SIMMs & OTHER MEMORY MODULES

In early generations of personal computer, the computer's RAM (random access, or read-write memory) was in the form of groups or 'banks' of standard DIL (dual-in-line) IC chips that were either soldered directly to the mother board or plugged into sockets on it. However as later computers began to need more and more memory, plug-in modules were developed as convenient replacements for this 'chip level' memory.

Most modern PCs are designed to have their memory in the form of these plug-in memory modules. The same modules are also used to expand the memory inside laser printers, etc. However it's easy to get confused, because there are now both a number of different types of module, and also many different kinds of memory chip used in them.

In this data sheet we'll try to give you a good basic understanding of memory modules and the technology in them, so you'll hopefully be able to pick the right module for your application. We'll look at the main types of memory module, in roughly the order they were developed. But first, let's clarify a few general terms you'll come across in connection with computer memory.

Bits and bytes

You're probably already aware that inside a computer the information is handled entirely in 1's and 0's — i.e., in binary digits, or *bits*. One bit is the smallest possible amount of information, and can only have a value of either 1 or 0.

To handle appreciable amounts of information efficiently, computers don't try to handle every bit individually. Instead they group them into 8-bit 'parcels' known as *bytes*. The first PCs (and many microcontrollers) did all of their arithmetic and data manipulation with these 8-bit bytes, and are therefore called '8 bit' computers.

Later PCs like the 80286-based 'AT' models handled two 8-bit bytes at the same time, and are not surprisingly called '16-bit' computers.

When 486-based PCs came along, these manipulated the data in four-byte words, and are thus described as '32-bit' computers.

And lastly (for the present) there are the even newer Pentium/PII/PIII class computers, which manipulate the data in 8-byte words or larger. These are therefore described as '64-bit' or '128-bit' computers, as the case may be.

Kilobytes & megabytes

Because modern computer memories are designed to store very large amounts of data, their storage capacity is usually given in terms of *kilobytes* or *megabytes* — same terms used to describe the capacity of floppy disks and hard disk drives, etc. But remember that in computer parlance, the prefixes 'kilo' and 'mega' are the *binary* versions, not the decimal variety.

So a *kilobyte* is 1024 bytes, not 1000. Similarly a megabyte is 1024 kilobytes, or 1,048,576 bytes. It's good practice to use 'KB' and 'MB' to represent kilobytes and megabytes respectively, as a reminder that we're using the binary multipliers and also talking about bytes rather than individual bits.

SRAM, DRAM and ROM

Strictly speaking all memory used in modern computers is *random-access memory* or 'RAM', meaning that the data in any particular memory location or cell is just as accessible as any other data, merely by specifying its unique 'memory address'. However by convention the term 'RAM' is only used to signify *read-write* random access memory, where data can be either written to or read from any address at random.

This is mainly to distinguish read-write memory from *read-only* memory or 'ROM', where the data can only be written to (i.e., stored in) each location once, and from then on only read out when it's needed. ROMs are still randomly accessible, of course.

As it happens there are also two main types of RAM chip: the *static RAM* or SRAM, and the *dynamic RAM* or DRAM.

SRAMs are essentially arrays of transistor flip-flops, each of which can store one bit of information indefinitely as long as the chip is supplied with power. Because of their complexity SRAMs are fairly expensive, but they're also capable of the fastest operation. That's why they're used in a computer's *cache memory* — the relatively small but very fast 'scratchpad' memory used to provide frequently-needed information to the CPU much faster than the main memory can supply it.

In contrast, the memory cells in DRAMs are little more than tiny MOS storage capacitors, where the bits are stored as the presence or absence of electrical charge. This makes DRAMs much cheaper memory than SRAMs, but because the charges tend to 'leak away' after a short time, the data in a DRAM must be regularly 'refreshed' (by reading and then re-writing it) every millisecond or so. Hence the term 'dynamic', reflecting the fact that the data in a DRAM needs constant attention to remain stored faithfully.

