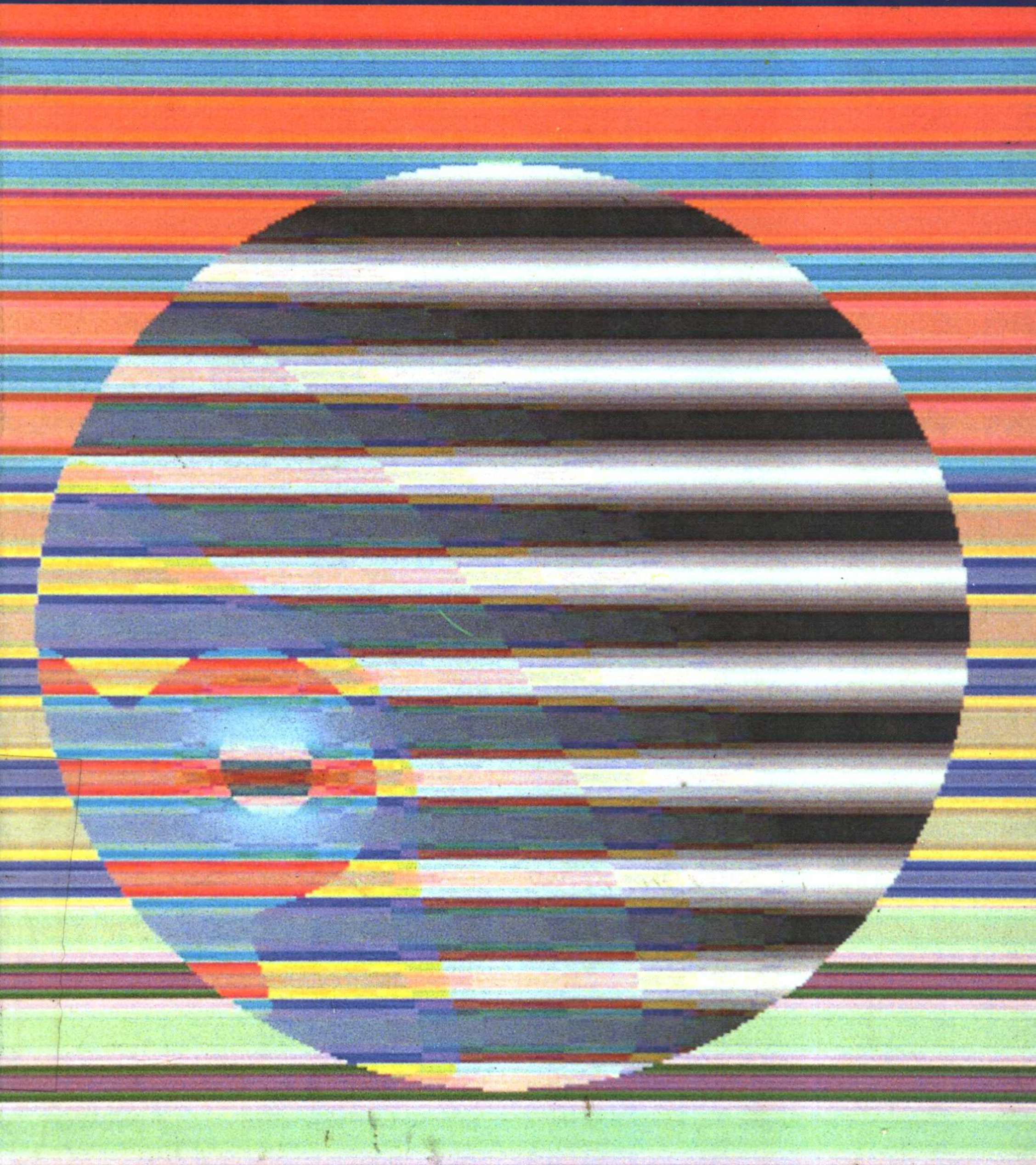
MICROELECTRONICS SYSTEMS DESIGN

The Physics of VLSI Systems

Robert W Keyes



THE PHYSICS OF VLSI SYSTEMS

ROBERT W. KEYES

IBM Corporation
Thomas J. Watson Research Laboratory
Yorktown Heights, NY

ADDISON-WESLEY PUBLISHING COMPANY
Wokingham, England · Reading, Massachusetts · Menlo Park, California
New York · Don Mills, Ontario · Amsterdam · Bonn · Sydney
Singapore · Tokyo · Madrid · Bogota · Santiago · San Juan

