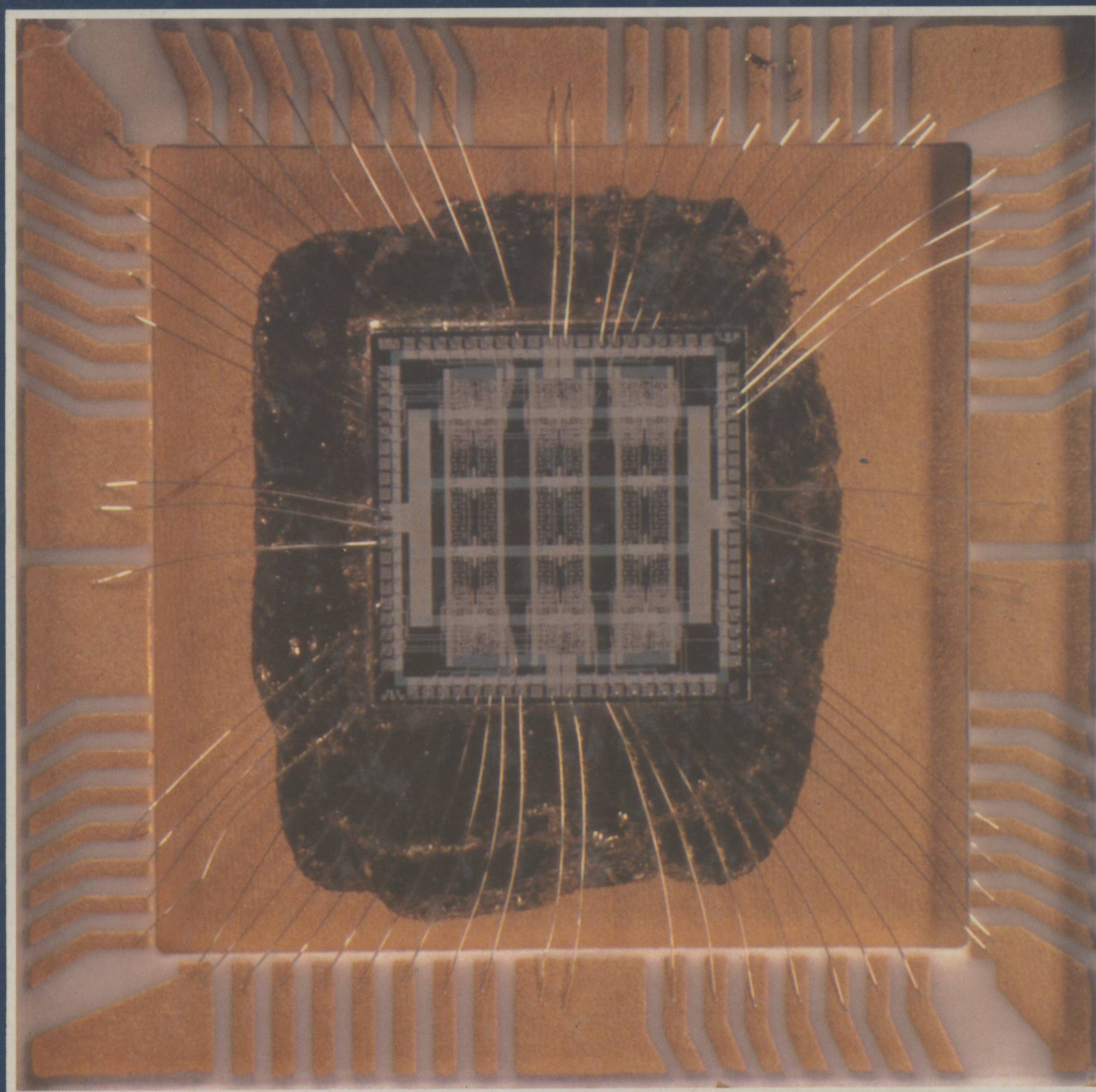
DISCOVERING COMPUTERS

by Mark Frank



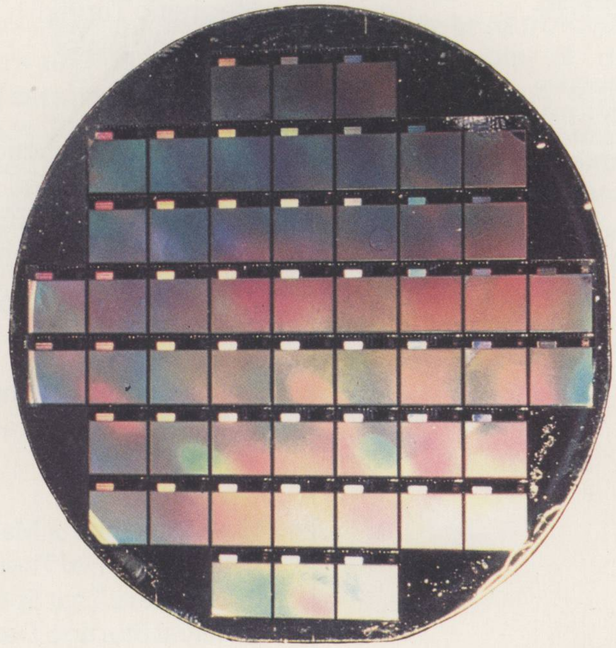
DISCOVERING COMPUTERS

Cover picture: The gray square in the center of the picture is a silicon chip 9/16 inch across. The integrated circuit on which it is mounted is shown in yellow, with the connecting wires clearly visible. The entire piece has been magnified about thirty-five times.

Opposite: A silicon wafer which actually measures three inches in diameter. Each of the small squares is a silicon chip 3/16 inch across, and each contains all the circuitry needed to function as a microcomputer.