Despite this shortcoming, DRAMs are the cheapest type of RAM and are also capable of storing much more data in a given chip area than SRAMs. That's why DRAMs are used in the 'main memory' of modern computers, and

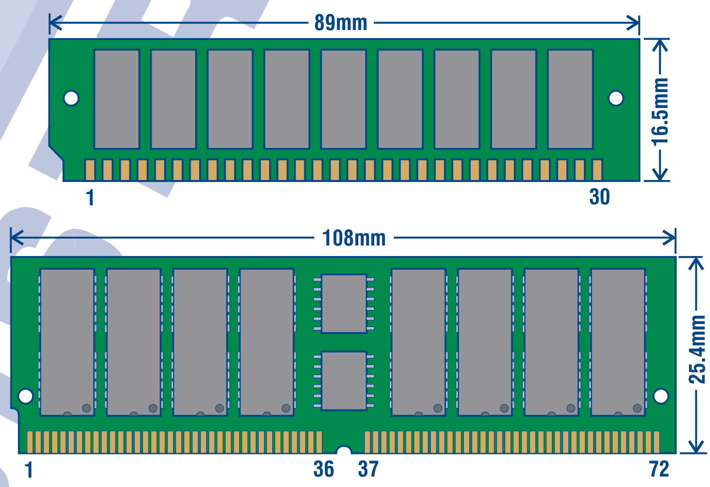

Fig.1: The two main kinds of SIMM that you'll come across. At top is the earlier 30-pin type, with the later (and still current) 72-pin type below it.

why it's this type of RAM chip (in one of a number of versions) that is used in the various memory modules we'll be looking at shortly.

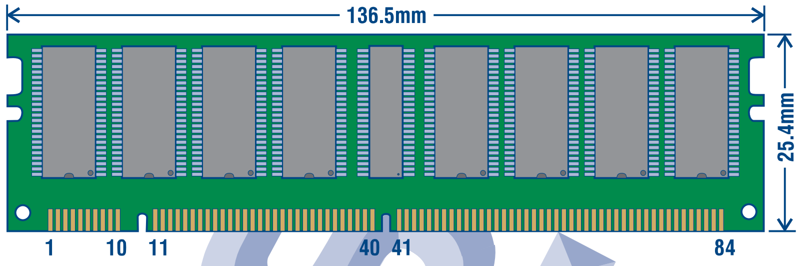

Fig.2: Larger again is the 168-pin DIMM, with 84 contacts on each side in three groups. DIMMs may look like SIMMs, but there's a crucial difference: the contacts on each side carry quite different signals. A DIMM can provide up to 512MB of memory.

Error checking & parity

Because it was difficult to make early DRAM chips that were 100% reliable in terms of data storage reliability, it was necessary to provide error-checking circuitry which would at least warn the computer's CPU if there was any suggestion that the data being read back from the memory chips wasn't identical to the data that had been stored in them.

In most cases the error-checking system used for this was *parity checking*, of either the 'odd' or 'even' type.

Parity checking involves storing an extra bit of data (the *parity* bit) along with every byte of 'real data'. The memory writing circuit checks the number of 1's in each data byte being stored, and makes the stored parity bit either a 1 or a 0 to make the *total* either an odd number (with odd parity) or an even number (with even parity). Then when the data is being read back out of the memory the circuitry checks the number of data 1's again and checks if they still make an odd (or even) number when the parity bit is added. If they don't, there's a strong chance that there has been a memory error — and the parity circuit will 'sound the alarm' to warn the CPU that the data has probably been corrupted.

Parity checking is thus a useful 'insurance' against

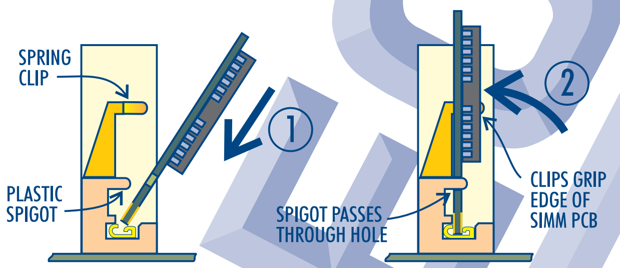

Fig.3: Fitting a SIMM into its socket is a two-step operation. First its contacts are push into the bottom contacts, at an angle; then it is swung up to a vertical position, with the spigots through the locating holes and the spring clips fastening the ends of the SIMM's PCB. The procedure is basically reversed to remove a SIMM from its socket.

memory chip faults, but it comes at a price: storing that ninth parity bit with each byte of real data makes the memory about 12.5% larger than it would otherwise be, and roughly that much more expensive as well. Or if you like, the memory is only 89% efficient, in terms of real data storage capacity.

