The Lincoln TX-2 Computer Development*

WESLEY A. CLARK†

Introduction

THE TX-2 is the newest member of a growing family of experimental computers designed and constructed at the Lincoln Laboratory of M.I.T. as part of the Lincoln program for the study and development of large-scale, digital computer systems suitable for control in real time. Although, in general characteristics and design philosophy, it owes a great deal to its predecessors, Whirlwind I and the Memory Test Computer, the Lincoln TX-2 incorporates several new developments in components and circuits, memories, and logical organization. It is the purpose of this paper to summarize these new features and to give some idea of the historical development and general design objectives of the TX-2 program. Fig. 1 shows TX-2 in its present development stage.

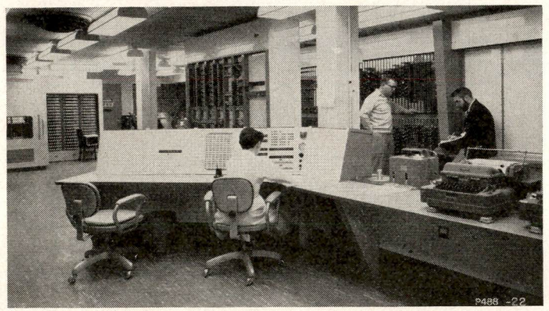


Fig. 1—The Lincoln TX-0 and TX-2 computers. Foreground: TX-0 console; middle center: TX-0 central computer frame; right rear: partially completed TX-2 frame showing plug-in unit construction; left rear: the 256×256 memory.

History

With the development by Lincoln and IBM engineers of the SAGE computer for air defense, real-time control computer systems had reached an impressive level of size, sophistication, and complexity. The highly successful 64×64 coincident-current, magnetic-core, memory array was in operation in the Memory Test Computer which had given up its earlier 32×32 array to Whirlwind. Vacuum tubes abounded in all directions. It was apparent that the further advances in system design which could be made by increasing memory size, eliminating vacuum tubes wherever possible, and organizing input-output buffering, control, and communications into more efficient forms, would be well worthwhile.

* The research reported in this document was supported jointly by the Army, Navy, and Air Force under contract with Mass. Inst Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

The development of a 256×256, switch-driven, magnetic-core memory array was begun and the Philco surface-barrier transistor made its appearance. After some very promising bench experiments with flip-flops and logic circuits, it became apparent that this transistor was potentially well-suited to use in large-scale systems and warranted further study. Accordingly, plans were laid for a succession of experimental digital systems of increasing size and complexity which would make possible the development and evaluation of circuits using the surface-barrier transistors, and which would lead to a computer of advanced design that would be capable of making efficient use of the 256×256 memory.

A double-rank shift register of eight stages and containing about 100 transistors was constructed and put on life-test in April, 1955. It has since been circulating a fixed pattern almost continuously with no known errors and no natural transistor failures.

As the next step, it was decided to build a small, high-speed, error-detecting multiplier and incorporate marginal checking and other system features. The value of a multiplier as a preliminary model had been well demonstrated by the 5-digit system built during Whirlwind's early development. The shift, carry, count, and complement operations, under closely controlled timing conditions, were felt to be representative of all of the operations in the manipulative elements of the type of computer planned. Accordingly, an 8-bit system using 600 transistors was designed and completed in August, 1955 and has been in nearly continuous operation since. Operating margins are periodically checked, and in steady state operation, the multiplier's error-rate has been about one every two months, or one error per 5×10^{11} multiplications at 10^5 multiplications per second. Most of these errors appear to have been caused by cracks in the printed wiring which open intermittently.

During this period, a better idea of the general characteristics of the projected computer began to develop and the engineers who were designing the 256×256 memory were encouraged to think in terms of a word of 36 bits. The notion of a logically separate input-output processor was examined and rejected in favor of a minimum buffering scheme in which data is transferred directly to and from the central memory of the computer. The possibility was recognized of programming these transfers by means of additional program sequences and associated program counters, thus taking advantage of the extensive facilities of the central machine itself for processing input-output data.

