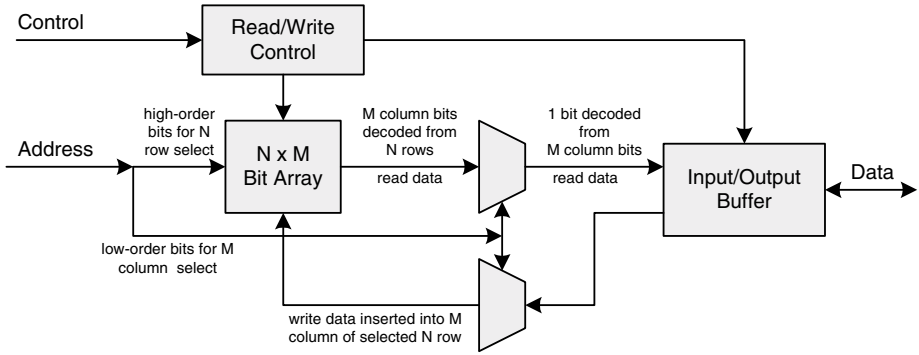

FIGURE 4.1 General memory device.

Despite the logical organization of the device, the internal bit array is usually less rectangular and more square in its aspect ratio. For example, a 131,072 × 8 memory (128 kB) may be implemented as 512 × 256 × 8. This aspect ratio minimizes the complexity of the address-decode logic and also has certain manufacturing process benefits. It takes more logic to generate 131,072 enable signals in one pass than to generate 512 and then 256 enables in two passes. The first decode is performed up-front in the memory array, and the second decode is performed by a multiplexer to pass the desired memory location.

Nonvolatile memory can be separated into two subcategories: devices whose contents are programmed at a factory without the expectation of the data changing over time, and devices whose contents are loaded during system manufacture with anticipation of in-circuit updates during the life of the product. The former devices are, for all practical purposes, write-once devices that cannot be erased easily, if at all. The latter devices are designed primarily to be nonvolatile, but special circuitry is designed into the devices to enable erasure and rewriting of the memory contents while the devices are functioning in a system. Most often, these circuits and their associated algorithms cause the erase/write cycle to be more lengthy and complex than simply reading the existing data out of the devices. This penalty on write performance reflects both the desire to secure the nonvolatile memory from accidental modification as well as the inherent difficulty in modifying a memory that is designed to retain its contents in the absence of power.

Volatile memory can also be separated into two subcategories: devices whose contents are nonvolatile for as long as power is applied (these devices are referred to as *static*) and devices whose contents require periodic refreshing to avoid loss of data even while power is present (these devices are referred to as *dynamic*). On first thought, the category of dynamic devices may seem absurd. What possible benefit is there to a memory chip that cannot retain its memory without assistance? The benefit is significantly higher density of memory per unit silicon area, and hence lower cost of dynamic versus static memory. One downside to dynamic memory is somewhat increased system complexity to manage its periodic update requirement. An engineer must weight the benefits and complexities of each memory type when designing a system. Some systems benefit from one memory type over the other, and some use both types in different proportions according to the needs of specific applications.

Memory chips are among the more complex integrated circuits that are standardized across multiple manufacturers through cooperation with an industry association called the Joint Electron Device Engineering Council (JEDEC). Standardization of memory chip pin assignments and functionality is important, because most memory chips are commodities that derive a large portion of their value by

being interoperable across different vendors. Newer memory technologies introduced in the 1990s resulted in more proprietary memory architectures that did not retain the high degree of compatibility present in other mainstream memory components. However, memory devices still largely conform to JEDEC standards, making their use that much easier.

4.2 EPROM

Erasable-programmable read-only-memory, EPROM, is a basic type of nonvolatile memory that has been around since the late 1960s. During the 1970s and into the 1990s, EPROM accounted for the majority of nonvolatile memory chips manufactured. EPROM maintained its dominance for decades and still has a healthy market share because of its simplicity and low cost: a typical device is programmed once on an assembly line, after which it functions as a ROM for the rest of its life. An EPROM can be erased only by exposing its die to ultraviolet light for an extended period of time (typically, 30 minutes). Therefore, once an EPROM is assembled into a computer system, its contents are, for all practical purposes, fixed forever. Older ROM technologies included programmable-ROMs, or PROMs, that were fabricated with tiny fuses on the silicon die. These fuses could be burned only once, which prevented a manufacturer from testing each fuse before shipment. In contrast, EPROMs are fairly inexpensive to manufacture, and their erasure capability allows them to be completely tested by the semiconductor manufacturer before shipment to the customer. Only a full-custom mask-programmed chip, a true ROM, is cheaper to manufacture than an EPROM on a bit-for-bit basis. However, mask ROMs are rare, because they require a fixed data image that cannot be changed without modifying the chip design. Given that software changes are fairly common, mask ROMs are relatively uncommon.

An EPROM's silicon bit structure consists of a special MOSFET structure whose gate traps a charge that is applied to it during programming. Programming is performed with a higher than normal voltage, usually 12 V (older generation EPROMs required 21 V), that places a charge on the floating gate of a MOSFET as shown in Fig. 4.2.

When the programming voltage is applied to the control gate, a charge is induced on the floating gate, which is electrically isolated from both the silicon substrate as well as the control gate. This isolation enables the floating gate to function as a capacitor with almost zero current leakage across the dielectric. In other words, once a charge is applied to the floating gate, the charge remains almost indefinitely. A charged floating gate causes the silicon that separates the MOSFET's source and drain contacts to electrically conduct, creating a connection from logic ground to the bit output. This means that a programmed EPROM bit reads back as a 0. An unprogrammed bit reads back as a 1, because the lack of charge on the floating gate does not allow an electrical connection between the source and drain.

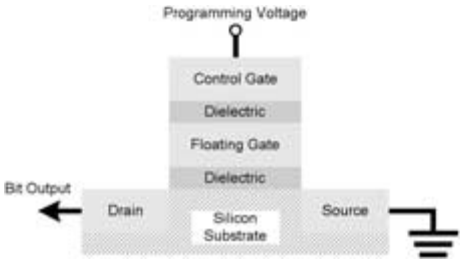

FIGURE 4.2 EPROM silicon bit structure.