As it became possible to manufacture memory chips of greater and greater reliability, the parity checking system was found largely unnecessary. The majority of modern PCs don't bother with it, at least as far as memory is concerned; they rely on the memory chips being 100% reliable. Network servers and other high-end computers that need 'absolute reliability' may use a more elaborate error checking system known as *Error Correction Code* (ECC), which can detect and correct single-bit errors.

Now let's look at specific memory modules.

30-pin SIMMs

This was the first type of plug-in memory module used in PCs. The name 'SIMM' is simply a contraction of *Single Inline Memory Module*, and it's essentially just a small PC board (89mm long by 16.5mm high) on which are mounted either eight or nine memory chips. A row of 30 contacts spaced at 0.1" (2.54mm) intervals along both sides of one long edge mate with the contacts of a socket on the mother board, when the SIMM is plugged in. (The contacts on both sides of the SIMM are electrically joined.) A notch at one end is used to ensure that the SIMM is fitted into the socket the correct way around, while round holes at both ends are used to locate the SIMM correctly in the matching socket.

30-pin SIMMs generally contain nine single-bit memory chips, where the ninth chip handles the parity bit. Occasionally they may have two four-bit RAMs and a single-bit chip for the parity. Either way the SIMM stores the actual data in bytes, and only eight of the 30 module connections are devoted to data. The rest are used for addressing, power, read/write control and so on.

30-pin SIMMs have storage capacities ranging from 256KB (256 kilobytes, or 262,144 bytes) up to about 4MB (four megabytes, or 4,194,304 bytes).

Because 30-pin SIMMs store the data in bytes, they must be used in pairs when used in 16-bit computers to store and retrieve the 16-bit wide data. Similarly groups of four are needed to handle the 32-bit wide data used in computers with 486-class CPUs.

30-pin SIMMs are mounted in a special socket with spring clips at the ends, designed to prevent them from working loose due to thermal cycling, but still allow easy removal and replacement. The SIMM is introduced to the socket at an angle of about 60° to the mother board, and then swung to the vertical position where it's captured by the spring clips (see below).

72-pin SIMMs

As time passed, PCs came to need more and more memory capacity. It also became possible to produce DRAM chips with larger internal capacity. To allow more memory capacity to be provided in a single memory module, 72-pin SIMMs were developed.

Like 30-pin SIMMs, the 72-pin type is basically just a small plug-in PC board with memory chips fitted to it, and with contacts along one of the longer sides. In this case however the SIMM is 108mm long and 25.4mm high, with 72 contacts along each lower side, spaced in this case on 0.05" (1.27mm) centres. The contacts on each side are again joined together, so although each side of the SIMM can have its own set of memory chips, they're addressed in common.

Because of the extra connections, a 72-pin SIMM can handle the data in four-byte words, or 32 data bits at a time. As a result this type of SIMM was first used as the memory for 486-class computers, either singly or in pairs when the memory was organised for 'interleaved' reading and writing. (This involves reading and writing to the two modules alternately, to speed things up.)

72-pin SIMMS come in both 'parity' and 'no parity' versions, with the parity type having additional chips to handle the four additional parity bits (one per stored data byte). The parity type are sometimes described as '36-bit SIMMs' because of these additional bits.

The memory capacity of 72-pin SIMMs ranges from about 1MB to 32MB. They mount in a spring-clip socket very similar to that used for the 30-pin SIMMs, and are inserted and removed in the same way.

168-pin DIMMs

The DIMM (Double Inline Memory Module) is a further development from the SIMM, designed to provide even greater memory capacity.