DISCOVERING COMPUTERS

by Mark Frank



Published by

STONEHENGE

in association with

The American Museum of Natural History

The author

Mark Frank is a computer instructor, teaching new graduates the fundamentals of computing. He graduated from Queens College, Cambridge, in 1972, with a BA in Philosophy, and gained an MSc in Management Science and Operations Research at Imperial College, London, a year later. Since then he has worked continuously in the computer industry, both as a systems engineer helping to install and operate large computer systems and as an instructor. Among Mark Frank's hobbies is operating his own personal computer.

The consultant

Myron H. Goldberg is Associate Professor and Chairman of the undergraduate and graduate Department of Computer and Information Systems (New York campus), Pace University. He designed and developed the BBA major in Management Information Systems at Pace. Professor Goldberg is a consultant to the New York State Department of Education and to industrial companies in the fields of curriculum planning, course development, and course packaging.

The American Museum of Natural History

Stonehenge Press wishes to extend particular thanks to Dr. Thomas D. Nicholson, Director of the Museum, and Mr. David D. Ryus, Vice President, for their counsel and assistance in creating this volume.

Stonehenge Press Inc.:
Publisher: John Canova
Editor: Ezra Bowen
Deputy Editor: Carolyn Tasker

Trewin Copplestone Books Ltd:
Editorial Consultant: James Clark
Managing Editor: Barbara Horn
Executive Editor: Penny Clarke

Created, designed and produced by
Trewin Copplestone Books Ltd, London

Library of Congress Card Number: 81-51991
Printed in U.S.A. by Rand McNally & Co.
First printing

ISBN 0-86706-005-0
ISBN 0-86706-057-3 (lib. bdg.)
ISBN 0-86706-026-3 (retail ed.)

Set in Monophoto Rockwell Light by
SX Composing Ltd, Rayleigh, Essex, England
Separation by Gilchrist Bros. Ltd, Leeds, England

Contents

The World of Computers

Over the next twenty years computers, more than any other invention, may well change our lives, as computers will perform ever more tasks. But most people still do not understand what computers are or what changes they will cause. This book is written for people who would like to understand how computers work and how they will develop, but have no previous knowledge of computers, mathematics or science.

Computers are becoming faster, cheaper, and smaller. An interesting way of looking at this is to compare the computer business with another industry that developed rapidly – the automobile industry, for example. Computers are developing faster than automobiles ever did. If cars had improved at the same rate as computers since 1950, a Cadillac might now cost $12, have a top speed of over a million miles per hour and run for more than 100,000 miles on a gallon of gas – and its engine would be the size of the period at the end of this sentence. This rate of change in computers is likely to continue for the foreseeable future because they are part of a new and developing technology.

When computers first appeared most people were not aware of their existence. Now computers have become so common that they are used in banks, stores, factories, warehouses, offices, and even in private homes.

Despite this, many people believe that they will never understand or be able to use computers, perhaps because they associate computers with mathematics. But in fact anyone who can add one and one together can understand enough about computers to appreciate how they work and what they are capable of doing. Computers are amazing machines, and some of the new developments in computer science surprise even the most experienced computer experts. Nevertheless, computers need not be a mystery to most people once the basis of their operation is understood. This book explains what computers are, how they work, and what they can and cannot do.

What Computers Do

Machines such as automobiles, telephones and televisions are very familiar, and most people use at least one of them every day. While computers are less familiar, they are affecting our lives just as much as other machines. But what are computers, and what do they do?

Suppose that a man keeps the scores at a track event. He receives telephone calls from the track telling him the details of competitors' performances, such as their names, numbers, and their times. It is his job to note the names of the competitors, record their positions, and display all the results on a scoreboard. This is the kind of task that a computer can do easily, and in fact computers have been used at major sporting events such as the Olympic Games for many years. In this example the scorer is taking in information by telephone, making calculations with that information and then displaying the results on the scoreboard. He is processing information, or data, and that is what computers do. Computers process data; and computer manufacturers say that they are in the data-processing industry.

Computers are not just one machine, but lots of devices connected to each other by electric cables. The data, in the form of electric signals, travels between devices along the cables. When not on the move the data is stored as an electric charge or a magnetic field.

Processing data can be thought of as taking place in stages, passing through various devices. Stage one is *input*. The data arrives at the computer from some outside source – in the case of the scorer it is a telephone call. A computer needs some way of collecting the data as input. It might be typed on a special typewriter which is connected to the rest of the computer by a telephone line.

Stage two is *storing*. The data has arrived in the computer. The scorer would probably write down the names and figures rather than try to remember them. He is storing the data before he works on it. In the same way the computer needs a place to store the data while it works on, or processes, that data. The data would probably go straight into main storage, which is called "main" because data usually has to pass through main storage as it goes from one device to another.

Stage three is *processing*. This stage processes the data for results. The scorer would probably do this mentally; the computer makes the calculations with a device called the arithmetic/logic unit. This is so called because arithmetic and logic are its main functions. In the case of the sports event the computer compares the times of the runners (logic) and works out the order of the winners (arithmetic). The results of the calculations are passed into main storage along with the original data.

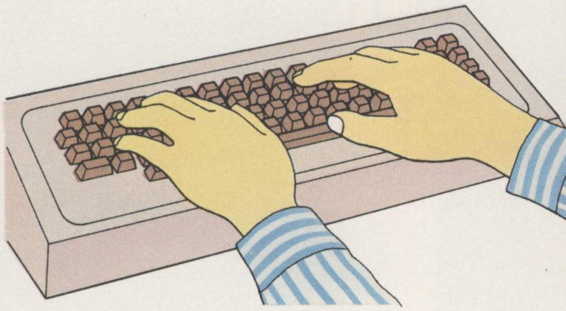
Stage four is *output*. Finally the results have to be put in some useful form. The scorer tells someone to put the names, times and the order of finish of the athletes on a scoreboard. The computer might display them on a screen similar to a television screen or print them on paper.