It was realized that another development step was desirable before attempting such an elaborate 36-bit system. The 8-bit multiplier had produced a certain

measure of confidence and familiarity with circuits, packaging, and techniques of logical design, but there remained the problems associated with communicating with memory units and input-output equipment operating at vacuum-tube levels over relatively large distances from a central machine which operated at transistor levels. It appeared that the memory development, which had now entered the construction phase, would also benefit by a preliminary evaluation of the 256×256 array and its switching, timing, and noise problems in an operating computer of some kind, possibly with a reduced word length. It was, therefore, decided to design and build next a simple machine—in fact, the simplest reasonable machine—in order to bring about an early intermediate closure of the various efforts within the program.

After some thought about the various possible minimal machines, a design was completed in which the word length would be 18 bits—a graceful half of the projected final form. We began to refer to this computer as the TX-0 and to the projected machine as the TX-2. Because the 256×256 memory array required 16 bits for complete addressing, the single-address instruction word of the TX-0 was left with 2 bits in which to encode instructions. The particular set of instructions chosen included three which required a memory address (*add*, *store*, and *conditional jump*) and one which did not. In this last instruction, the remaining 16 bits were used to control certain necessary and useful primitive operations such as clearing and complementing the accumulator, transferring words between registers, and turning on and off input-output equipment.

The TX-0, equipped with a Flexowriter, a paper-tape reader, and a cathode-ray tube display system was completed, except for the memory, in April, 1956. Twenty planes of the 256×256 memory array were installed the following August and the TX-0, now containing about 3600 transistors and 400 vacuum tubes, began to function as a complete computer. Since that time, it has been used to run a variety of testing and demonstration programs, and a symbolic address compiler and other utility programs have been constructed and are currently in use.

Not only has the TX-0 served the evaluational purposes for which it was built, but it has also demonstrated an effectiveness as a usable computer that is somewhat surprising in view of its simplicity. Its relatively high speed of about 80,000 instructions per second and its 65,536-word memory compensate in large measure for the limitations of its instruction code and logical structure.

With the successful completion of the TX-0, the final steps in the development were undertaken in packaging, circuit refinement, and logical design of the TX-2. A great deal had been learned about the performance of the transistors and memory, the types of logical circuits which are practical, techniques of marginal checking, and the lesser system problems such as color scheme selection and the proper location of pencil sharpeners. As design work progressed, the TX-2 took form as a system of about 22,000 transistors and 600 vacuum tubes. It is an interesting fact that at each step of the development since the shift register, the number of transistors involved was about 6 times the number in the preceding step. This is graphically shown in Fig. 2. At the time of writing, approximately 16 million transistor-hours have accumulated in the shift register, multiplier, and TX-0. There have been two natural deaths and a dozen or so violent ones, primarily due to contact shorting with clip leads and probes.

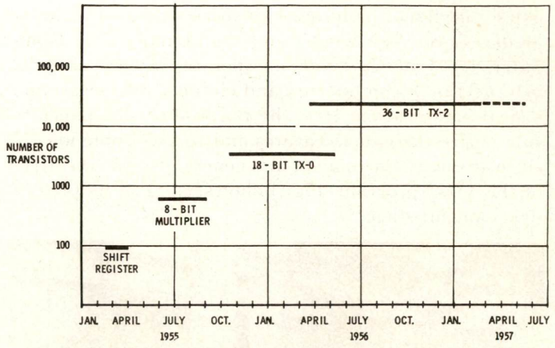


Fig. 2—Steps in the Lincoln TX-2 development program.

Design Objectives

In describing design objectives, it should be pointed out that speed of operation was *not* the primary consideration to which all other attributes were sacrificed. It would have been possible, at the expense of a few more logic circuits, to increase the speed of multiplication, division, and shift-type operations. Similarly, the operation of the index register system could have been made more efficient at the cost of an additional small, fast memory. The principal objective was rather that of achieving a balance among the factors of speed, reliability, simplicity, flexibility, and general virtue.

