Memory Types*

Charles B. Cameron

March 21, 2005

1 Memory Systems

1.1 Elements of a Memory System

The following items are needed in a memory system

- Address lines
- Data lines to carry data written to memory
- Data lines to carry data read from memory
- Control lines to carry commands and status information between the processor and the memory. Examples of these signals are:
  - A Read or Write command. These signals may share the same line ($R/\overline{W}$) or use separate lines ($\overline{R}$ and $\overline{W}$).
  - Enable and disable commands. These signals may be called chip select ($\overline{CS}$), memory select ($\overline{MS}$), chip enable ($\overline{CE}$), or some other comparable term.

The data-write and the data-read buses usually share the very same lines. In some systems, the address bus also shares the same lines. The trade-off is this: the more distinct buses there are, the faster the system runs but the more expensive it is to make.

1.2 Write Operations

To store data in memory the processor and the memory system must perform this sequence of operations:

1. Processor outputs the address of the target location.
2. Processor places data on the bus.
3. Processor issues the write command and any other associated control signals.
4. Memory decodes the address.
5. Memory stores the data on the bus in the addressed location.

1.3 Read Operations

To read data from memory the processor and the memory system must perform this sequence of operations:

1. Processor outputs the address of the target location.
2. Processor issues the read command and any other associated control signals.
3. Memory decodes the address.
4. Memory retrieves the data in the addressed location and places it on the bus.
5. Processor reads data from the bus.

1.4 Physical Arrangments

Within a memory device, addressable locations are typically arranged in rows and columns. Part of the address is used to designate a row and part to designate a column where the addressed location resides physically.

Collections of memory devices also often are arranged in rows and columns to provide even larger arrays of data.

*Course notes for EE461 Microprocessor-based Digital Design

Factor	Name	Symbol	Origin		Derivation	
2^{10}	kibi	Ki	kilobinary:	$(2^{10})^1$	kilo:	$(10^3)^1$
2^{20}	mebi	Mi	megabinary:	$(2^{10})^2$	mega:	$(10^3)^2$
2^{30}	gibi	Gi	gigabinary:	$(2^{10})^3$	giga:	$(10^3)^3$
2^{40}	tebi	Ti	terabinary:	$(2^{10})^4$	tera:	$(10^3)^4$
2^{50}	pebi	Pi	petabinary:	$(2^{10})^5$	peta:	$(10^3)^5$
2^{60}	exbi	Ei	exabinary:	$(2^{10})^6$	exa:	$(10^3)^6$

Table 1: Prefixes for Binary Multiples: adapted from [1]

Individual memory devices typically have sizes which are a power of 2, such as $2^{10} = 1024$ or $2^{20} = 1,048,576$. It is very common in the literature to see such quantities referred to as 1 k or 1 M because they lie close to the actual values of $1\,\mathrm{k} = 1,000$ and $1\,\mathrm{M} = 1,000,000$. Because there is a high risk of confusion in the use of these SI-like abbreviations (does 1 kByte mean 1,000 Bytes or 1,024 Bytes?), in December 1998 the International Electrotechnical Commission (IEC) approved as an IEC International Standard names and symbols for prefixes for binary multiples for use in the fields of data transmission and data processing. These are shown in Table 1. The IEC suggested that the first syllable of the name of the binary-multiple prefix should be pronounced in the same way as the first syllable of the name of the corresponding SI prefix, and that the second syllable should be pronounced as bee. For example, the unit in 1 Kibit = 1 kibibit would be pronounced "kee-bee-bit" and the unit in 5 MiB = 5 mebibytes would be pronounced "meh-bee-bytes". (The suffix *-bi* is evocative of the word *binary*.)

2 Terms

Note: Terms appearing in **boldface** are references to their definitions, distinct entries in the list of terms.

Access Time In the case of a **Read Operation** the access time is the delay from the instant when an address is stable to the time the data is stable.

Address The n-bit value used to specify a particular location in an address space. For example, if $n = 10$ then there are $2^{10} = 1024$ possible addresses. Each may be specified by a unique 10-bit address in the range from 0 to $2^{10} - 1 = 1023$.

Auxiliary Memory Non-volatile memory used to fill RAM when power is first applied to a system is used as auxiliary memory. Examples of such memory include **Magnetic Disk Storage**, tape, CDROM, optical disk, and various other kinds of mass storage units produced commercially over the years. Auxiliary memory is far slower than RAM, but also much cheaper than other forms of faster non-volatile storage such as EEPROM. Also known as **Mass Storage**.