So each device that belongs to the computer performs something that the scorer does, although, unlike the scorer, the computer parts work with the speed of light. And there is another vital difference between the scorer and the computer. The scorer himself is capable of deciding what to do with the data, and does not need to be given instructions. All the devices that make up the computer, on the other hand, need very detailed instructions at every stage to know what to do with the data. These instructions have to be put into main storage before all the data, and these instructions are what is meant by a computer program. Every computer has to have a program in main storage before it can do anything.

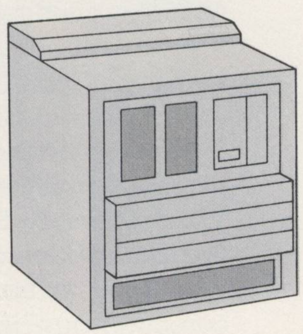
This means that there has to be an extra part of the computer that reads the instructions, works out what they mean, and directs them to the appropriate device. This component is called the control unit. It is usually in the same box as the arithmetic/logic unit, and the two together are called the processing unit or processor.

The four stages of a computer processing cycle:

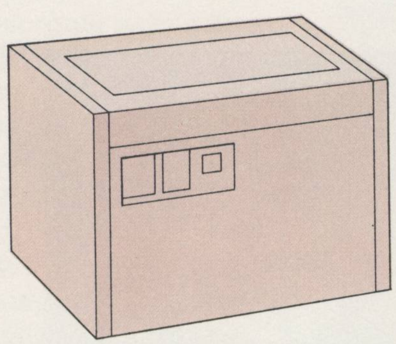
(1) Input – A scorer at a track event receives the names, numbers and times of the competitors by telephone; the computer receives the same information when someone types it on a typewriter like keyboard that is linked to the computer.



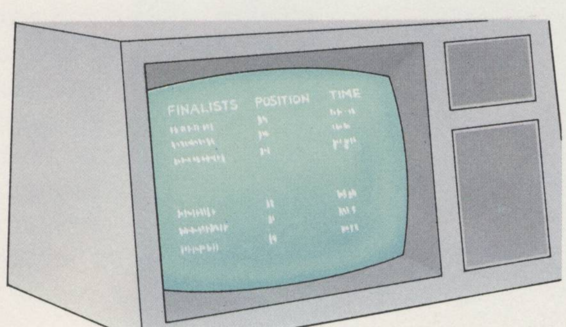
(2) Storing – The scorer writes down the information so he can refer to it while he works on it; the computer passes the information into main storage for the same reason. Main storage looks similar to a refrigerator.



(3) Processing – The scorer compares the times of the competitors and works out the order of finish; the computer makes the same calculation in the arithmetic/logic unit that is contained within the processor.



(4) Output – The scorer tells someone to display the results on a scoreboard; the computer passes the information to an output machine called a VDU (visual display unit) that shows the results on a screen similar to a TV screen.



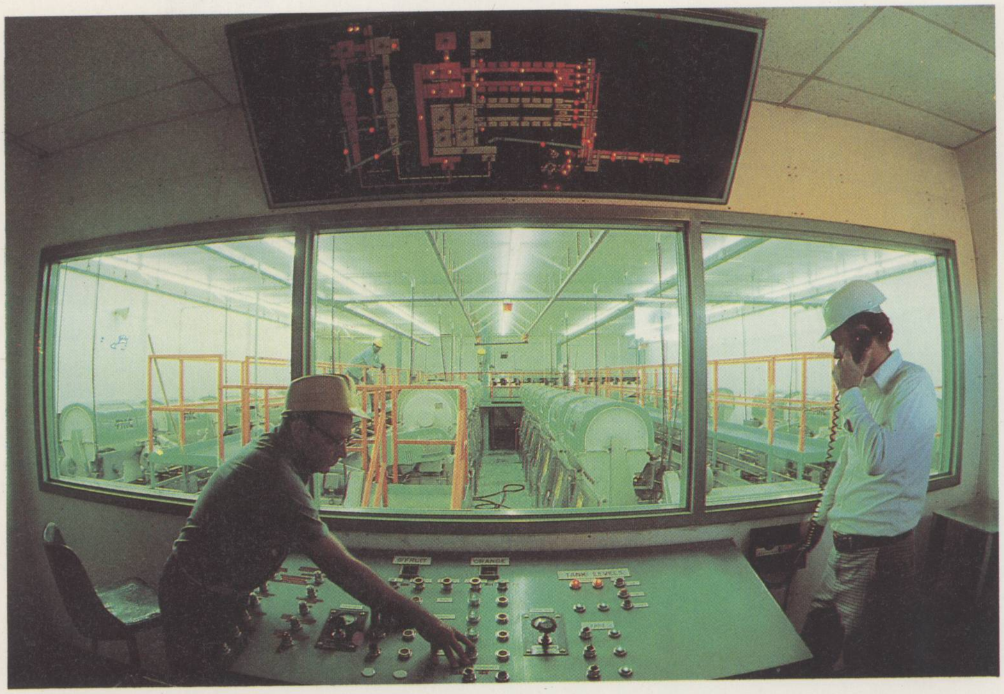
What Computers Are Used For

The first modern computer was in operation in 1946. By the mid-1950s a few companies were manufacturing computers commercially. The companies making these early computers each believed they would sell only about ten in the entire world. They thought computers would be used only by scientists to make long and complicated calculations. And, indeed, scientists still use computers extensively. Without the help of computers, men would never have reached the moon, for example, and flights into space would be impossible. Scientists need computers to calculate the route a spaceship will take and to make corrections in the course of the journey, among other things.