A key aspect is that of *expandability* which, in an experimental computer in an active environment, certainly ranks with the foregoing qualities in importance. The address structure in the TX-2 permits an expansion of the memory by about a factor of 4, partly to allow for new memory developments, such as the transistor-driven 64×64 array which was begun following the completion of TX-0. New instructions and pieces of terminal equipment will certainly be added during the course of future operation. Extra space and spare plugs have been artfully distributed about in constructing the computer frame. Finally, modular construction will permit a fairly easy physical expansion when required.

The result of all this activity has been a computer of relatively large capability. In addition to incorporating high-speed transistor circuits and a large magnetic-core memory array, the Lincoln TX-2 has two major and distinguishing design characteristics:

1) The structure of the arithmetic element can be altered under program control. Each instruction specifies a particular form of machine in which to operate, ranging from a full 36-bit computer to four 9-bit computers with many variations. Not only is such a scheme able to make more efficient use of the memory in storing data of various word lengths, but it also can be expected to result in greater over-all machine speed because of the increased parallelism of operation.

Peak operating rates must then be referred to particular configurations. For addition and multiplication, these peak rates are given in Table I.

2) Instead of one instruction counter, the TX-2 has 32 such counters which are assigned separately to different users of the computer, who then compete for operating time from instruction to instruction. A special part of the machine selects a particular user based partly on a predetermined priority schedule and partly on the current needs of that user. This multiple-sequence operation, in which many essentially independent instruction sequences interrupt and interleave one another, is an extension of the breakpoint operation found in DYSEAC of the National Bureau of Standards.

TABLE I

Peak Operating Speeds of TX-2

Word Lengths (in bits)	Additions per second	Multiplications per second
36	150,000	80,000
18	300,000	240,000
9	600,000	600,000

The value of these features will have to be assessed during the course of future machine operation. The features themselves are discussed in more detail in the next two papers.

Discussion

P. C. Miller (Logistics Research): When several programs are being run simultaneously, are there any provisions to prevent each of the users from stealing another's storage space?

Mr. Clark: This is a provision which, of course, has to be made by the programmer himself. Actually, suppose we were talking about the multisequence type machine that will be discussed in the next paper, I might mention that there are certain things that we can't acquire in the control of the programmer, if the programmer does try to use the same area of storage that another programmer is using, and there is no way at all that we are going to try to attempt to detect this.

F. S. Preston (Norden Labs.): What uses are planned for the TX-2?

Nelson Blachman (Sylvania EDI) and **Howard Bedford** (North American Aviation): Will the TX-2 Computer be used in any manner by the SAGE system?

Mr. Clark: What motivated the computer was an obvious desire to make something better. Is this a direct part of the SAGE development? The answer to that is, "No direct part," This is a development project of experimental systems, one of many development projects at the Lincoln Laboratory, and the Lincoln Laboratory is, of course, very deeply involved in the SAGE program, but no other form of connection exists. This machine is to be used in simulating physical systems.

W. A. Farrand (Autonetics Div. of N.A.A., Bellflower, Calif.): I would like to know what checking methods are used?

Mr. Clark: I am not sure just what checking methods are meant here. So far as the circuits go, we check the memory system by means of a parity loop; this is a very simple check. We expect the memory to be quite reliable in that we do check all of the core memories with the parity; otherwise, there is no checking while the machine is running. We have to check the machine before the machine is actually doing the programming, but this will be gone into later.

L. Kolbo (RAND Corp.): Does this machine make use of extract and deposit operations by use of masks?

Mr. Clark: The three floating functions that I mentioned originally, the and/or instructions, are not masked instructions; the only one which is, is the masked stored instruction, which is not, strictly speaking, a logical instruction but a digit and memory substitution type instruction.