Cover graphic by kind permission of Digital Pictures Ltd, London.
Printed in Great Britain by The Bath Press, Avon.
First printed 1987

British Library Cataloguing in Publication Data

Keyes, Robert W.
The physics of VLSI systems. –
Microelectronics systems design series).
1. Electronic digital computers – Circuits
2. Integrated circuits – Very large scale integration
I. Title II. Series
621.395 TK7888.4

ISBN 0-201-15402-1

Library of Congress Cataloging in Publication Data

Keyes, Robert W., 1921-
The physics of VLSI systems.

(Microelectronics systems design series)
Bibliography: p.
Includes index.
1. Integrated circuits – Very large scale integration.
2. Semiconductors. 3. Digital electronics. I. Title.
II. Series.
TK7874.K44 1987 621.395 87–19289

ISBN 0-201-15402-1

Preface

The pace of improvement of semiconductor products is unprecedented in the history of technology. Changes are measured in factors of two or four or in orders of magnitude rather than in percentages. These advances in the value of semiconductor devices have enabled great forward strides in the information-processing power and the availability of large systems, systems containing hundreds of thousands to many millions of electronic components. A large body of auxiliary technology is needed to combine the capabilities of a great many semiconductor devices into a system that enables their power to be brought to work towards a single goal. The link is conditioned by the laws of physics and the principles involved may be understood by applying the methods of physics. Thus, an overview of the entire computing system from the physicist's perspective is the theme of this book.

The purpose of the book is to enhance interdisciplinary communication. The computer scientist's awareness of the limitations and the potential of the machinery behind the console should be increased. The device engineer will be introduced to those aspects of a large system that make possible the use of his efforts as part of a larger whole. The material does not form the basis for an academic course, but is intended to supplement courses in computer-related subjects, such as those in semiconductor devices or computer architecture. Knowledge equivalent to first-year university physics is assumed in the reader.

The development of semiconductor technology and of computer technology, which should be related by a physical view of systems, is reviewed in Chapter 1. Chapter 1 also serves to introduce some of the terminology that is needed later. Chapter 2 provides a brief introduction to the properties of semiconductors, the uniqueness of which makes solid state electronics possible. Qualitative description of a few essential elementary quantum mechanical concepts is necessary here. The chapter goes on to describe the physical basis of semiconductor devices and their use in digital circuits. The experienced semiconductor practitioner may wish to skip this chapter.

Chapters 3 and 4 establish in a rather general way the representation of information as a physical quantity and what must be required of devices that manipulate it in a large computing system. Then the outstanding example of Very Large Scale Integration, semiconductor random access memory, is examined in some detail in Chapter 5. It makes available rapid access to very large amounts of information at low cost, a function vital to the success of the modern computer. Other technologies, aimed at providing a similar function at even lower cost, have also profited from the miniaturization and integration that have characterized semiconductor circuits, and from the tools and methods that enabled the steady advance of VLSI. The bubble memory is dependent on quite different physical principles, but is technologically very similar to semiconductor VLSI. The rate of miniaturization of storage in disk files has kept pace with that of semiconductor memory. Disk storage has become an essential part of computing systems, and is also treated.

The application of VLSI to large-scale computing is constrained and limited by the ability of technology to fabricate structures, the physics of semiconductors and semiconductor devices, and the nature and requirements of systems. These constraints are considered in Chapters 6 to 8. The need for a large amount of communication among components that are to work together as a unified system has an important bearing here.

The ultimate worth of a computer lies in its ability to process information. This can be measured by the rate at which rather imprecisely definable instructions can be executed. A physical view of a system that permits such a rate to be estimated is presented in Chapter 9. The physical model relates performance to the technological parameters of the system components. Past progress in the improvement of parameters can be extrapolated into the future to project performance and to examine alternative scenarios.