However, it did not take long for commerce and industry to catch on to the scientists' new tool, and to use computers differently. Instead of making long calculations on a few items of data – a physicist using a complicated mathematical formula, for example – they need to make a few calculations again and again on lots of data – to keep account of the different types of stock in a company's warehouse or to record the number of items delivered and dispatched.

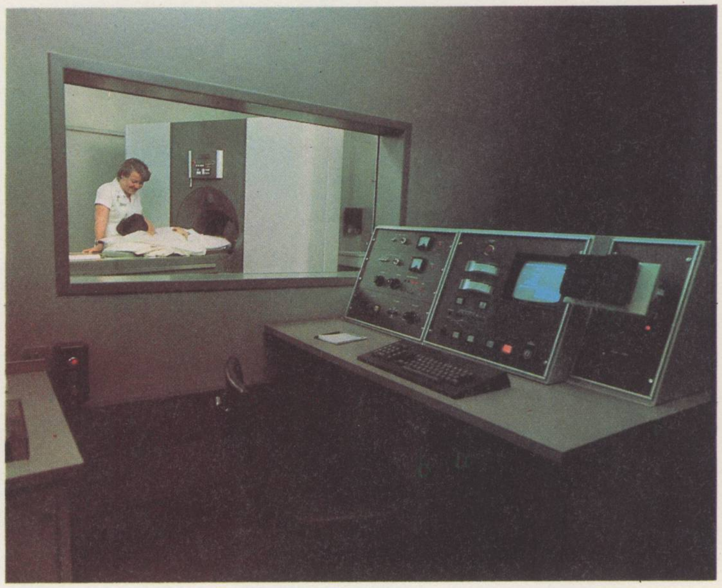
In fact, for a long time computers were divided into two types: scientific, which were quick at doing arithmetic but did not have many input and output facilities, and commercial, which were slower at arithmetic but had more flexibility in input and output. Then in the middle of the 1960s computers became so flexible that manufacturers started to make one that was good at both kinds of work. Nowadays most computers are like this.

Computers were first used commercially to help to do jobs such as keeping a record of how much



The panel of a computer that is used to control the production of orange juice. It indicates the levels of the juice in the tanks and controls the extraction process. The lighted panel above the window shows the flow of juice throughout the factory.

The business management student is using a computer-linked visual display unit, similar to a TV screen. The computer poses questions and then flashes answers up on the screen. The student touches the answer he thinks is correct on the sensitized screen.



A computerized scanning machine takes detailed X-ray pictures of a patient, seen through a window of protective glass talking to an attendant. The screen on the operating console in the foreground displays data, while the X-ray picture is seen through the black viewer at right.

money each customer owed a company and printing the bills, or working out the shortest route for trucks delivering goods from a factory to the stores. Now, most companies have computers to do jobs such as these which would take a great deal of time if people had to do them. In fact, almost everything you buy today has been produced by an organization that relies on computers.

Computers are also used because they can react very quickly to a change of circumstance. They can physically control a process – the flow of oil in an oil refinery, for example. Electric sensing devices tell the computer the state of the process. If anything unusual happens – the amount of oil increasing or decreasing suddenly – the computer will react either by displaying a message on a screen as a warning or by sending instructions directly to some other machine, such as a valve, which alters the supply of oil so that it flows normally once more. Such a use is called a process control, and the important thing is that the computer can work fast enough to keep the process controlled.

A technician at the control board of a computer directs drilling machines that make underground pipeline tunnels. The screen and dials show the progress of the operation.

Main Storage and Bits

It is the ability to store vast amounts of information and to submit that information for fast processing that gives the computer its power. Main storage is at the heart of the computer system because a computer has to store the data and programs in main storage before the processor can work on the data. Symbols have to be used to store the data and programs in main storage. The scorer of the track and field event (pages 6–7) might store a competitor's time by writing down "21 seconds." In this case he is using eight different symbols, the digits 1 and 2, and the six letters c, d, e, n, o, s. This is typical of the method we employ to store data in written form. We use a selection from the ten digits (zero through nine), the twenty-six letters of the alphabet and other symbols such as punctuation and plus and minus signs. All the digits and all the letters are known as alphanumeric characters.

Computers, however, use only two symbols when they store data: zero and one. It is perfectly possible to use just these two symbols and record all the data needed. This is done by replacing all the characters with a code of zeros and ones, in much the same way that Morse code replaces letters and numbers with a code of dots and dashes. Computer people call zero and one "bits" (from *b*inary dig*its*, as the binary system of mathematics uses only the digits zero and one). All computers store data as bits using just these two digits to represent every letter of the alphabet and number in the decimal system.