W. Heising (IBM): What is the time to execute typical floating "add" (programmed)?

Mr. Clark: The reason why we did not wire floating-point operations into the machine is because we found that with configuration control it is very easy to program these instructions. Floating addition, for instance, takes 7 to 9 instructions, depending on how many there are going to be of them, to be actually executed in sequence to develop in 10 microseconds per instruction. That is, about 70 microseconds are required to execute an interpretive floating addition operation. Multiplication and division are much shorter: they take 3 or 4 instructions apiece.

R. Frohman (National Cash Register Co.): It appears that the TX-2 was well and thoroughly planned. Could you please indicate about what amount of time was spent in preliminary planning?

Mr. Clark: The preliminary planning was done largely in the previous systems which were developed. The actual manpower which went into the design and building of the computer is roughly broken down: three engineers doing logical design; three engineers doing the memory work; and three engineers designing and building the hardware. This is besides shop facilities, drafting facilities, and a good number of very good technicians. It was approximately nine people for one year.

Reprinted from the Proceedings of the Western Joint Computer Conference
Los Angeles, California, February 1957

PRINTED IN THE U.S.A.

The Lincoln TX-2 Computer Development*

WESLEY A. CLARK†

Introduction

THE TX-2 is the newest member of a growing family of experimental computers designed and constructed at the Lincoln Laboratory of M.I.T. as part of the Lincoln program for the study and development of large-scale, digital computer systems suitable for control in real time. Although, in general characteristics and design philosophy, it owes a great deal to its predecessors, Whirlwind I and the Memory Test Computer, the Lincoln TX-2 incorporates several new developments in components and circuits, memories, and logical organization. It is the purpose of this paper to summarize these new features and to give some idea of the historical development and general design objectives of the TX-2 program. Fig. 1 shows TX-2 in its present development stage.

Fig. 1—The Lincoln TX-0 and TX-2 computers. Foreground: TX-0 console; middle center: TX-0 central computer frame; right rear: partially completed TX-2 frame showing plug-in unit construction; left rear: the 256×256 memory.

History

With the development by Lincoln and IBM engineers of the SAGE computer for air defense, real-time control computer systems had reached an impressive level of size, sophistication, and complexity. The highly successful 64×64 coincident-current, magnetic-core, memory array was in operation in the Memory Test Computer which had given up its earlier 32×32 array to Whirlwind. Vacuum tubes abounded in all directions. It was apparent that the further advances in system design which could be made by increasing memory size, eliminating vacuum tubes wherever possible, and organizing input-output buffering, control, and communications into more efficient forms, would be well worthwhile.

* The research reported in this document was supported jointly by the Army, Navy, and Air Force under contract with Mass. Inst Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

The development of a 256×256, switch-driven, magnetic-core memory array was begun and the Philco surface-barrier transistor made its appearance. After some very promising bench experiments with flip-flops and logic circuits, it became apparent that this transistor was potentially well-suited to use in large-scale systems and warranted further study. Accordingly, plans were laid for a succession of experimental digital systems of increasing size and complexity which would make possible the development and evaluation of circuits using the surface-barrier transistors, and which would lead to a computer of advanced design that would be capable of making efficient use of the 256×256 memory.

A double-rank shift register of eight stages and containing about 100 transistors was constructed and put on life-test in April, 1955. It has since been circulating a fixed pattern almost continuously with no known errors and no natural transistor failures.

As the next step, it was decided to build a small, high-speed, error-detecting multiplier and incorporate marginal checking and other system features. The value of a multiplier as a preliminary model had been well demonstrated by the 5-digit system built during Whirlwind's early development. The shift, carry, count, and complement operations, under closely controlled timing conditions, were felt to be representative of all of the operations in the manipulative elements of the type of computer planned. Accordingly, an 8-bit system using 600 transistors was designed and completed in August, 1955 and has been in nearly continuous operation since. Operating margins are periodically checked, and in steady state operation, the multiplier's error-rate has been about one every two months, or one error per 5×10^{11} multiplications at 10^5 multiplications per second. Most of these errors appear to have been caused by cracks in the printed wiring which open intermittently.

