/>

Kumanoya, M., T. Ogawa, and K. Inoue. 1995. Advances in DRAM interfaces. *IEEE Micro* 15.6: 30-36.

LostCircuits. (November 2000) Inside the EDDR Chip. *LostCircuits.* <http://www.lostcircuits.com/memory/>

NEC Electronics Inc. (1998) Virtual Channel Memory Product Brief <http://www.necel.com/home.nsf/ViewAttachments/M13154EU2V0PB00/$file/13154e20.pdf>

N. R. (April 1998) Everything you wanted to know about random access memory. <http://www.cyberwalker.net/columns/apr98/0423.html>

Rambus INC. (Janurary 2001) Rambus Home Page.
<http://www.rambus.com>

Randall, Neil. (10/21/97) A RAM Primer. *PC Magazine.*
<http://www.zdnet.com/pcmag/pctech/content/16/18/tu1618.001.html>

Risley, David. (January 2001) RAM Primer. *Hardware Central.*
<http://www.hardwarecentral.com/hardwarecentral/tutorials/52/3/>

Risley, David. (January 2001) Video Memory. *Hardware Central.*
<http://hardwarecentral.com/hardwarecentral/tutorials/68/2/>

Simoncelli, Federico. (July 2000) Types of RAM.
<http://www.scit.wlv.ac.uk/~c9921190/types_of_ram.htm>

Spock. (January 2001) RAM Information.
<http://www.spock.mem.net/computer/ram.html>

Tang, Stanley. (October 1996) Computer Info.
<http://www.core.binghamton.edu/~crsovr/webdoc9.htm>

Tyson, Jeff. (July 2000) How Random Access Memory (RAM) Works.
<http://www.howstuffworks.com/ram3.htm>

x2xtreme. (January 2001) RAM (Random Access Memory)
<http://x2xtreme.homeip.net/ram.htm>