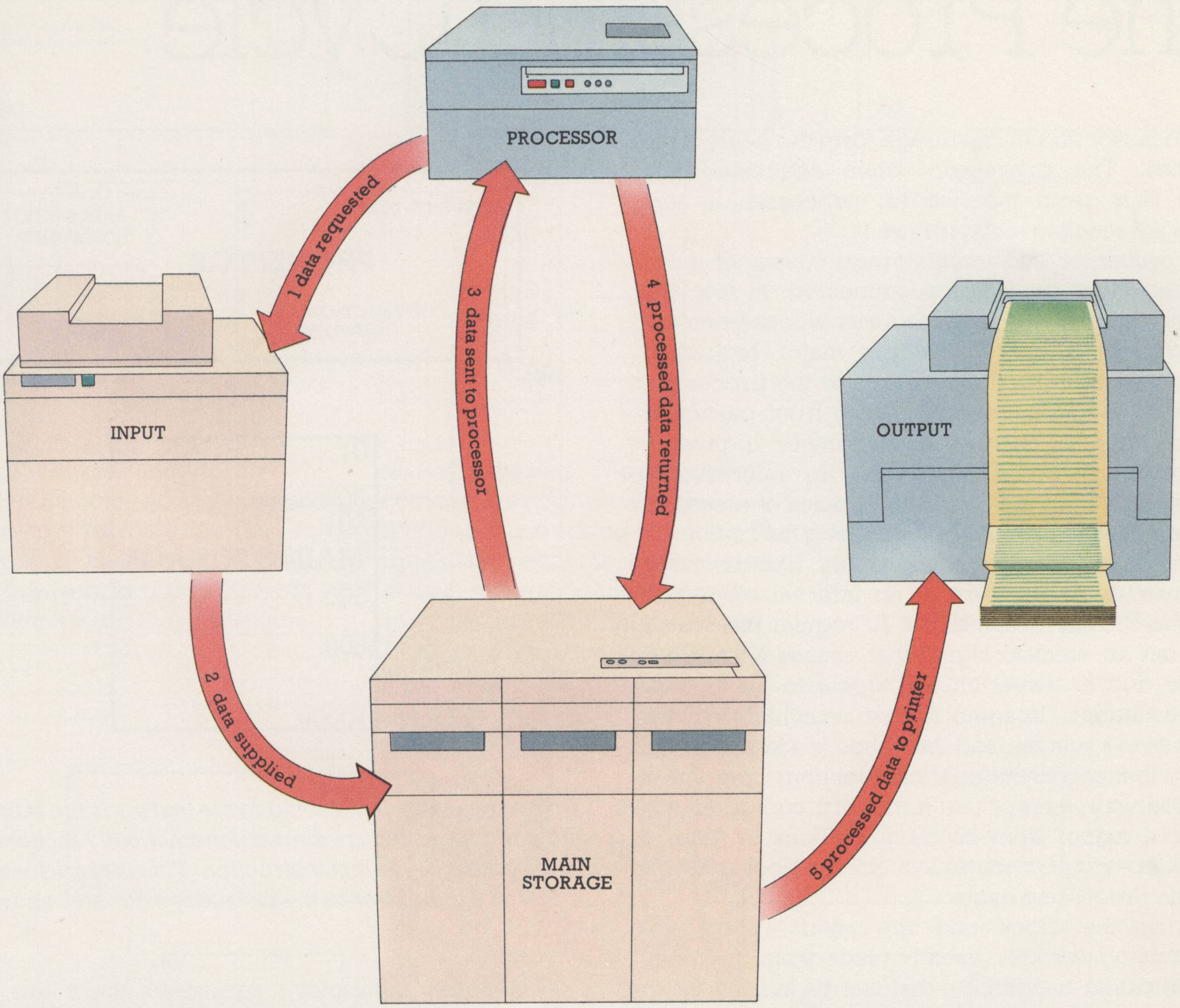
Modern main storage consists of thousands of tiny pieces of metal that can be either full of electricity (charged) or empty of electricity (uncharged). The state of these pieces of metal is used to record data and programs. The computer designer decides either that being charged will count as the one bit and being uncharged will count as the zero bit, or the other way around. It does not matter which way. There are other storage devices besides main storage, and they also use mechanisms that can be programmed into one of two states: for example, a magnetic tape with spots that can be magnetized or unmagnetized.

When the data and program are in main storage, there has to be a means of reaching the appropriate data when it is needed. If you want to get information from a library you do not go through every shelf until you find the right book. There is normally a catalog system that directs you to the shelf where the book you want is, In the same way, computer main storage is split into locations, each of which holds one item of data and has a number called its address. The address is permanent; it is given to the location when the main storage is manufactured.

The program, once stored, will instruct the other parts of the computer system to work on the data by using the address of a location. For example, the program might instruct the input device to store an item of data at address 200. Then it might tell the processor to add the data at address 200 to the data at address 300 and put the result into address 200 in place of the original information. Finally it might tell an output device to take this new data from location 200 and print it. So a computer program has

Addresses	Bits in Locations
1	11110101
2	10111101
3	10001110
4	10101011
5	11110110
6	10110110
7	10111101
8	11010110
9	01011001
10	01110100
11	11101000
999995	01111011
999996	11101110
999997	01110101

This stylized presentation of main storage shows how main storage is divided up into small areas called locations. Every location has a numbered address that is used to refer to the location. The zeros and ones within each location are the bits that represent one item of data.



to keep careful track of the address of each item of data. As well as the data it works on, the program itself is in main storage, so all the program's instructions will also be in locations and have addresses.

The fastest storage devices work at the same speed as the processor, so that the processor does not have to wait for information. But such storage is expensive to provide in large enough quantities. Because of this, slower and cheaper devices that back up the main storage are used for long-term storage. These devices keep the data and programs until needed, when they are transferred into main storage.

This diagram illustrates how main storage operates doing a simple computer exercise: the addition of two numbers. Once a program has been started by an operator, the processor first instructs the input machine to supply main storage with a number. On receipt, main storage sends that number to the processor along with another number already in main storage and instructs the processor to add the two numbers. The result comes back; main storage reads it in and then instructs a printer to print the result. Because all instructions and items of data must be in main storage before they can go to any other device, main storage is central to all computer operations.

