PROCEEDINGS OF THE EASTERN JOINT COMPUTER CONFERENCE

PAPERS AND DISCUSSIONS PRESENTED
AT THE JOINT COMPUTER CONFERENCE
NEW YORK, N. Y., DECEMBER 10–12, 1956

THEME: NEW DEVELOPMENTS IN COMPUTERS

SPONSORS
AMERICAN INSTITUTE OF ELECTRICAL ENGINEERS
Committee on Computing Devices
THE ASSOCIATION OF COMPUTING MACHINERY
THE INSTITUTE OF RADIO ENGINEERS
Professional Group on Electronic Computers

Published by
American Institute of Electrical Engineers
33 West 39th Street, New York 18, N. Y.

Design Objectives for the IBM Stretch Computer

S. W. DUNWELL

IT IS APPARENT to every user of an electronic computer that there are areas in which major improvements are desired. Present computer speeds, remarkable as they are, still fall short of the requirements of many of today's business and technical applications. The labor of programming often delays or prevents the application of new problems to the machines. Maintenance continues to consume a significant fraction of the working day, and requires highly skilled engineers to locate and repair machine faults.

The experience gained by the builders and users of computers, together with the advent of new solid-state components, now makes possible computers which are significantly better in all of these respects. Project Stretch has the goal of producing such a system for both technical and business use. As the name implies, we are endeavoring to employ the most advanced techniques and components possible with today's technology Many of these techniques are still in the research phase of their development.

International Business Machine Corporation recently contracted to deliver the first Stretch system to the Los Alamos Scientific Laboratory of the Atomic Energy Commission. A joint mathematical planning group has been set up to examine the advanced problems which future machines will be required to solve, and to assist in developing the Stretch system along these lines.

Since the project is still in the research phase, a number of important design decisions are still to be made. While confident that continued research will provide a satisfactory solution in each case, it is recognized that these decisions may have some influence on the ultimate speed of the computer system.

It is estimated that the operating speed of the system on typical technical applications will be at least 100 times that of the fastest general purpose computer now in use. This will be obtained through a combination of higher speeds for arithmetic and logical operations, multiplexing to allow several parts of the system to operate concurrently, and new instructions designed to reduce the number of steps required for common operations.

It is evident that a computer system operating 100 times faster than today's fastest machines and consuming in the process a million or more instructions a second, cannot be programmed effectively by any process which demands that a human write each instruction to be executed. The program will have to be assembled largely by the computer itself from a much simpler statement in mathematical and logical notation. A system of computer-assisted programming, or "automatic programming" as we call it, will be developed concurrently with the development of the instruction system for the Stretch computer, so that the least

S. W. DUNWELL is with the International Business Machines Corporation, Poughkeepsie, N. Y.

possible human effort be required for programming.

Experience with solid-state components indicates that their use alone will greatly increase the reliability of the Stretch system over that of vacuum tube machines. This, combined with a greater ease of locating and correcting those falts which do occur, will have an important bearing both on the amount of useful work which the system can accomplish and on the size of the engineering staff required to maintain it. It is planned that the computer include a thorough system of automatic checking, together with means for indicating within narrow limits, the location of any part of the mechanism which is creating errors. Since automatic checking cannot insure the accuracy of the input data or the program, it cannot insure the accuracy of the computed results. Instead, its presence will be primarily to facilitate the repair of the machine. Automatic error correction also will be provided in several areas. It will, for example, be used in association with the magnetic tapes to insure a level of accuracy which cannot be obtained in any other way.

The system will consist of three major computer sections, all of which communicate with one multi-section memory. The operations required of the computer system will be divided so that each section performs a portion of the over-all task. The three sections of the computer are an input-output section to maintain communication with the individual input-output units, a serial computer section for editing the flow of input-output data, and a high-speed parallel arithmetic section to operate upon the organized data of the problem. The division of the computer into separable major sections is sufficiently unconventional to deserve explanation here. From the standpoint of engineering design, it allows the choice of the most appropriate mechanism in each part of the system. This is evidenced by the choice of a serial arithmetic mechanism for the complex editing of data, and of a parallel arithmetic system for high-speed operation upon previously organized data. Furthermore, separation of the computer into sections enables us to schedule efficiently the design and manufacture of each section.

