1.5 Technologies for building processors and memory

Processors and memory have improved at an incredible rate, because computer designers have long embraced the latest in electronic technology to try to win the race to design a better computer. The table below shows the technologies that have been used over time, with an estimate of the relative performance per unit cost for each technology. Since this technology shapes what computers will be able to do and how quickly they will evolve, we believe all computer professionals should be familiar with the basics of integrated circuits.

Table 1.5.1: Relative performance per unit cost of technologies used in computers over time (COD Figure 1.10).

Year	Technology used in computers	Relative performance / unit cost
1951	Vacuum tube	1
1965	Transistor	35
1975	Integrated circuit	900
1995	Very large-scale integrated circuit	2,400,000
2013	Ultra large-scale integrated circuit	250,000,000,000

Source: Computer Museum, Boston, with 2013 extrapolated by the authors.

A *transistor* is simply an on/off switch controlled by electricity. The integrated circuit (IC) combined dozens to hundreds of transistors into a single chip. When Gordon Moore predicted the continuous doubling of resources, he was forecasting the growth rate of the number of transistors per chip. To describe the tremendous increase in the number of transistors from hundreds to millions, the adjective very large scale is added to the term, creating the abbreviation VLSI, for *very large-scale integrated circuit*.

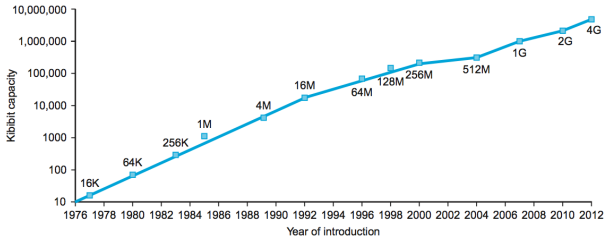
Transistor: An on/off switch controlled by an electric signal.

Very large-scale integrated (***VLSI***) circuit: A device containing hundreds of thousands to millions of transistors.

This rate of increasing integration has been remarkably stable. The figure below shows the growth in DRAM capacity since 1977. For 35 years, the industry has consistently quadrupled capacity every 3 years, resulting in an increase in excess of 16,000 times!

Figure 1.5.1: Growth of capacity per DRAM chip over time (COD Figure 1.11).

The y-axis is measured in kibibits (2^{10} bits). The DRAM industry quadrupled capacity almost every three years, a 60% increase per year, for 20 years. In recent years, the rate has slowed down and is somewhat closer to doubling every two years to three years.



PARTICIPATION ACTIVITY 1.5.1: Technologies in processors and memory.

1) Very large-scale integrated circuits combine dozens to hundreds of vacuum tubes into a single chip.
   - ○ True
   - ○ False

2) From the 1970s to 1990s, DRAM capacity has quadrupled every three years.
   - ○ True
   - ○ False

3) DRAM capacity today continues to quadruple every three years.
   - ○ True
   - ○ False

To understand how to manufacture integrated circuits, we start at the beginning. The manufacture of a chip begins with *silicon*, a substance found in sand. Because silicon does not conduct electricity well, it is called a *semiconductor*. With a special chemical process, it is possible to add materials to silicon that allow tiny areas to transform into one of three devices:

- Excellent conductors of electricity (using either microscopic copper or aluminum wire)
- Excellent insulators from electricity (like plastic sheathing or glass)

- Areas that can conduct or insulate under special conditions (as a switch)

Silicon: A natural element that is a semiconductor.

Semiconductor: A substance that does not conduct electricity well.

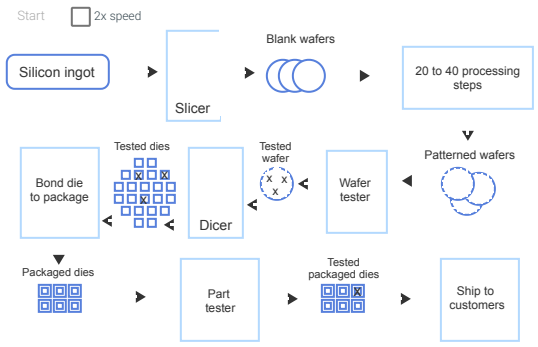
Transistors fall into the last category. A VLSI circuit, then, is just billions of combinations of conductors, insulators, and switches manufactured in a single small package.

The manufacturing process for integrated circuits is critical to the cost of the chips and hence important to computer designers. The figure below shows that process. The process starts with a *silicon crystal ingot*, which looks like a giant sausage. Today, ingots are 8-12 inches in diameter and about 12-24 inches long. An ingot is finely sliced into *wafers* no more than 0.1 inches thick. These wafers then go through a series of processing steps, during which patterns of chemicals are placed on each wafer, creating the transistors, conductors, and insulators discussed earlier. Today's integrated circuits contain only one layer of transistors but may have from two to eight levels of metal conductor, separated by layers of insulators.

Silicon crystal ingot: A rod composed of a silicon crystal that is between 8 and 12 inches in diameter and about 12 to 24 inches long.