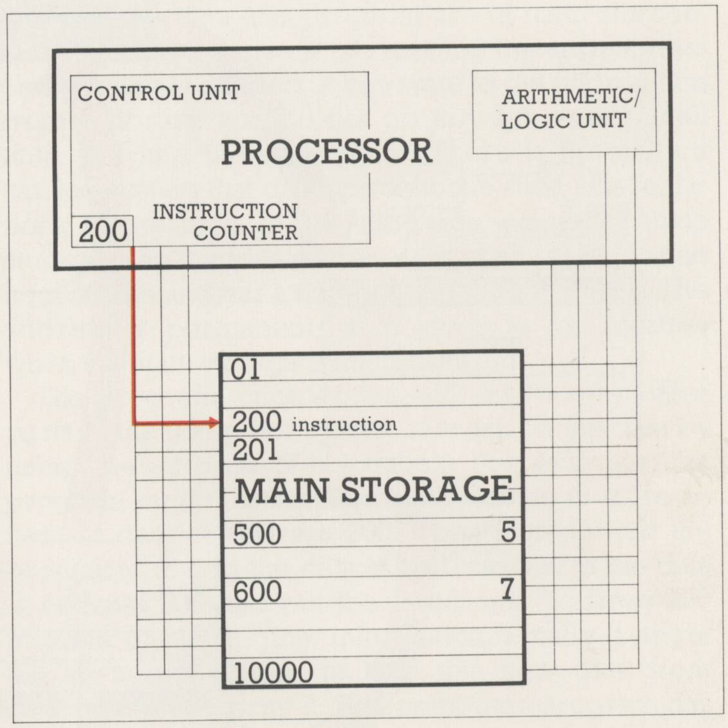
The Processing Cycle

The processor and main storage form the heart of the computer. The processor, when requested, retrieves data from main store, processes it, and returns the result to main storage.

The processor and main storage consist of many parts and these must all be connected. In fact, the computer can be compared to a city whose parts are connected by roads. There are major highways, such as the one from main storage to the processor, and minor footpaths, such as those from particular locations in main storage to the major highways. The roads can be cables, wires, or microscopic pieces of metal on a silicon chip. Pulses of electrical energy, representing the bits, go along the roads.

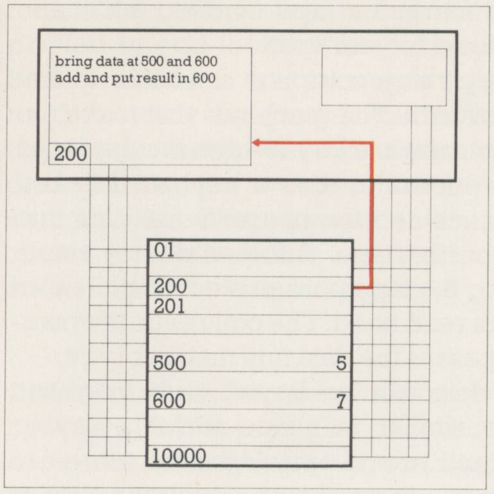
The bits do not travel continuously, like the traffic in a town. Instead, there is an internal clock that regulates the traffic of the bits. At regular intervals it sends out an electric signal that causes all the bits that are due to travel on this signal to go to their next destination. Imagine a town in which a clock chimes every minute, and the chime is a signal for all the cars to move to the next intersection. Computers work this way, except that a modern computer will receive a signal from its clock millions of times a second. In every computer system this clock controls the basic processing cycle.

Besides the clock and the roads, there are thousands of switches, usually made from tiny components called transistors, that can be turned on or off at every signal. They are used to control which bits are read or written. For example, there are switches that are used to select locations in main storage. If the switch for a location is on, bits can travel to or from that location. If the switch is off, then the bits in that location will not be read. In the processor other transistors are connected together in groups that are used to process the data. For example, one group adds two numbers together. When the program instructs the processor to add the two numbers, the bits that represent the numbers are sent through this group of transistors. The result returned to main storage is the sum of the two numbers.

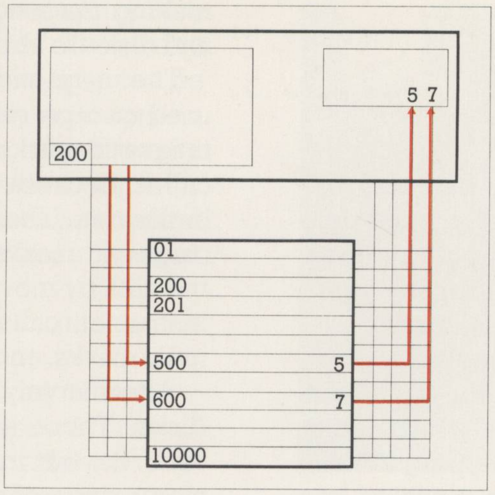


The steps in a processing cycle that adds the numbers 5 and 7 together are shown schematically in these six diagrams. (1) Select instruction. This is at address 200 as the number in the instruction counter shows.

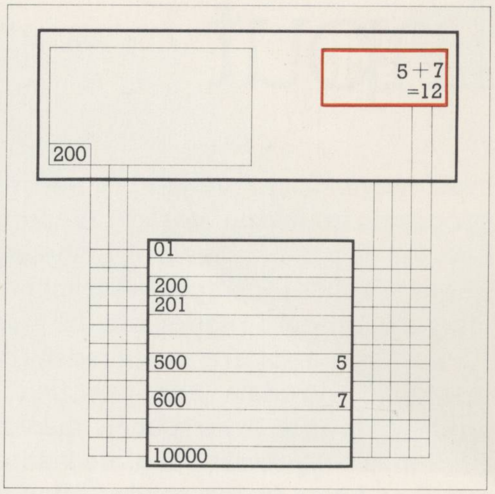
In every computer's processor there are small storage areas called registers. They hold bits in the same way that main storage does; but they are used only for special purposes. For example, there is one called the instruction counter which is used to hold the address of the next instruction to be obeyed after the present instruction has been completed. Say a computer is carrying out a program. It has just finished performing the instruction at address 199 and now it is about to execute the instruction at address 200. This is an instruction to add the number at address 500 to the number at address 600 and put the result back in address 600. The instruction counter contains the address of the instruction to be executed, that is, 200.