The section which provides communication between the various input-output units and the computer memory is known as the "exchange" because of its functional resemblance to a telephone exchange. The mechanism is capable of communicating simultaneously with many input-output units. These will include the various kinds of computer input-output devices which we know today, such as magnetic tapes, paper tapes, printers, manual keyboards and typewriters, card readers and card punches. New input-output devices being designed for the Stretch system include a magnetic disk memory capable of holding one million words of data, and of communicating with the computer at the rate of one word each 4 microseconds. These disk memories will be available in multiples to supplement the high-speed ferrite memory. Also, magnetic tape units capable of operating at nearly the same speed are planned for availability by the time the first system is delivered.

Particular attention is being given to making the input-output system as general as possible, so that it can readily accept the input-output devices of the future. It is anticipated that these will include electronic printers and plotters, units for direct visual display of data, and devices for the direct reading of typed input data. The interpretation of speech and handwritten data are recognized as being less immediately achievable, but it is assumed that they will become possible within the life span of the Stretch system. The system will be capable of operating on-line with analog and digital devices of many kinds, and of being fitted into business systems requiring direct over-wire interrogation of large memory reservoirs from remote locations. As an example, an interrogation typewriter might communicate with the computer over long distance telephone lines to obtain company-wide availability data on any desired product.

The serial computer section will operate in both decimal and binary arithmetic modes. It will include the full complement of instructions necessary for fixed- and floating-point operation, and for handling data with fields and records of variable length. Its use to edit the flow of data into and out of the high-speed arithmetic section will be facilitated by instructions designed specifically for converting between binary and decimal radices, and for converting between the codes used by the individual input-output units and a preferred code within the computer. It will also include means for accepting instructions and data words in the original format in which they were transcribed by humans, and of condensing them into the more efficient format used within the computer. In a similar manner, results will be expanded into the format necessary for printing and ready for human comprehension.

Because the arithmetic unit of the serial computer section will operate on data words of varying lengths, the speed of arithmetic operations can be expressed best as ranges. Addition will require 2–3 microseconds, and multiplication 5–15 microseconds. Logical operations will be more comprehensive than the transfer of control operations of present computers, and will often perform the equivalent of several present logical operations in 2 microseconds.

The complex nature of typical editing processes requires that the serial section provide all of the classes of operations of which the system as a whole is capable and perform them at considerable speed. It is, in fact, a very versatile computer within a computer. The serial computer alone has an operating speed at least 10 times that of present day vacuum tube machines, such as the International Business Machine types 704 and 705.

The high-speed arithmetic mechanism has the function of rapidly executing the operations called for by the mathematician in his description of the problem. Since the other sections of the system relieve it of all secondary responsibilities, the high-speed computer section will be able to perform useful work on the arithmetic portion of the problem at between 75 per cent and 90 per cent of its theoretical maximum rate. This compares very favorably with present computers, which spend a large part of their time on secondary operations. Addition and subtraction in floating point notation will require 0.6 microsecond. Multiplication will require 1.2 microsecond. Approximately 0.2 microsecond must be added to these times for each data word transferred over the bus system between the arithmetic unit and the memory. Specific times cannot be given for logical and housekeeping operations, since they will, to a large extent, be executed simultaneously with the arithmetic of the problem.