Like SIMMs, DIMMs are basically a low profile plug-in PC board with contacts along the lower long edge. However in this case the contacts on each side of the board are not linked together, but used for different signals (hence *dual* in line). The most commonly used type of DIMM has a total of 168 connections (2 x 84), and is used to provide memory organised in 8-byte words (i.e., 64 bits at a time). This makes them very suitable for computers using Pentium/PII/PIII and similar class processors.

The 168-pin DIMM measures 5.375" x 1" (136.5 x 25.4mm) and has the contacts arranged in three unequal groups. It has notches at both ends, and slides vertically down into a 'slot' socket fitted with swing-away capture levers at the ends. Memory capacities range from 32MB up to 512MB.

Many of the latest models of desktop PC are designed to take memory in 168-pin DIMMs.

100-pin DIMMs

A shorter but taller (3.5" x 1.25") DIMM with only 100 connections (2 x 50) along the lower edge is used to provide plug-in expansion memory in some of the newer laser printers.

SODIMMs

Reduced size or 'small outline' DIMMs — dubbed 'SODIMMs' — have recently been developed to provide very compact memory modules for laptop computers, where space is particularly limited.

Currently there are two different sizes of SODIMMs, one with 72 pins (2 x 36) and the other with 144 pins (2 x 72). They're both about 26mm high, but about 2.375" (60mm) and 2.625" (67mm) long respectively. As with the larger DIMMs the connections along both sides of the board are quite separate.

Memory speed

Generally speaking and regardless of whether you use memory chips or modules, it's important to make sure that its *access time* is equal to — or less than — that of any existing memory already in a computer. Otherwise the memory won't be able to keep up with the computer's CPU.

Today's memory chips and SIMMs generally come with rated access times of either 60 or 70ns (nanoseconds). The slower 70ns type is generally fast enough for 386-based and earlier computers, while the faster 60ns type may be needed for 486 and early Pentium based machines. Later Pentium computers may need even faster 50ns memory.

By the way, the last two digits of the type number printed on the memory chips themselves (including the chips on a SIMM or DIMM) usually represent a code for their access time, especially when they're preceded by a dash. For example chips whose type number ends in '-07' generally have a 70ns access time, while those ending in '-06' are usually 60ns chips. Similarly '-05' represents 50ns chips. Older memory chips ending in '-10' and '-15' will usually have access times of 100ns and 150ns respectively.

Recent very fast Pentium-class computers with a 100MHz or 133MHz system bus need to be fitted with faster SIMM, DIMM or SODIMM modules again. These are usually rated directly in terms of the bus speed they're compatible with — i.e., PC100 or PC133.

Types of memory

As the demand grew for faster and faster memory, memory manufacturers developed improved DRAM chip technology to allow faster storage and retrieval of data.

The first development was a modified chip architecture known as *fast page mode*, which gave a worthwhile speedup over the original type of DRAMs. Then came *extended data out* or 'EDO' memory chips, offering 10-25% faster accessing again — but only in computers where the mother board and support chips were designed for it.

A further development again was *synchronous DRAM* or 'SDRAM' technology, where reading and writing of the memory chips is locked or synchronised to the main CPU clock. SIMMs and DIMMs using SDRAM chips allow memory accessing up to 25% faster than those with EDO, although again this is only true if the motherboard and support chips are designed to be compatible with SDRAM technology.

A further and more recent development again is *double data rate* or 'DDR' SDRAMs, where the chips can read data on both the rising and falling edges of the clock waveform — approximately doubling the data transfer rate over standard SDRAMs. DDR SDRAMs are also called 'SDRAM II' technology.

Still more recent developments in memory technology are 'Rambus' DRAMs, which use a proprietary architecture to achieve speeds up to 10 times faster than conventional DRAMs, and 'synclink' or SLDRAM which is an extension of SDRAM architecture. These are only just starting to be used in high-end systems.

The main thing to remember about all of these enhanced memory chip technologies is that a computer won't be able to take advantage of their capabilities unless its mother board and support chipset have been designed to do so. In other words, you always need to check compatibility before fitting any SIMM or DIMM to your computer. Fitting one of the latest ultra-fast EDO or SDRAM memory modules to an older computer may not make it run faster; in fact it might easily prevent it from booting up and running at all.