The success of electronic information processing has rested on the transistor. Novel effects that work well in laboratory experiments and can be made to reproduce the function of the transistor in some sense are often proposed as bases for future computing systems. Thus, in conclusion, the ideas behind a set of these proposals for alternative devices that have attracted substantial attention and support are examined from the perspective of the earlier chapters.

The author's education in the subject matter of this book was obtained as a staff member of the IBM Research Division and he wishes to acknowledge the patience and assistance of many colleagues there and the opportunities and stimulation available in that environment. Leonard D. Harris typed most of the manuscript.

Yorktown Heights, New York Robert W. Keyes

Contents

1

Microelectronics and integration

The road to VLSI began with the invention of the transistor in 1947. The importance of the invention was recognized immediately and it attracted wide interest. The concept was quickly translated into manufacturable devices and incorporated in many commercial products. A stream of new circuits and improved structures and production methods emerged. The rapid pace of innovation and change in technology that began around 1950 has never abated. Even the most optimistic predictions of the future of semiconductor electronics turn out to be overly conservative.

Interest in automatic computation was heightened during World War II. Computational machinery of the 1940s was based on relays. Wartime developments proved that vacuum tubes could be the elements of larger, faster machines. Commercial general purpose computers built from vacuum tubes became widely available in industrial and business enterprises in the 1950s. The power and versatility of these machines opened a market for even more powerful processors. Transistors emerged to replace the tube and appeared in commercial computers in the late 1950s.

Changes in technology since the introduction of the transistor have occurred in smaller steps; the advances may be described as evolutionary. However, taken together the changes are so vast and significant and have occurred in such a short space of time that they are not inappropriately called revolutionary.

The high reliability and low power dissipation of transistors made it possible to assemble them into very large systems. All aspects of such systems, from a single transistor to a completed system, are governed by the laws of physics. The physics of the solid state, in particular the electronic properties of semiconductors on which the transistor is founded, is reasonably well understood. Many other subdisciplines, e.g., optics, defect crystallography, surface physics, and the properties of insulators are involved in the fabrication and assembly of systems. The physics of the system considered as a collection of macroscopic units, such as circuits, chips, boards, wires, connectors, and cooling fluids, is less well developed. Nevertheless, the application of VLSI in large systems is susceptible to investigation by the methodology of physics.

1.1 Integration

The ever-falling cost has been made possible by integration, the manufacture of many components on a single piece of semiconductor material. Each step in the manufacturing process is performed on a great many components simultaneously, and its cost is, consequently, shared by an enormous number of them. The elementary unit of circuitry is the chip, a thin block of silicon one to one hundred square millimeters in area on which electrical components such as transistors and capacitors and the wires that interconnect them have been fabricated. The chip is the entity that is sold and handled and connected to other chips and to power supplies to form systems.

Thus the progress of integrated circuit technology is appropriately measured by the content of the chip. We illustrate the advance of the art in Fig.1.1, a plot of the number of transistors on the most advanced commercial chips as a function of time. Perhaps the most remarkable aspect of the progress illustrated is that it has not been accompanied by increases in the cost of a chip. A chip containing, say, 100 000 transistors may be sold by the manufacturer for $2, a few millicents per transistor, much less than, for example, a sheet of disposable paper tissue. The decrease in cost per component is, indeed, the economic force driving the advances in technology that are responsible for the events summarized by Fig.1.1. The continually decreasing cost opens new applications and markets to electronics. The trend shown in Fig. 1.1 shows no sign of slowing. In fact, a chip content one thousand times larger than the last point in the figure has been projected. Progress far into the future can be anticipated with some confidence.

