Supplementary Notes 4:

Microprocessor & Random Access Memory

Microprocessor - The Brain of the Computer.

At the heart of all personal computers and most workstations sits a microprocessor. A microprocessor is designed to perform arithmetic and logic operations that make use of small number-holding areas called registers. Typical microprocessor operations include adding, subtracting, comparing two numbers, and fetching numbers from one area to another.

These operations are the result of a set of instructions that are part of the microprocessor design. When the computer is turned on, the microprocessor is designed to get the first instruction from the Basic Input/Output System (BIOS) that comes with the computer as part of its memory. After that, either the BIOS, or the operating system that BIOS loads into computer memory, or an application progam is "driving" the microprocessor, giving it instructions to perform.

Characteristics

Three basic characteristics differentiate microprocessors:

- **Instruction set**: The set of instructions that the microprocessor can execute.
- **Bandwidth**: The number of bits processed in a single instruction.
- **Clock speed**: Given in megahertz (MHz), the clock speed determines how many instructions per second the processor can execute.

In both cases, the higher the value, the more powerful the CPU. For example, a 32-bit microprocessor that runs at 50MHz is more powerful than a 16-bit microprocessor that runs at 25MHz.

In addition to bandwidth and clock speed, microprocessors are classified as being either **RISC** (reduced instruction set computer, eg: Motorola.) or **CISC** (complex instruction set computer, eg: Intel and AMD).

Clock speed

Also called clock rate, the speed at which a microprocessor executes instructions. Every computer contains an internal clock that regulates the rate at which instructions are executed and synchronizes all the various computer components. The CPU requires a fixed number of clock ticks (or clock cycles) to execute each instruction. The faster the clock, the more instructions the CPU can execute per second.

Clock speeds are expressed in megahertz (MHz), 1 MHz being equal to 1 million cycles per second. The CPUs of Personal computers have clock speeds of anywhere from 33 MHz to over 300 MHz.

The internal architecture of a CPU has as much to do with a CPU's performance as the clock speed, so two CPUs with the same clock speed will not necessarily perform equally. Whereas an Intel 80286 microprocessor requires 20 cycles to multiply two numbers, an Intel 80486 or later processor can perform the same calculation in a single clock tick. (Note that clock tick here refers to the system's clock, which runs at 66 MHz for all PCs.) These newer processors, therefore, would be 20 times faster than the older processors even if their clock speeds were the same. In addition, some microprocessors are superscalar, which means that they can execute more than one instruction per clock cycle.

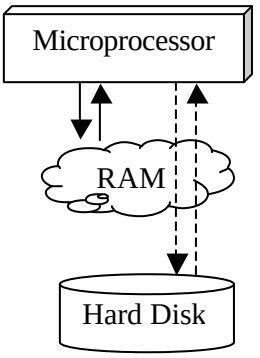
Like CPUs, expansion buses also have clock speeds. Ideally, the CPU clock speed and the bus clock speed should be the same so that neither component slows down the other. In practice, the bus clock speed is often slower than the CPU clock speed, which creates a bottleneck. This is why new local buses, such as AGP, have been developed.

Random Access Memory

RAM (random access memory) is the place 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 the 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 your computer is running. 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 (*eg: parent's birthday*) and the hard disk to the long-term memory (*eg: girlfriend's phone number*). 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. Unlike the hard disk which can become completely full of data so that it won't accept any more, RAM never runs out of memory. It keeps operating, but much more slowly than you may want it to.



How Big is RAM?

RAM is small, both in physical size (it's stored in microchips) and in the amount of data it can hold. It's much smaller than your hard disk. A typical computer may come with 32 million bytes of RAM (32 MB) and a hard disk that can hold 10 billion bytes. RAM comes in the form of "discrete" (meaning separate) microchips and also in the form of modules that plug into holes in the computer's motherboard. These holes connect through a bus or set of electrical paths to the processor. The hard drive, on the other hand, stores data on a magnetized surface that looks like a phonograph record.

Today's personal computers come with 64 or more megabytes of RAM (*except at your current college*), usually increasing in multiples of 8 megabytes. Users of graphic applications usually need 64 or 96 megabytes of memory. Most personal computers are designed to allow you to add additional RAM modules up to a certain limit (for example, up to 64 or 128 megabytes). Having more RAM in your computer reduces the number of times that the computer processor has to read data in from your hard disk, an operation that takes much longer than reading data from RAM. (RAM access time is in nanoseconds; hard disk access time is in milliseconds.)

Why Random Access?

RAM is called "random access" because any storage location can be accessed directly. RAM is organized and controlled in a way that enables data to be stored and retrieved directly to specific locations. 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 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 so that the operating system can be loaded into RAM each time the computer is turned on.

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 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.

Kinds of RAM

RAM can be divided into (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.

(1) Main RAM

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

- **Static RAM (SRAM)**

  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. (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.

- **Dynamic RAM (DRAM)**

  Dynamic RAM uses a kind of capacitor that needs frequent power refreshing to retain its charge. 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. DRAM is the least expensive kind of RAM.

Fast Page Mode DRAM (FPM DRAM)

Prior to newer forms of DRAM, Fast Page Mode DRAM (FPM DRAM) was the most common kind of DRAM in personal computers. *Page mode* DRAM essentially accesses a row of RAM without having to continually respecify the row. A row access strobe (RAS) signal is held active while the column access strobe (CAS) signal changes to read a sequence of contiguous cells. This reduces access time and lowers power requirements. 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).

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

Extended Data Output RAM (EDO RAM) or Extended Data Output Dynamic RAM (EDO DRAM) is up to 25% faster than standard DRAM and reduces the need for level-2 cache memory.

Synchronous DRAM (SDRAM)

Synchronous DRAM (SDRAM) is a generic name for various kinds of DRAM that are synchronized with the clock speed that the microprocessor is optimized for. This tends to increase the number of instructions that the processor can perform in a given time. The speed of SDRAM is rated in MHz rather than in nanoseconds (ns). This makes it easier to compare the bus speed and the RAM chip speed. You can convert the RAM clock speed to nanoseconds by dividing the chip speed into 1 billion ns (which is one second). For example, an 83 MHz RAM would be equivalent to 12 ns.

PC100 SDRAM

PC100 SDRAM is SDRAM that states that it meets the PC100 specification from Intel. 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 will improve 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).

Enhanced SDRAM (ESDRAM)

Enhanced SDRAM (ESDRAM), made by Enhanced Memory Systems, includes a small static RAM (SRAM) in the SDRAM chip. This means that many accesses will be from the faster SRAM. In case the SRAM doesn't have the data, there is a wide bus between the SRAM and the SDRAM because they are on the same chip. ESDRAM is apparently competing with DDR SDRAM as a faster SDRAM chip for Socket 7 processors.

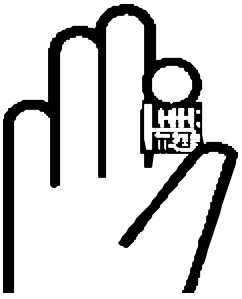
Direct Rambus DRAM (DRDRAM)

Direct Rambus DRAM (DRDRAM) is a proprietary technology proposed by Rambus, Inc. in partnership with Intel. Like SLDRAM, it promises RAM speed up to 800 MHz. It has a smaller bus width (16 bits compared to 64 bits) than current SDRAM designs.

(2) 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). 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. Most forms of video RAM are *dual-ported*. While the processor is writing a new image to video RAM, the display is reading from video to refresh its current display content. The dual-port design is the main difference between main storage RAM and video RAM.

Adapted from:
webopedia.com and **whatis.com**



The important thing is never to stop questioning
Albert Einstein