During this period, a better idea of the general characteristics of the projected computer began to develop and the engineers who were designing the 256×256 memory were encouraged to think in terms of a word of 36 bits. The notion of a logically separate input-output processor was examined and rejected in favor of a minimum buffering scheme in which data is transferred directly to and from the central memory of the computer. The possibility was recognized of programming these transfers by means of additional program sequences and associated program counters, thus taking advantage of the extensive facilities of the central machine itself for processing input-output data.

It was realized that another development step was desirable before attempting such an elaborate 36-bit system. The 8-bit multiplier had produced a certain

measure of confidence and familiarity with circuits, packaging, and techniques of logical design, but there remained the problems associated with communicating with memory units and input-output equipment operating at vacuum-tube levels over relatively large distances from a central machine which operated at transistor levels. It appeared that the memory development, which had now entered the construction phase, would also benefit by a preliminary evaluation of the 256×256 array and its switching, timing, and noise problems in an operating computer of some kind, possibly with a reduced word length. It was, therefore, decided to design and build next a simple machine—in fact, the simplest reasonable machine—in order to bring about an early intermediate closure of the various efforts within the program.

After some thought about the various possible minimal machines, a design was completed in which the word length would be 18 bits—a graceful half of the projected final form. We began to refer to this computer as the TX-0 and to the projected machine as the TX-2. Because the 256×256 memory array required 16 bits for complete addressing, the single-address instruction word of the TX-0 was left with 2 bits in which to encode instructions. The particular set of instructions chosen included three which required a memory address (*add*, *store*, and *conditional jump*) and one which did not. In this last instruction, the remaining 16 bits were used to control certain necessary and useful primitive operations such as clearing and complementing the accumulator, transferring words between registers, and turning on and off input-output equipment.

The TX-0, equipped with a Flexowriter, a paper-tape reader, and a cathode-ray tube display system was completed, except for the memory, in April, 1956. Twenty planes of the 256×256 memory array were installed the following August and the TX-0, now containing about 3600 transistors and 400 vacuum tubes, began to function as a complete computer. Since that time, it has been used to run a variety of testing and demonstration programs, and a symbolic address compiler and other utility programs have been constructed and are currently in use.

Not only has the TX-0 served the evaluational purposes for which it was built, but it has also demonstrated an effectiveness as a usable computer that is somewhat surprising in view of its simplicity. Its relatively high speed of about 80,000 instructions per second and its 65,536-word memory compensate in large measure for the limitations of its instruction code and logical structure.

With the successful completion of the TX-0, the final steps in the development were undertaken in packaging, circuit refinement, and logical design of the TX-2. A great deal had been learned about the performance of the transistors and memory, the types of logical circuits which are practical, techniques of marginal checking, and the lesser system problems such as color scheme selection and the proper location of pencil sharpeners. As design work progressed, the TX-2 took form as a system of about 22,000 transistors and 600 vacuum tubes. It is an interesting fact that at each step of the development since the shift register, the number of transistors involved was about 6 times the number in the preceding step. This is graphically shown in Fig. 2. At the time of writing, approximately 16 million transistor-hours have accumulated in the shift register, multiplier, and TX-0. There have been two natural deaths and a dozen or so violent ones, primarily due to contact shorting with clip leads and probes.

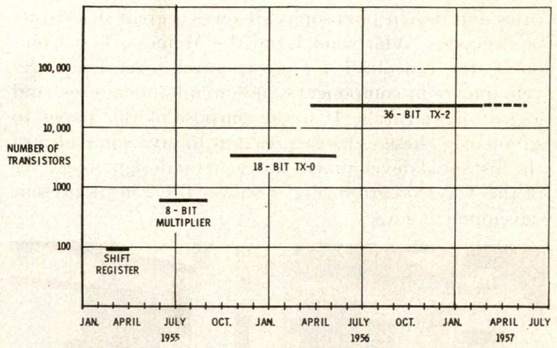


Fig. 2—Steps in the Lincoln TX-2 development program.