The state of integrated circuit technology is frequently described by certain more or less ill-defined acronyms. The early integrated circuits consisted of a few transistors interconnected on a chip, sometimes with a few resistors included. The abbreviation IC was introduced to characterize the new technology. The realization that the methods could be extended to hundreds of components caused the term Large Scale Integration (LSI) to be coined. When efforts to increase the amount of circuitry on a chip succeeded in developing chips that could store 2^{10} (= 1024 or 1 K) bits of information, and there was reason to believe that progress in the same direction could continue for some time to come, practitioners of the techniques involved began to refer to the chips of the future with the added adjective Very and to talk of VLSI, Very Large Scale Integration. In the hope of adding an element of uniqueness to their endeavours, some now speak of Ultra Large Scale Integration, ULSI. Very important changes in technology are involved in progressing from ten transistor chips to thousand transistor chips and to the hundred million transistor chips forecast for the future, so perhaps the distinctions are not entirely artificial. New products are needed to utilize the new capabilities effectively, new kinds of manufacturing tools must be used, and the design of such large and complex entities requires constant innovation in techniques.

The basic technological trends underlying the increase of the number of components on a chip may be loosely divided into three categories. One is

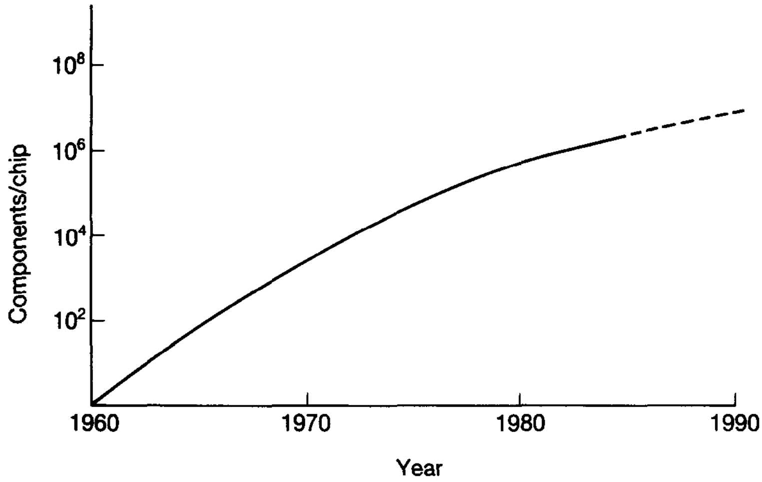


Figure 1.1 The number of transistors on advanced chips as a function of the year in which they were first produced.

miniaturization, reducing the dimensions of everything fabricated on a chip. The term microelectronics refers to the small size of components. A second trend is the increasing size of the chip. The development of these two parameters since the introduction of the integrated circuit is shown in Fig. 1.2. The state of miniaturization for the purposes of the figure is characterized by the minimum dimension of structures fabricated on a chip. The third ingredient of increasing integration may be called compaction, and consists of making better use of the available area through improved designs and circuits. We illustrate the role of this factor by dividing the device count presented in Fig. 1.1 by the areal factor obtainable from Fig. 1.2. In other words, the square of the minimum dimension is taken as the unit of area or pixel available to the designer, and the ratio of the area of the chip to this pixel area is the number of pixels available on the chip. Comparing this with the data in Fig. 1.1 yields the number of pixels per transistor, which is plotted in Fig. 1.3 to show the influence of compaction.

Devices and circuits on highly integrated chips must be characterized by, in addition to low cost, low power dissipation and high reliability. The power that is supplied to a chip is dissipated as heat and must be removed without permitting the temperature of the chip to rise enough to impair the functioning of devices. The power that can be removed from a chip is limited, and high levels of integration would not be possible without a continuing reduction of the power per device. High reliability is necessary because if even one device fails to operate the chip may be useless.

Integration and miniaturization have also proved to be the key to these desirable properties. Miniaturization achieves low power dissipation because the volumes in which electric fields are created and destroyed are reduced; in other words, capacitances are reduced. And it has turned out that the connections made by integrated circuit techniques are far more reliable than those made by conventional solder joints and mechanical plugs. Miniaturization and integration have, in addition, proved to be the key to high speeds, a consequence of the reduction of capacitances and the shorter distances that signals travel between devices.