In order to achieve the level of performance which has been described, it is necessary that all of the active components in the system be new and of types not now in use. The transistors will be what we know as the 10-megapulse class, by which is meant that they can be used with a 10-megacycle repetition rate in common devices such as adders. These transistors in both the p-n-p and n-p-n types are available in limited quantities for circuit design, where they are exhibiting, under test, rise and fall times short enough to permit their use by Stretch. The transistor circuits to be used are discussed in a paper by Mr. R. A. Henle.[1]

Two classes of ferrite memory will be used. Both of these memories are discussed in a paper by Mr. W. W. Lawrence.[2] One class of memory will be provided in units of 8,192 words and will have a full cycle time of 2 microseconds. When reading, a data word will be available for use at the end of 0.8 microsecond. The second class of memory will be provided in units of 512 words and will have a full cycle time of 0.5 microsecond. When reading from this memory, a data word will be available for use at the end of 0.2 microsecond of its cycle. To further increase the speed of operation of the memory, multiple sections in each class will be provided, with each section capable of operating concurrently with the others. It is recognized that a large memory is one of the most important needs of the computer of the future. For this reason, the addressing system of the Stretch computer will be designed for the ultimate use of randomly addressable memories of up to a million words capacity. In addition, the external memories in the form of magnetic disks and of magnetic tape devices may ultimately provide data in blocks up to a total capacity of possibly 100 million words.

The many investigations necessary to carry the project through its present research phase are proceeding in close conformity to the original schedule, and we look forward with considerable confidence to being able to put the Stretch computer into production in the near future.

References

1. Proceedings of the Eastern Joint Computer Conference. AIEE Special Publication T-92. "High-Speed Transistor Computer Circuit Design," **R. A. Henle.** May 1957, pp. 64–66.

2. *Ibid.* "Recent Developments in Very-High-Speed Magnetic Storage Techniques," **W. W. Lawrence, Jr.** May 1957, pp. 101–03.

High-Speed Transistor Computer Circuit Design

R. A. HENLE

THE REALIZATION of the computer described by S. W. Dunwell[1] calls for high-frequency circuits capable of passing a signal through five sequential logical stages in 100 millimicroseconds. At the start of the project a study of available devices and techniques indicated that this goal would be difficult to reach without significant improvements either in the devices or the circuits, or both. Such an improvement came about in the device area with the availability of drift type transistors for research work. The characteristics of these transistors made necessary a new approach to circuit design which has resulted in circuits which not only meet the speed requirements, but also are relatively simple, reliable, and insensitive to noise.

Drift Transistor Characteristics

Typical electrical characteristics of drift transistors are shown in Table I. It can be noted that the transistor has a number of advantages and disadvantages when compared with other high frequency transistors. The major advantages are the following: 1. A very high frequency response: 200 to greater than 500 megacycles. 2. A high collector breakdown voltage: greater than 50 volts. 3. A low base restistance: less than 50 ohms.

These characteristics represent significant improvements in some of the switching parameters of transistors. The disadvantages of the transistor are a high collector saturation resistance: 50 to 150 ohms; and low reverse emitter breakdown: 0.5 to 5 volts.

The seriousness of the disadvantages depends to a large extent on the circuit design philosophy chosen. The collector saturation resistance becomes of major importance in saturating type circuits, and the low emitter breakdown voltage does not allow certain types of emitter-follower logical circuits to be used without modification. However, in some types of circuits, where it is desirable to limit the reverse bias on an emitter, this characteristic can be used to advantage. An example of this is an inverter circuit design such as is shown in Fig. 1. In the usual circuit of this type a diode is used to limit the reverse bias on the emitter junction, thereby reducing the circuit's turn-on delay. With drift transistors, the low emitter breakdown voltage can be used for this purpose, eliminating the need for the diode.

Additional information about the d-c and a-c parameters of the transistor is given in Fig. 2. This figure shows the collector V–I plot of a p-n-p unit. Also plotted are contours of constant cut-off frequency and constant collector capacitance. It will be noted that the contours of constant frequency response take on roughly the form of hyperbolas with frequency response decreasing as either the low current or the low voltage regions are approached. Collector capacitance decreases as the collector voltage becomes higher. As with other transistors, hole storage delays the response when the transistor is driven out of the collector saturation region. For fast operation one must either control the degree of saturation or stay out of the saturation region completely. It follows from these considerations that the most desirable operating region for this transistor lies in an area which avoids either low values of collector voltage or very low values of collector current.

The collector V–I plot of an n-p-n drift transistor is shown in Fig. 3. This transistor has electrical characteristics similar to those described for the p-n-p unit.