Wafer: A slice from a silicon ingot no more than 0.1 inches thick, used to create chips.

PARTICIPATION ACTIVITY 1.5.2: The chip manufacturing process (COD Figure 1.12).



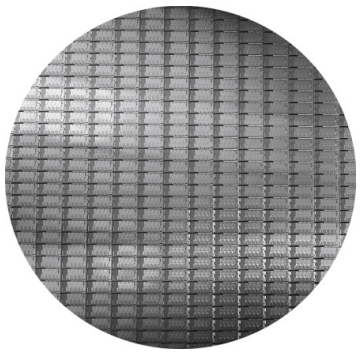
A single microscopic flaw in the wafer itself or in one of the dozens of patterning steps can result in that area of the wafer failing. These *defects*, as they are called, make it virtually impossible to manufacture a perfect wafer. The simplest way to cope with imperfection is to place many independent components on a single wafer. The patterned wafer is then chopped up, or *diced*, into these components, called *dies* and more informally known as *chips*. The figure below shows a photograph of a wafer containing microprocessors before they have been diced; an earlier figure (COD Figure 1.9) shows an individual microprocessor die.

Defect: A microscopic flaw in a wafer or in patterning steps that can result in the failure of the die containing that defect.

Die: The individual rectangular sections that are cut from a wafer, more informally known as ***chips***.

Figure 1.5.2: A 12-inch (300mm) wafer of Intel Core i7 (Courtesy Intel) (COD Figure 1.13).

The number of dies on this 300 mm (12 inch) wafer at 100% yield is 280, each 20.7 by 10.5 mm. The several dozen partially rounded chips at the boundaries of the wafer are useless; they are included because it's easier to create the masks used to pattern the silicon. This die uses a 32-nanometer technology, which means that the smallest features are approximately 32 nm in size, although they are typically somewhat smaller than the actual feature size, which refers to the size of the transistors as "drawn" versus the final manufactured size.



PARTICIPATION ACTIVITY 1.5.3: Chip manufacturing process.

1) Silicon is _____.
   - a rare element
   - a common element found in sand

2) The manufacturing process begins with a silicon _____.
   - ingot
   - wafer
   - die

3) Blank wafers undergo 20 to 40 chemical processing steps to create _____.
   - transistors, conductors, and insulators
   - silicon
   - dies

4) A wafer tester evaluates the patterned wafer for the presence of _____.
   - patterns
   - defects
   - components

5) After the patterned wafer is chopped up into dies, the dies are _____.
   - packaged
   - shipped
   - diced

Dicing enables you to discard only those dies that were unlucky enough to contain the flaws, rather than the whole wafer. This concept is quantified by the *yield* of a process, which is defined as the percentage of good dies from the total number of dies on the wafer.

Yield: The percentage of good dies from the total number of dies on the wafer.

The cost of an integrated circuit rises quickly as the die size increases, due both to the lower yield and to the fewer dies that fit on a wafer. To reduce the cost, using the next generation process shrinks a large die as it uses smaller sizes for both transistors and wires. This improves the yield and the die count per wafer. A 32-nanometer (nm) process was typical in 2012, which means essentially that the smallest feature size on the die is 32 nm.

Once you've found good dies, they are connected to the input/output pins of a package, using a process called bonding. These packaged parts are tested a final time, since mistakes can occur in packaging, and then they are shipped to customers.

Elaboration

The cost of an integrated circuit can be expressed in three simple equations:

$$\text{Cost per die} = \frac{\text{Cost per wafer}}{\text{Dies per wafer} \times \text{yield}}$$

$$\text{Dies per wafer} \approx \frac{\text{Wafer area}}{\text{Die area}}$$

$$\text{Yield} = \frac{1}{(1 + (\text{Defects per area} \times \text{Die area}/2))^2}$$

The first equation is straightforward to derive. The second is an approximation, since it does not subtract the area near the border of the round wafer that cannot accommodate the rectangular dies (see the figure above). The final equation is based on empirical observations of yields at integrated circuit factories, with the exponent related to the number of critical processing steps.

Hence, depending on the defect rate and the size of the die and wafer, costs are generally not linear in the die area.

PARTICIPATION ACTIVITY 1.5.4: Check yourself: Yield and chips costs.

A key factor in determining the cost of an integrated circuit is volume. Which of the following are reasons why a chip made in high volume should cost less?

1) With high volumes, the manufacturing process can be tuned to a particular design, increasing the yield.
   - True
   - False

2) It is less work to design a high-volume part than a low-volume part.
   - True
   - False

3) The masks used to make the chip are expensive, so the cost per chip is lower

for higher volumes.

- ○ True
- ○ False

4) Engineering development costs are high and largely independent of volume; thus, the development cost per die is lower with high-volume parts.

- ○ True
- ○ False

5) High-volume parts usually have smaller die sizes than low-volume parts and therefore have higher yield per wafer.

- ○ True
- ○ False

Provide feedback on this section