Design Objectives

In describing design objectives, it should be pointed out that speed of operation was *not* the primary consideration to which all other attributes were sacrificed. It would have been possible, at the expense of a few more logic circuits, to increase the speed of multiplication, division, and shift-type operations. Similarly, the operation of the index register system could have been made more efficient at the cost of an additional small, fast memory. The principal objective was rather that of achieving a balance among the factors of speed, reliability, simplicity, flexibility, and general virtue.

A key aspect is that of *expandability* which, in an experimental computer in an active environment, certainly ranks with the foregoing qualities in importance. The address structure in the TX-2 permits an expansion of the memory by about a factor of 4, partly to allow for new memory developments, such as the transistor-driven 64×64 array which was begun following the completion of TX-0. New instructions and pieces of terminal equipment will certainly be added during the course of future operation. Extra space and spare plugs have been artfully distributed about in constructing the computer frame. Finally, modular construction will permit a fairly easy physical expansion when required.

The result of all this activity has been a computer of relatively large capability. In addition to incorporating high-speed transistor circuits and a large magnetic-core

memory array, the Lincoln TX-2 has two major and distinguishing design characteristics:

1) The structure of the arithmetic element can be altered under program control. Each instruction specifies a particular form of machine in which to operate, ranging from a full 36-bit computer to four 9-bit computers with many variations. Not only is such a scheme able to make more efficient use of the memory in storing data of various word lengths, but it also can be expected to result in greater over-all machine speed because of the increased parallelism of operation.

Peak operating rates must then be referred to particular configurations. For addition and multiplication, these peak rates are given in Table I.

2) Instead of one instruction counter, the TX-2 has 32 such counters which are assigned separately to different users of the computer, who then compete for operating time from instruction to instruction. A special part of the machine selects a particular user based partly on a predetermined priority schedule and partly on the current needs of that user. This multiple-sequence operation, in which many essentially independent instruction sequences interrupt and interleave one another, is an extension of the breakpoint operation found in DYSEAC of the National Bureau of Standards.

TABLE I
Peak Operating Speeds of TX-2

Word Lengths (in bits)	Additions per second	Multiplications per second
36	150,000	80,000
18	300,000	240,000
9	600,000	600,000

The value of these features will have to be assessed during the course of future machine operation. The features themselves are discussed in more detail in the next two papers.

Discussion

P. C. Miller (Logistics Research): When several programs are being run simultaneously, are there any provisions to prevent each of the users from stealing another's storage space?

Mr. Clark: This is a provision which, of course, has to be made by the programmer himself. Actually, suppose we were talking about the multisequence type machine that will be discussed in the next paper, I might mention that there are certain things that we can't acquire in the control of the programmer, if the programmer does try to use the same area of storage that another programmer is using, and there is no way at all that we are going to try to attempt to detect this.

F. S. Preston (Norden Labs.): What uses are planned for the TX-2?

Nelson Blachman (Sylvania EDI) and **Howard Bedford** (North American Aviation): Will the TX-2 Computer be used in any manner by the SAGE system?

Mr. Clark: What motivated the computer was an obvious desire to make something better. Is this a direct part of the SAGE development? The answer to that is, "No direct part," This is a development project of experimental systems, one of many development projects at the Lincoln Laboratory, and the Lincoln Laboratory is, of course, very deeply involved in the SAGE program, but no other form of connection exists. This machine is to be used in simulating physical systems.

W. A. Farrand (Autonetics Div. of N.A.A., Bellflower, Calif.): I would like to know what checking methods are used?

Mr. Clark: I am not sure just what checking methods are meant here. So far as the circuits go, we check the memory system by means of a parity loop; this is a very simple check. We expect the memory to be quite reliable in that we do check all of the core memories with the parity; otherwise, there is no checking while the machine is running. We have to check the machine before the machine is actually doing the programming, but this will be gone into later.