It should not be hard to imagine that the actual design and construction of integrated chips involves an enormous amount of detail. This detail is handled differently by different manufacturers and users. There is really no such thing as a description of a chip that is universally applicable. Figures such as our Figs. 1.1 and 1.2 above are generalizations based on the assessment of the state of the art at points in time and do not apply quantitatively to any particular factory or line of development. Nevertheless, the progress that they represent is real and typical of advanced commercial practice. It should also be realized that much more advanced technologies than those represented by the products in our figures are being investigated in laboratories, produced in pilot lines, and described in the technical press.

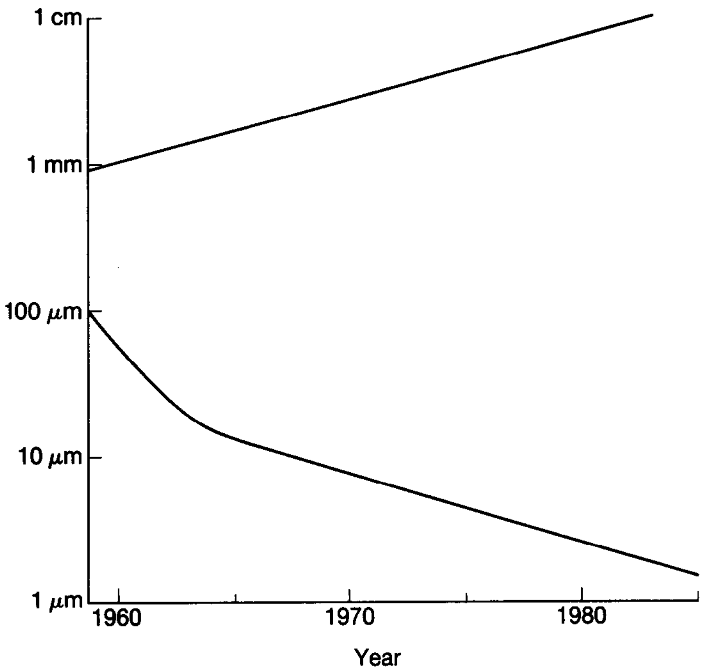


Figure 1.2 Trends in two dimensional parameters that characterize integrated circuit technology. The upper line represents the chip edge while the lower shows the smallest dimension of a structure fabricated on the chip.

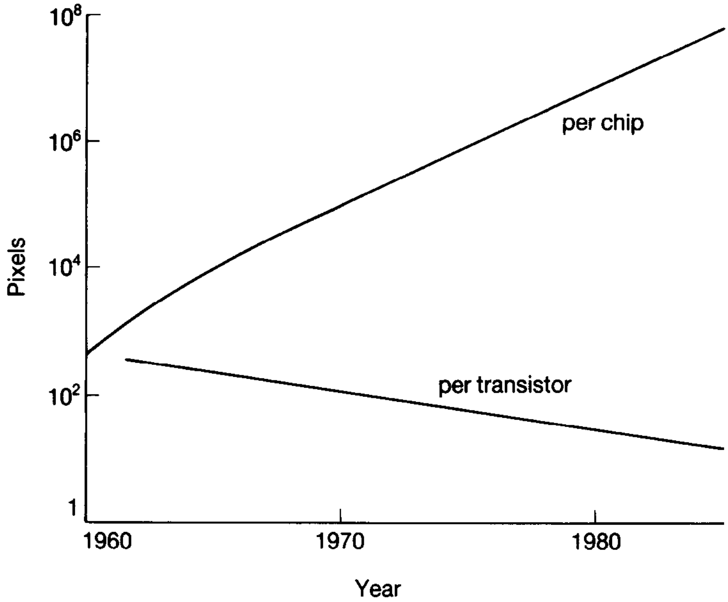


Figure 1.3 Upper curve: The number of resolvable elements on a chip, obtained by dividing the chip area by the square of the minimum dimension, the lower line in Fig. 1.2. Lower curve: The reduction in the number of resolvable elements needed to make a transistor.

1.2 Technology

Silicon is a semiconductor. Transistors are made possible by the fact that the properties of a semiconductor can be controlled by small amounts of impurity. The unique properties of the transistor arise from the existence of two kinds of charge carriers, positive and negative, holes and electrons, that can be produced by doping with impurities. The holes and electrons are very mobile, they acquire a high velocity in an electric field. In a transistor, charges of one sign are held in a place where they attract carriers of the opposite sign that can move through the device. Many more electrons (or holes) can be moved through the device than the number of holes (or electrons) that were needed to induce the mobile carriers; there is a large gain.

