Memory Devices

This lecture describes common semiconductor memory devices. You should learn to: select the appropriate type of memory device for different applications, combine ICs to form memory arrays, and design address decoders using SSI decoders, PALs and PROMs.

Memory Technologies

A wide variety of IC memory devices are available. They vary in terms of data permanence, the way data is stored in the device, power consumption, cost, capacity, and access time.

SRAM

Static RAM. Volatile read/write memory. Data stored as the state of a flip-flop. Contents lost when power is removed. CMOS devices have very low power consumption when not being accessed and can be used with battery or capacitor backup. Bipolar devices have higher power consumption but feature the shortest access times.

DRAM

Dynamic RAM. Data stored as charge on a capacitor. Lowest cost per bit for RAM. Contents lost if memory locations not accessed (refreshed) every few milliseconds. Capacity about 4 times that of same-generation SRAM. 16 Mbit 60ns devices currently in volume production. Will be discussed in detail in next lecture.

Mask-Programmed ROM

Non-volatile read-only memory. Data stored as connections between gates. Memory contents determined at time of manufacture. Lowest cost per bit of any memory but have large NRE costs so only suitable for large volume (>10k) applications.

EPROM

Field-programmable non-volatile ROM that can be erased by exposure to UV light. Data stored as charge on “floating gates.” Typically have byte-wide organization. Not as fast or dense as RAM. OTP devices are less expensive since the packages don’t have windows (and can’t be erased).

EEPROM

Non-volatile memory that can be written like RAMs. Write cycles are relatively long (hundreds of microseconds). Relatively small capacity. Limited to several thousand write cycles. Mostly used to store infrequently-changed configuration information.

Flash EEPROM

Similar to EEPROMs but have with larger capacity. Whole sections of the chip must be erased and then re-written.

Specialized Memories

Video RAM

RAM used in video display circuits. They have a second independent data output used that sequentially reads out a complete row of data from the memory array. This “dual ported” arrangement allows both the CPU and the video signal generator to access the RAM at the same time. This reduces contention and improves performance.

Serial EEPROMs

EEPROMs whose contents must be read or written one bit at a time from start to end. The serial interface reduces chip size, pin count and cost (e.g. a 128x8 EEPROM in an 8-pin DIP for $1). Typical application is storing configuration information. The slow serial access is not a drawback since the device is typically used only when the product is turned on (or off).

Exercise:

What type of memory device(s) would likely be used in a popular PC for the following purposes: storage for the power-on boot code? the “CMOS” configuration memory? the main data/program RAM? the main memory cache? the boot program in a prototype of this PC? the video display memory?

What type of memory device(s) would likely be used for the following applications: non-volatile memory for user programs in a calculator? a font card for a laser printer? user-upgradeable storage for the firmware in modem? the mileage reading in a digital car odometer? a video game cartridge?

Memory Organization

A memory IC that stores B bits in 2^N memory locations is described as a 2^N by B memory. For example a 16384 by 4 (16kx4) RAM would have 14 address inputs and 4 data read/write pins. If the width (number of bits available in parallel) of the device is less than the CPU bus width, devices are combined in parallel by connecting their data bits to successive bits of the data bus. To make available more memory than can be provided by one such set of ICs, B-bit wide blocks of devices are selected by a circuit called an address decoder.

Exercise: Eight (8) 1Mx4 devices are to be connected to a CPU with a 16-bit data bus. How many address and data bits does each IC require? What is the total memory size in MBytes? Draw a block diagram showing the address and data bus connections to the different ICs. What is the load on each CPU bus pin? How many enables will be required from the address decoder?

The storage elements inside the IC are arranged as a square matrix of 2^{N-1} rows and 2^{N-1} columns of memory elements. Half of the address bits select one row and the other half drive a multiplexer that select one of the columns.

Address Decoding

The memory space of a computer system is an array of 2^N bytes where N is the number of CPU address bits. Within the address space the system must fit the memory and i/o devices. Combinational circuits called *address decoders* select which devices respond to which addresses. The decoders' inputs are address bits and its outputs are logic signals that enabled devices when the address falls within the desired address range. The three most common ways to implement address decoders are SSI decoders, PALs and PROMs.

Address decoders are invariably decode regions that start at and have lengths of powers-of-two. The decoded region can therefore be written as a bit pattern that the decoder responds to. For example, a decoder for addresses from ($ indicates a hexadecimal value) $02 0000 to $02 7FFF (32 kBytes) would respond to addresses of the form 0000 0010 0XXX XXXX XXXX XXXX where the X's are "don't cares."

Designs can be simplified by allowing the decoder to respond to multiple addresses. This is called partial decoding. For example, if the above decoder responded to addresses of the form 0000 001X XXXX XXXX XXXX XXXX the decoded region would extend from $00 0000 to $01 FFFF (128kB) and the 32kB region would appear replicated in all four 32k blocks in that region. This partial decoding "wastes" part of the processor's address space because it is now unavailable for other devices. The advantage is that the extra "don't-care" bits leads to simpler implementation.

SSI Decoders

A decoder such as the 74LS138 3-to-8 decoder can be used to divide up an address range.

Exercise: How would '138 be used to divide up the 68000's 24-bit (16 MByte) address space into 8 2-MByte regions?

PAL Decoders

Since PALs can generate complex combinational functions they can be used to divide up a memory space into regions of different sizes.

Exercise: We want to design a PAL decoder that selects one 512kB region from $000000 to $03FFFF and one 4 MB region from $F0000 to $FFFFFF out of the 68000's address space. How may input bits are required? How many outputs? Write out the logic equations.

PROM Decoders

PROMs can also be used to implement arbitrary combinational functions. The PROM's address inputs are connected to CPU address bus lines and the PROM outputs are the enables for the desired memory areas. Decoder PROMs must have very short access times so the main memory access time is not affected.

Exercise: What size (NxM) PROM would be needed to split up a 64k address space into 8 memory blocks the smallest of which is 1kBytes?