L. Kolbo (RAND Corp.): Does this machine make use of extract and deposit operations by use of masks?

Mr. Clark: The three floating functions that I mentioned originally, the and/or instructions, are not masked instructions; the only one which is, is the masked stored instruction, which is not, strictly speaking, a logical instruction but a digit and memory substitution type instruction.

W. Heising (IBM): What is the time to execute typical floating "add" (programmed)?

Mr. Clark: The reason why we did not wire floating-point operations into the machine is because we found that with configuration control it is very easy to program these instructions. Floating addition, for instance, takes 7 to 9 instructions, depending on how many there are going to be of them, to be actually executed in sequence to develop in 10 microseconds per instruction. That is, about 70 microseconds are required to execute an interpretive floating addition operation. Multiplication and division are much shorter: they take 3 or 4 instructions apiece.

R. Frohman (National Cash Register Co.): It appears that the TX-2 was well and thoroughly planned. Could you please indicate about what amount of time was spent in preliminary planning?

Mr. Clark: The preliminary planning was done largely in the previous systems which were developed. The actual manpower which went into the design and building of the computer is roughly broken down: three engineers doing logical design; three engineers doing the memory work; and three engineers designing and building the hardware. This is besides shop facilities, drafting facilities, and a good number of very good technicians. It was approximately nine people for one year.

Reprinted from the Proceedings of the Western Joint Computer Conference
Los Angeles, California, February 1957

PRINTED IN THE U.S.A.

A FUNCTIONAL DESCRIPTION OF THE LINCOLN TX-2 COMPUTER

BY

J. M. FRANKOVICH AND H. P. PETERSON

Reprinted from the PROCEEDINGS OF THE WESTERN JOINT COMPUTER CONFERENCE
Los Angeles, California, February 1957

PRINTED IN THE U.S.A.

A Functional Description of the Lincoln TX-2 Computer*

J. M. FRANKOVICH† AND H. P. PETERSON†

Introduction

THE TX-2 is a large scale digital computer designed and built at the Massachusetts Institute of Technology Lincoln Laboratory utilizing new memory and circuit components and some new logical design concepts. The computer will be applied as a research tool in scientific computations, and in data-handling and real-time problems. The design of the computer reflects not only the characteristics of the components available, but also the nature of the intended applications. This paper explains the functional and organizational aspects of the computer which are important from the user's point of view.

General Structure of TX-2

TX-2 is a parallel binary computer with a 36-digit word length. The internal memory is all random-access and will initially consist of 69,632 registers of parity checked magnetic-core memory and about 24 additional toggle switch and flip-flop registers. About 150,000 instructions can be executed per second. Instructions are of the indexed single-address type, and a fixed-point, signed-fraction, one's complement number system is used.

Several unusual ideas incorporated in the system organization reduce the amount of information unnecessarily manipulated during program sequences. Furthermore, the system organization facilitates the execution of several operations simultaneously, thereby increasing the effective speed of the computer.

The principal registers and information paths in the computer are illustrated schematically in Fig. 1. *A*, *B*, *C*, *D*, *E*, *F*, *M*, and *N* are the 36-bit flip-flop registers in the machine. *M* and *N* are memory buffer registers, each of which has a parity flip-flop and associated circuitry used to check the parity of memory words. *P*, *Q*, and *X* are 18-digit registers; *X* also has a parity digit which is used to check the parity of words in the *X* memory. Control flip-flops are not shown in Fig. 1.