Boot ROM Embedded systems have a start-up program known as a *Boot Program* stored in non-volatile semiconductor memory known as Boot **ROM**. Whenever power is applied to the system, this program is given control. Because it is not economical to store very large programs such as operating systems in such storage, large programs are stored in **Auxiliary Memory** and retrieved from there by the Boot Program. Once the operating system has been retieved and placed in **Main Memory**, it is given control and the Boot Program has completed its only task.

Cache Memory Very fast memory designed to hold a subset of the contents of **Main Memory**. The goal is to minimize the number of times when the processor must look for data in the Main Memory because it is missing from the cache, as cache memory is generally considerably faster than Main Memory. Only registers on the same

chip as the microprocessor itself are faster than cache memory.

CDROM Compact Disk **ROM**. A compact disk formatted to contain arbitrary digital data. Just as with compact disks formatted for audio use, data are laid down in spiral tracks, not in concentric rings as with **Magnetic Disk Storage**. The data begin near the center of the disk and are recorded sequentially outward toward the perimeter. (This is opposite to the direction of recording in phonograph records, where the sounds begin near the perimeter and proceed toward the center.)

Core Dump The contents of memory, usually printed in hexadecimal on lengthy listings. Sometimes the only way to debug an errant program is to examine a core dump taken at the time of the failure. Such debugging is very tedious and highly error prone. People who can do it are often regarded as minor gods.

Core Memory A non-volatile form of magnetic memory. Each bit of data was stored magnetically in a small, doughnut-shaped ring called a core. The cores were arranged in three-dimensional arrays, suspended by the wires which were threaded through the center of each core and used to transfer data to and from them. The memory was bulky, by present-day standards, and fragile. So dominant was this memory technology in the early days of computers that the term Core Memory is often used synonymously with **Main Memory**, even today. It gave rise to the term **Core Dump**.

Density Describes the capacity of a memory. Expressed in units of quantity of data per unit area, unit volume, memory slot, *etc.*, as in 20 Gbits/inch^2 or 20 Mbits/slot.

Dynamic RAM (DRAM) RAM whose contents do vanish, even with continous power available, unless their contents are refreshed frequently. Unappealing as this trait is, the fact that DRAM can achieve much higher density than SRAM makes it attractive in large-memory systems. However, it generally is also slower than SRAM. The opposite of DRAM is **Static RAM (SRAM)**.

EEPROM Electrically Erasable **PROM**. The user is responsible for storing the data. However, individual address locations in the device can be erased electrically and rewritten. These memories are very attractive for prototyping and for storing information (such as configuration choices) which it is desirable not to lose when power is removed. Unfortunately, the number of erasures is finite and not very large (on the order of 100,000 at the time of writing, adequate for many purposes.) As with all ROMs whose contents can be modified at all, writing the data is substantially slower than reading it is. Often, though, an EEPROM can be reprogrammed in the same circuit it is used in. Programming an EEPROM requires application of a high voltage to one input of the device to put it into write mode.

EPROM Erasable **PROM**. The user is responsible for storing the data. However, the device can be re-initialized by exposure to ultraviolet (UV) light for around 20 minutes. Because such radiation is harmful, UV erasers are enclosed in metal and generally provide a small tray on which the device to be erased may be placed. These memories are useful for prototyping but the lengthy erase time is inconvenient. Unfortunately, the number of erasures is finite and not very large (on the order of 100,000 at the time of writing, adequate for many purposes.) Programming an EPROM requires application of a high voltage to one input of the device to put it into write mode.

Fetch See **Read Operation**.

Flash Memory Similar to **EEPROM** except that individual addresses cannot be erased and rewritten. Rather, entire blocks (also called sectors) of the memory are erased at one time. In fact, some flash memories can only be rewritten in their entirety. The fact that individual addresses cannot be selectively altered makes flash

memories less useful for storing varying data during program execution than for program memories. As with all ROMs whose contents can be modified at all, writing the data is substantially slower than reading it is. Often, though, a flash memory can be reprogrammed in the same circuit it is used in.