Fig. 4 shows the grounded-base transient response of a drift transistor at two values of collector voltage. The applied current input of 0 to 5 milliamperes is shown in 4(a); 4(b) shows the output wave form taken with a load resistance of 200 ohms and a supply voltage of $1^1/_2$ volts; and 4(c) is the output wave form into the same load with a collector supply voltage of 10 volts. (The minimum rise time of the oscilloscope was 7 millimicroseconds.) It will be noted that an appreciable improvement in response is obtained by operating at a higher collector voltage in the region of higher frequency response and lower collector ca-

R. A. HENLE is with International Business Machines Corporation, Poughkeepsie, New York.

The work reported here represents the combined efforts of a number of people at the International Business Machines Corporation Research Center.

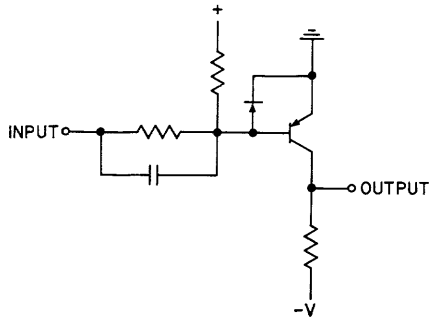


Fig. 1. Inverter designed for minimum turn on delay

pacitance. Only a very slight improvement in response has been noted in operating the transistor entirely class *A*, avoiding the very low collector current region.

Circuit Design

The need for very fast operation of circuits makes it imperative that careful consideration be given to all aspects of circuit design. One consideration is the signal voltage swing at the output of the circuit. As the speed of switching circuits becomes higher, the power consumption of the circuits must go up according to the equation:

$$\text{power consumption} = C\ V^2\ F$$

where C is the shunt load capacitance, V the signal voltage swing, and F the pulse repetition frequency. For high-speed circuits, then, the voltage swing should be as small as is consistent with reliable operation. It is apparent also that every attempt must be made to keep capacities low.

Since small voltage swings are desirable for high-speed operation, careful consideration must be given to the method of performing logic. Diodes, because of their small size and fast operation, warrant consideration as logical elements. Several problems are encountered, however, in their use in high-speed circuits. At low signal levels the voltage drop across the diode becomes an appreciable percentage of the voltage swing. For example, if one has diodes with a 0.3 volt forward drop and attempts to use voltage swings of 1.5 volts, the resulting level shifts of the signal through the diode represents 20 per cent of the usable signal swing. Also, relatively large currents must be used to switch diode logical circuits at millimicrosecond rates. This limits the number of logical stages that may be driven by one transistor.

Both of these disadvantages of diodes can be overcome by using transistors as logical elements. Transistors will switch completely on small signal swings, and the power gain of the transistor can be used to increase the parallel cascading factor of each stage.

The characteristics of drift transistors are such that the most favorable operation is found in a higher-voltage, higher-current region. It follows from this that the circuits will be nonsaturating. It has also been shown that the circuits must be capable of operating at a low signal voltage. An obvious requirement is that the circuits control the operating region of the transistor to keep it within its voltage and current rating.

To meet these requirements a philosophy of circuit design based on controlled current switching was investigated. In the past, most switching circuits have placed close tolerances on the voltage swings. This has proved an acceptable mode of design with both vacuum tubes and transistors. However, a current switch has some desirable features. Fig. 5 illustrates a current driver connected to a load, R_L, through a series inductance, a noise generator, and a resistance. The value of the current received at the load is not affected by these series elements. The voltage developed at the load is that required to pass the current I, while the voltage developed at the current generator may be many times higher. From this line of reasoning,

Table I. Drift Transistor Characteristics

f_{co}	200 → 500 megacycles
α_{cb}	20 → 80
I_{co}	2 → 10 microampere
R_{sat}	50 → 150 ohms
C_c	1 → 8 micromicrofarads
V_{eb}, max	> 50 volts
V_{eb}, min	> 0.5 volts
r_b'	< 50 ohms
Transit time delay	< 5 millimicroseconds

it became evident that one could develop a set of switching circuits by using a transistor as a current generator driving other transistors. This system would be tolerant of noise and would require very low voltage swings. It is plain that shunt elements would deteriorate the operation of these circuits, and care would have to be taken to keep them to a minimum.