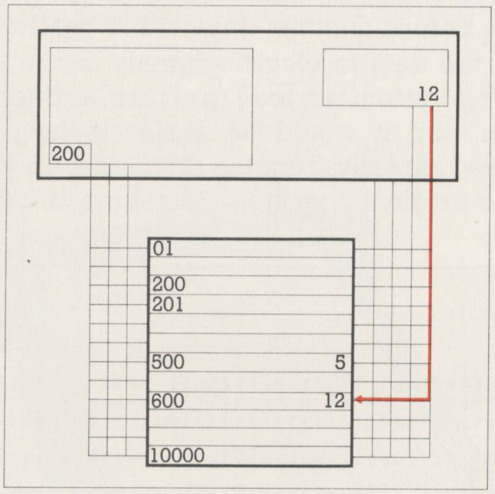
(2) *Bring the instruction. The switches that give access to address 200 are opened and allow the instruction contained in the location to go to the instruction register in the control unit of the processor.*



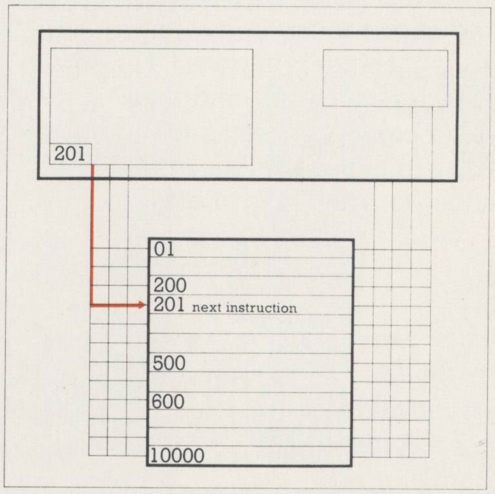
(3) *Bring the data. The instruction contains the addresses of the numbers to be added together. The processor sends these addresses to main storage, which opens the switches that allow the data in the locations to pass to the arithmetic/logic unit.*



(4) *Add the numbers. A group of transistors in the arithmetic/logic unit adds the two numbers together and puts the result in a register.*



(5) *Store the result. The control unit sends to main storage the address of the location in which the data is to be stored. This address opens switches that allow the data to pass from the arithmetic/logic unit into the location of that address. The data replaces the data originally in the location; the location at address 500 remains unchanged.*



(6) *Point to the next instruction. At the end of the cycle the instruction counter will contain the address of the next instruction to be executed, 201.*

Input

Input devices are needed to get the data and the programs into main storage. They convert the data into the electric signals that the computer needs. There are different input devices, but for many years the punched-card reader was the most common.

Punched cards are so called because holes representing the data are punched into them. An operator works a keypunch machine which looks similar to a typewriter but, instead of typing letters, punches holes in the cards. After the machine has punched the holes, the cards are placed in an input device called a card reader, which is part of the computer system. The card reader converts the data that is in the form of holes into electric signals, which will then pass into storage as zero and one bits. It does this in one of two ways. One way uses small metal brushes to "feel" the holes, which are then converted by the card reader into electric currents. The more modern method shines a light onto the card as it moves through the machine and registers a current whenever the light shines through a hole. In each case, the card reader converts the data into signals and passes the signals into main storage.

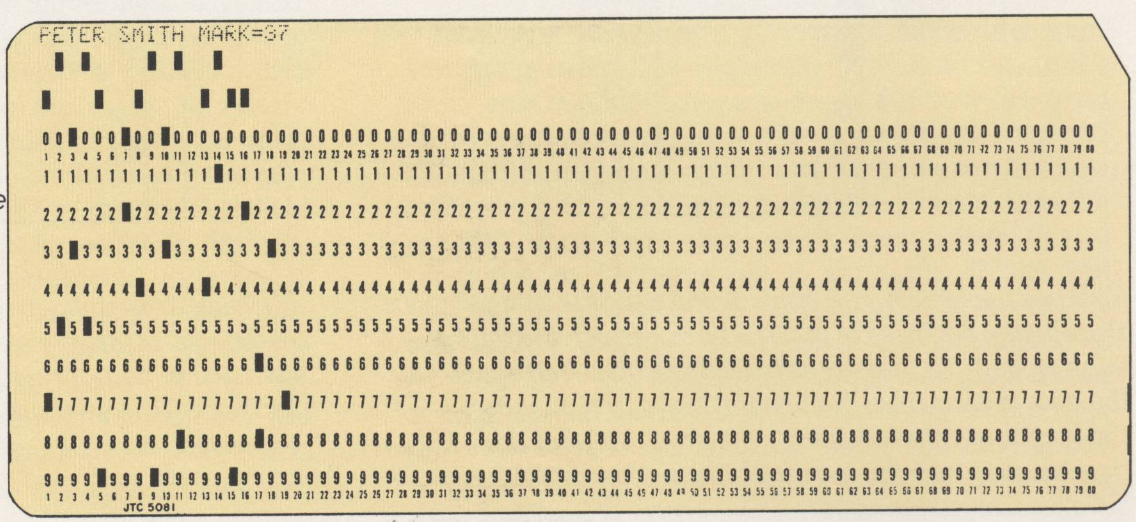
A fast card reader can read and convert twenty cards a second. In computer terms this is slow, as a processor can do a million additions in that time.

The alternative to punching holes in cards is making magnetic marks on tape or disk. Both tape and disk are also used for storage.

The magnetic tape on a computer is similar to that used on tape recorders. The machines that make the magnetic marks on tape are key-to-tape machines, so called because an operator uses a keyboard to key in the data; the machine then converts the data into magnetic marks on the tape. Another device reads the data by moving the tape past an electric coil on an instrument called a read head. The coil reads the magnetic marks, and passes the data into main storage.

Alternatively, data can be keyed onto magnetic disks. These are similar in shape to long-playing records, but instead of one spiral groove there are many rings of magnetic marks on which the data is recorded. There are two types of magnetic disks: large rigid disks and small flexible ones. These flexible disks are often known as floppy disks or diskettes. They can be carried easily and even sent by mail. Both types of disks can be used for input in a similar way to magnetic tapes. The data is keyed onto a disk by one machine and then read into the computer by a different machine that moves the surface of the disk past an electric coil.