Transistors are connected to form logic gates, circuits containing about five transistors that accept electrical signals that represent digits and transmit an output that is some logical function of the inputs. The functional capability of a chip is measured by the number of logic gates that it contains. Sufficient connections to the chip must be available to allow the gates to be used as part of a larger system. The chips are mounted on other surfaces connected with one another. The physical hardware beyond the chip is called the package.

The development of silicon devices into a major industry has provoked a vast body of basic research aimed at clarifying the physics and chemistry of semiconductor materials, and particularly of silicon. Thus silicon has become the most-studied and best-understood solid. The research is greatly assisted by the fact that device technology has developed means for preparing silicon of unusual quality; silicon is available in higher purity and more perfect crystalline form than any other substance. The research and technological efforts complement one another; even properties that would be regarded as insignificant detail in another material can be of great importance in a process that performs hundreds of operations on a wafer that is to contain millions of electronic devices.

Integrated circuitry is based on planar technology, the fabrication of components and interconnections by operations performed on one face of a silicon slice. The operations introduce impurities into the silicon substrate to form devices, and deposit and remove material selectively to form connections and insulation.

A large measure of the success of silicon as a material for electronics is due to its oxide, SiO_2. A layer of oxide is easily formed on the surface of silicon by heating it in an oxidizing atmosphere. The oxide is an excellent electrical insulator and is chemically inert, protecting the silicon from attack by other substances. The inertness of the oxide is used to advantage in device fabrication processes.

1.3 Systems

The low cost and high reliability of integrated electronics has made the collection of many thousands of logic gates into a single computing system possible. The success of semiconductor electronics in rapidly decreasing the cost of elementary components, thereby allowing ever-greater numbers of components to be assembled into powerful systems, is evidenced by the computer revolution of the latter part of the twentieth century. The earliest electronic computers were intended primarily for scientific and engineering studies. However, the steady decrease in cost of use has allowed them to be applied to a very wide spectrum of human activity, spanning the navigation of spacecraft and ships to management of supermarket inventories, word processing, economic forecasting, and banking.

Solid state electronics has also made a great variety of computing machinery available, ranging from the largest supercomputers with prices in the tens of millions of dollars to inexpensive single chips that can be used in video games and automotive ignition controllers.

The growth of the size of the largest general purpose computers in terms of the number of logic gates is shown in Fig. 1.4. These large machines demand rapid access to large amounts of information, and the quantity of electronically readable memory that they use has increased even faster, as shown in Fig. 1.5. These increases in number of components have been accomplished with substantial decreases in the physical size of systems.

The collection of logic chips forms the central processing unit, the CPU, of the computer. The CPU performs the data processing functions and a control function, the regulation and coordination of the flow of information in the system, of which the CPU is only one part. The nature of the complete system is shown in Fig. 1.6. It includes a collection of information storage devices of various types and with varying characteristics. The low cost per bit devices, magnetic tapes and disks, can deliver information to the CPU in times varying from milliseconds to seconds. Access to more expensive electronic memories can be obtained in nanoseconds to microseconds. The system also includes other peripheral devices that interface to the human user: terminals with keyboards and displays and perhaps their own CPU, and printers and plotters. The technology of these components is intertwined with the discipline "human factors" and will not be touched upon here. The human factors aspect of computer systems has not been affected very much by the advent of VLSI, aside from the increasing memory and intelligence available in terminals. Miniaturization of the human interface is not possible; keyboards must match fingers and displays and printed symbols must be large enough to be interpreted by the eye. Still other devices for input and output may be occasionally encountered, such as automated sensors of physical quantities, readers of bar codes and magnetic stripes, pilot lights, and readers of cards and card punches.

Ways to measure the success of microelectronics in meeting the aim of processing information, and of doing so at low cost, are necessarily sought.