Circuits which satisfy the necessary conditions may be either a-c or d-c coupled. Both modes of operation have advantages and disadvantages, and it is difficult to prove conclusively that one is superior. Indeed, it could well be that the optimum system is a combination which incorporates the better features of each. In general, we have tended toward the d-c philosophy because of ease of servicing, avoidance of pulse transformers, and less critical timing relationships.

Circuits

A number of circuit schemes have been investigated which satisfy the requirements of operating on small signal swings, keeping the transistor out of saturation, allowing operation of the transistor in its most favorable operating region, and operation as a current switch. Fig. 6 shows a way in which transistors may be coupled together, by means of alternate n-p-n and p-n-p stages, without the use of interstage coupling networks. The voltage swing at the collector of the p-n-p stage need only be from $V+$

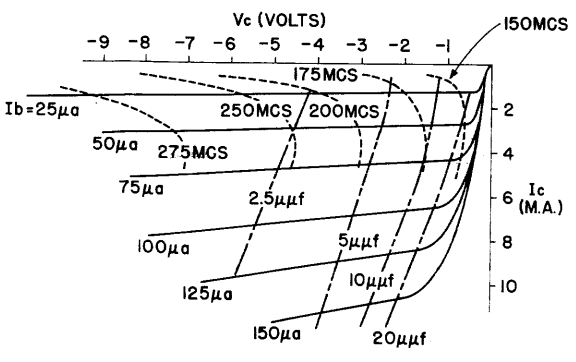
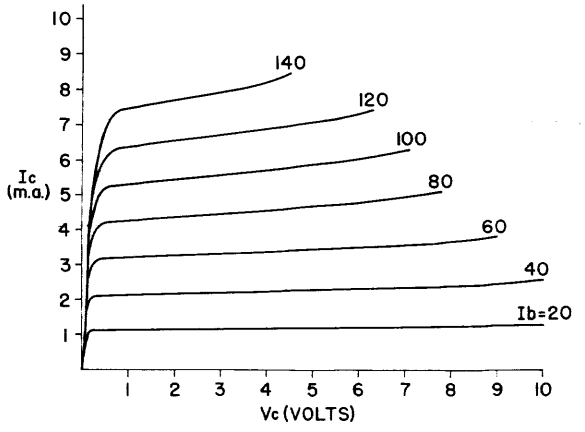