All these methods have something in common. A person has to key the data onto cards, magnetic tape, or disks before another device can read them and convert the data to electric signals inside the computer. This system can lead to errors, as people make mistakes, and it would be easier if the computer could read directly from the document on which the data was originally written. Sometimes it can. If you

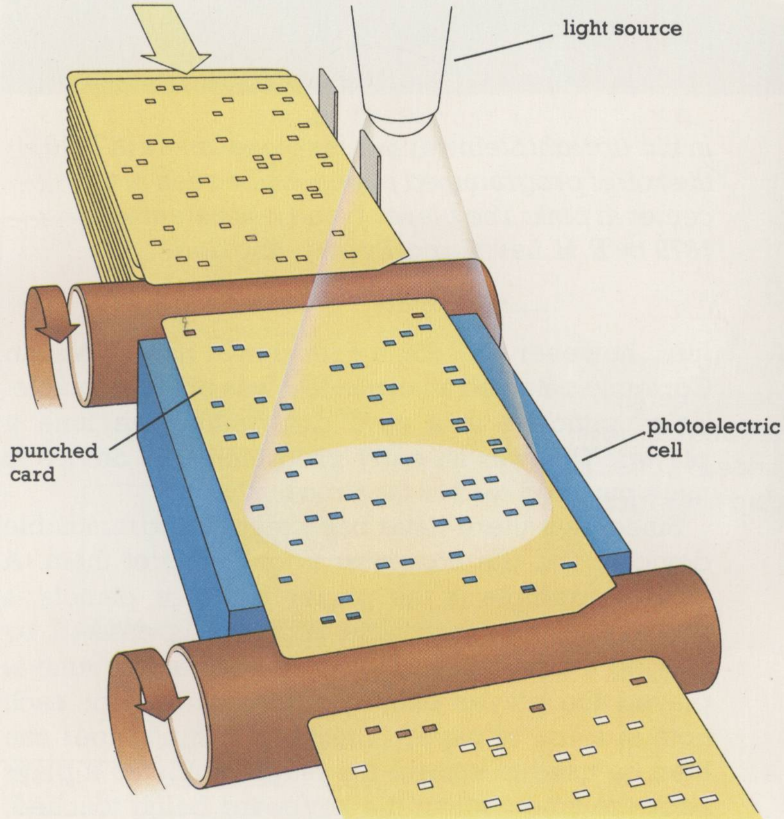


This punched card carries the name "Peter Smith" and his examination mark "37." The eighty columns are numbered in small figures in the second row from top and in the bottom row. Each column of holes represents one character. The characters are printed at the top of the card so it is possible to read what the holes represent.

look at the bottom of a check you will see characters printed in odd-looking lumpy numbers. They have been printed in a special magnetic ink so they can be read not only by people but also by a magnetic character reader on the computer, which easily distinguishes the numbers by the pattern of their straight sides. There are also machines called optical character readers that can read ordinary printed documents or even handwriting.

This keypunch operator is typing letters and numbers on the keyboard. The machine converts the letters and numbers into holes it punches in the cards.

Optical card readers are modern versions of the punched-card reader. The punched card still has rows of holes, each row representing a different character. The reader, instead of having metal brushes, has a series of photoelectric cells, each cell corresponding to a row of holes. As the card moves through the reader across these cells, a light shines down on it. Where there are holes, the light shines through and causes an electric current in the corresponding photoelectric cell. The current is turned into the bits representing the character which are then transmitted to main storage.



An increasingly popular form of input is the floppy disk, here shown being inserted into the disk drive of a small computer system. This disk can hold 576,000 characters of data, others can hold as many as 1,100,000 characters.

Programs

The program is a series of instructions written to make the computer perform a task, such as working out what raw materials a factory has to buy in order to make the following month's products. At first the instructions may be written on paper by people, and then they may be stored on magnetic tape or disk, but they will eventually have to be copied into main storage, probably by another program already in main storage, before the computer can obey them.

A large program, such as the one that controlled the space shuttle, can involve a very large number of programmers for many years and cost a great deal of money. In most computer systems the programs that are instructing the computer cost more to produce than the machines themselves. What is more, the cost of the machinery is decreasing while the cost of programming is increasing as a result of labor charges. There was a time when manufacturers gave away most of their programs free with the machinery. Nowadays this is rare, and programs are bought and sold like any other product.

There are terms for all the types of products, or wares, sold by computer manufacturers. Hardware is the physical machinery in a computer system, which includes processors, disk drives, main storage, card readers, magnetic tapes and magnetic disks. Software denotes the programs that go with it.

Software is as important as hardware, and the boundary between the two is not as sharp as it might seem. There is a middle road, called firmware. In this case the processor is built to obey only a few simple instructions, such as moving a bit from one location to another. Then the manufacturer writes a program that uses these simple instructions to perform more complex instructions, such as dividing one number into another. This firmware is a part-program that is in the machine all the time, and forms the basis of all subsequent programs.

The computer is said to be a programmable device, and without a program it cannot work. There is no such thing as a computer that does not need a program.

Computers are not the only programmable machines, however. In 1852 a Frenchman named Martin Corteuile patented an organ that was programmable. Holes punched in a card determined the tune it played. This was an early programmable device to use a punched card as the program input.

Since then there have been many programmable devices, and the computer is only one of them. A famous example is the player piano or pianola, a programmable piano. The program is created by playing a tune on the keyboard, and as the tune is played the player piano punches a hole for each note in a roll of paper. The same roll of paper can then be used to control the player piano so it plays back the tune without the keyboard being touched.

In this upright Steinway player piano, made in 1910, the roll of programmed paper can be seen in the center in pink. The player piano was patented in 1879 by E. H. Leveaux of Surrey, England.