Magnetic Disk Storage Memory which uses non-volatile magnetic coatings on flat disks for storing data. The data are typically organized in concentric rings (called tracks) on the surfaces of the platters. A location can therefore be specified by mentioning its track number, its surface number, and its circumferential position. Magnetic disk storage is therefore partly random and partly sequential because although read and write heads can move from one track to another, the system must wait for data to rotate under the heads and it may take the head an appreciable time to go from, say, the innermost track to the outermost track. The first of these two types of delay is referred to as *latency* and the second is referred to as *seek time*.

Main Memory RAM. The fastest memory is the registers on the processor chip. The next fastest is **Cache Memory**. RAM is the next fastest memory in general use and it comprises the main memory (also known as **Working Memory** of a computer system. **SRAM** and **DRAM** are the dominant forms of main memory today. Formerly it was comprised of **Core Memory**.

Mask ROM The data are stored at the factory at the time of manufacture. The term *mask* comes from the use of photographic masks used to create the semiconductors used in this type of memory. The cost to create the mask is significant but if large numbers of units are required, the per-unit costs are very low.

Mass Storage See **Auxiliary Memory**.

Non-Volatile Memory Memory which loses its contents upon the removal of power.

Non-Volatile RAM (NVRAM) Usually this is implemented with a standard static RAM and a battery back-up.

One-Time Programmable ROM See **PROM**.

Optical Disks This term actually applies to a **Magnetic Disk Storage** with an additional optical element. The data are stored magnetically as usual. However, the magnetic fields used for writing the data are ordinarily insufficient to change the magnetic polarity on the disk. By shining a high-power laser onto the surface and heating the magnetic medium above the Curie point, the coercivity of the medium is lowered locally (in the immediate vicininty of the very small laser beam), permitting the data to be stored there. As the cross-sectional area of the laser beam is so very small, the data density of an optical disk is substantially larger than that of a standard magnetic disk.

Programmable Logic Devices (PLD) These come in various forms. Their principal characteristic is that not all possible combinations of inputs are decoded. A **ROM**, in contrast, fully decodes every address. Consequently, PLDs do not really qualify as memory devices, although the technology is very similar.

PROM Programmable Read-Only Memory. The user is responsible for putting the contents into the memory. This can be done only once and usually entails burning fuses selectively in the device. Programming a PROM requires application of a high voltage to one input of the device to put it into write mode.

Random Access Memory (RAM) Refers to memory which can be written to or modified and which can be accessed randomly. Random access implies that **Access time** is independent of the **address**. The term RAM is something of a misnomer inasmuch as **ROM** and other types of memory also can be accessed randomly.

Read-Only Memory (ROM) Non-volatile memory which can be read only. There are several

variants of ROM, all of which differ in the way in which data is stored in them in the first place and in the way that data may be altered. If it can be altered at all, writing data always takes considerably longer than reading it does. Common uses for ROM include tables of unchanging data (such as character-generator tables or tables of sines), **Boot ROMs**, microprograms for running a processor, and permanent configuration information.

Read Operation Refers to a transfer of data from a device to the master. Also known as a **Fetch**. A transfer is termed a **Read** or a **Write** according to the viewpoint of the master, not the slave.

Read/Write Memory (RWM) A memory which can be both read and written. Although this term may be more accurate than the term **RAM**, the latter term is nonetheless more commonly used to refer to such memory.

Sequential Access Memory (SAM) Memory whose access time is directly proportional to the difference between the address location sought and that most recently accessed. Magnetic tape is an example of sequential memory. Another example is video memory, which is always read sequentially when its contents are being transferred to a cathode ray tube, even though the data may be stored randomly.

Static RAM (SRAM) RAM whose contents do not vanish, provided power is supplied. The opposite of SRAM is **Dyamic RAM (DRAM)**.

Store See **Write Operation**

Volatile Memory Memory which does not lose its contents upon the removal of power.

Working Memory See **Main Memory**

Write Operation Refers to a transfer of data from the master to a device. Also known as a **Store**. A transfer is termed a **Read** or a **Write** according to the viewpoint of the master, not the slave.

References

[1] International Electrotechnical Commission, *IEC 60027-2 Letter Symbols to Be Used in Electrical Technology — Part 2: Telecommunications and Electronics*. IEC, November 2000.

[2] V. P. Heuring and H. F. Jordan with a contribution by Miles Murdocca, *Computer Systems Design and Architecture*, 2nd ed. Upper Saddle River, New Jersey: Pearson Education, Inc., 2004.

[3] R. J. Tocci and N. S. Widmer, *Digital Systems: Principles and Applications*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 2001.