Fig. 2 (left). P-n-p collector V_c–I_c characteristic

Fig. 3 (right). N-p-n collector V_c–I_c characteristic



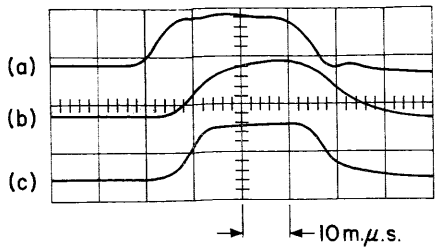


Fig. 4. Drift-transistor transient response

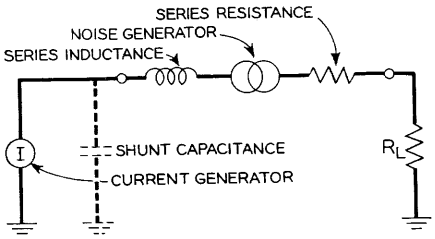


Fig. 5. Current generator driving a load

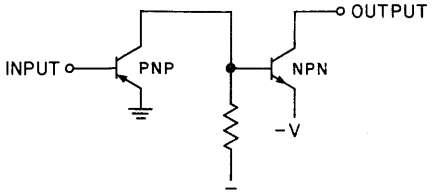


Fig. 6. Direct coupled p-n-p and n-p-n transistors

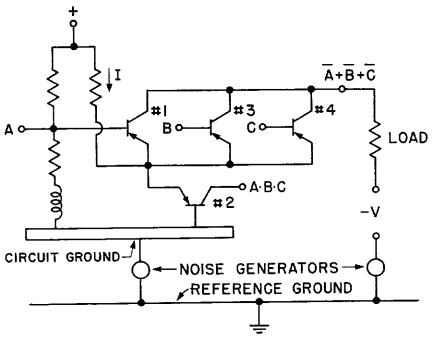


Fig. 7. Current-switching circuit

0.3 to $V-0.3$ volt to completely switch the n-p-n transistor. To limit power dissipation in the p-n-p transistor, it is necessary in some way to limit the collector current. A scheme for accomplishing this is shown in Fig. 7. Here the emitter current of transistors no. 1, no. 2, no. 3, and no. 4 is limited to I. If the inputs to A, B, and C are at an up level, transistor no. 2 conducts all of the current, I, and gives an output which is logically $A \cdot B \cdot C$. If one or more inputs to either A or B or C are at their down level, these transistors will conduct the current I and give an output which is logically $\bar{A}+\bar{B}+\bar{C}$.

It will be noted that $\bar{A}+\bar{B}+\bar{C}$ is the complement of $A \cdot B \cdot C$ and, in general, circuits of this type will generate both a given logical function and its complement. When a number of circuits are combined, logic is done on two voltage levels depending on whether an n-p-n or a p-n-p stage is being driven. The output of an n-p-n is used to drive p-n-p inputs and vice versa. Since any logical expression can be constructed with either type of input, complete freedom in constructing logical expressions is retained. This method of coupling allows us to choose the operating voltage of the collector on the basis of speed and power considerations. The circuits are nonsaturating and the voltage swing need be no greater than that required to switch the transistor.

Fig. 7 is shown with several noise generators introduced into the circuits. With an input circuit such as is shown at input A to transistor no. 1 and the driving signal supplied from a current source, noise generators in the circuit do not affect either the bias on transistor no. 1 or the current supplied to the load.

Conclusions

Circuits of this type are one way of satisfying most of the requirements which were found desirable for high-speed switching. Results with circuits similar to these, using commercially available high frequency transistors, indicate that signal delays average about 16 millimicroseconds per logical stage. Loaded rise times are of the order of 20 millimicroseconds. Delays are cumulative when going through successive logical stages; rise times are not.

High-speed logical circuits, using drift transistors and current-switching techniques, show a great deal of promise. Drift transistors appear capable of operating reliably at speeds well above what was previously possible with economical vacuum tube circuitry.

Reference

1. Proceedings of the Eastern Joint Computer Conference. *AIEE Special Publication T-92*. "Design Objectives of the IBM Stretch Computer," S. W. Dunwell. April 1957, pp. 20–22.

Discussion

M. Cooper (Motoroia, Inc.): Are the transistors described in your talk commercially available, and if so, from whom?

Mr. Henle: These are not commercially available transistors. The transistors were made by the Component Research Group in the Poughkeepsie Laboratories.

N. Prywes (Remington Rand Univac): On which scope, and at what rise time was Fig. 4 taken?

Mr. Henle: Fig. 4 was taken from a Tektronix 517 oscilloscope. Feeding a signal directly into the plates of this oscilloscope, rise times have been measured, of unloaded circuits, of 2 to 3 millimicroseconds.

Mr. Chang (Sylvania Electric Products, Inc.): Is the transistor a p-n-i-p structure or n-p-i-n structure?

Mr. Henle: The transistor is a graded-base transistor with a drift field in the base and a very thin base region, both of which improve the transistor's transient response.

T. R. Finch (Bell Laboratories Incorporated): Is the design philosophy proposed weighted heavily by speed consideratins, reliability, or economy as compared to Dr. Angell's direct-coupled transistor logic proposal, as outlined in his paper.

Mr. Henle: We feel that this technique has much of the simplicity of the direct-coupled transistor logic circuit approach, and probably represents a faster way of using the transistor because it is non-saturating and it does keep the transistor out of its most unfavorable operating regions.