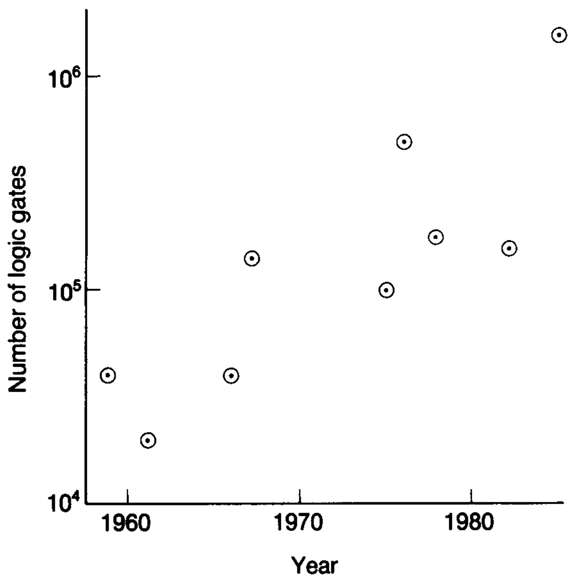


Figure 1.4 The size of large computers measured in number of logic gates as a function of time.

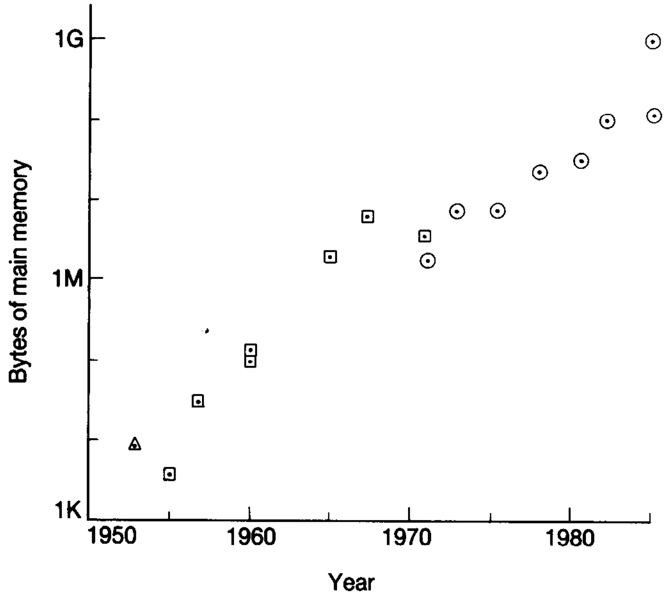


Figure 1.5 The increase in the amount of electronically accessible memory of large computers. Squares represent core memory, circles, semiconductor memory, and the triangle shows charge storage in a cathode ray tube.

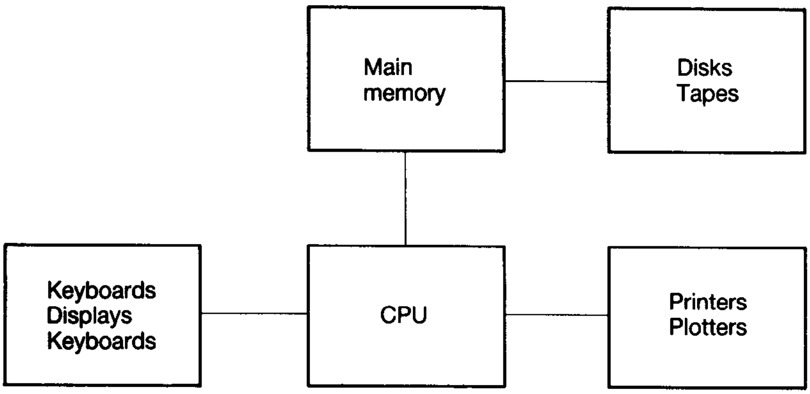


Figure 1.6 The components of a large computing system.

Various measures are given the name "performance." Performance generally refers to the speed with which a task is performed, and it will be used in this sense herein. The task may range from carrying out a single operation on a chip, which can be simply characterized by an average delay, to the time for a large computing system to carry out some step that cannot be quite so objectively defined. The latter may be called an instruction or an operation and performance described in terms of MIPS (Millions of Instructions Per Second),