Instructions are full memory words and are placed in the Control Element during the instruction memory cycle. During the operand memory cycle, an operand is usually transmitted between the Memory Element and some other element—always through the Exchange

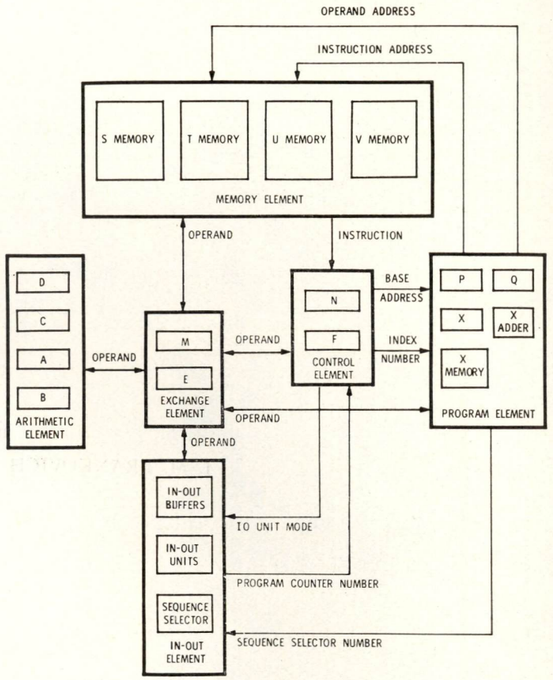


Fig. 1—TX-2 system schematic, showing the principal registers and transfer paths.

Element. The 36-digit configuration of the memory is not, however, maintained throughout the computer during operation timing. A programmer can, in effect, control several independent, shorter operand word length computers simultaneously during the execution of each instruction. This flexibility is realized by specifying a particular system *configuration* with each instruction.

The computer communicates with the outside world through units in the In-Out Element, several of which can be simultaneously operated. Whenever an input or output information transfer can occur, signals to the Program Element from the In-Out Element automatically call into operation the associated instruction sequence. This *multiple-sequencing* aspect of the computer will not be described in this paper.[1]

* This research was supported jointly by the Army, Navy, and Air Force under contract with the Mass. Inst. Tech., Cambridge, Mass.

† M.I.T. Lincoln Lab., Lexington, Mass.

[1] J. W. Forgie, "The Lincoln TX-2 in-out system," this issue· p. 156.

Memory Element

The availability of a large, fast, core memory for TX-2 permitted an emphasis on the design of a machine with an all random-access memory which could be as large as 262,144 words. The homogeneous aspect of so large a memory system simplifies the programmer's coding problems and permits continued high-speed operation regardless of the program location in the internal memory.

The TX-2 Memory Element (see Fig. 2) is divided into four independently operating memories, each containing up to 65,536 36-digit words. The operating speed of TX-2 is determined by the cycle time for the memories: the 65,536-word *S Memory* is expected to have a cycle time of between six and seven microseconds; and the 4096-word *T Memory*, a cycle time between five and six microseconds. Both memories are parity checked.

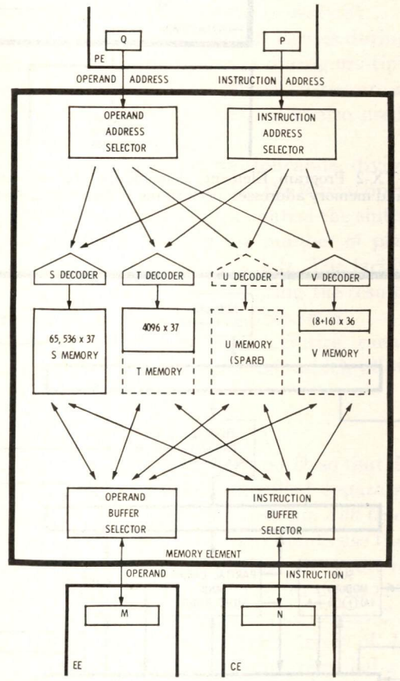


Fig. 2—TX-2 Memory Element. Two address and two buffer registers are used to permit simultaneous operation of any two of the four memories.

Although the *U Memory* currently is not specified, it may contain a 4096-word core memory in the initial system. The *V Memory* consists of 8 flip-flop registers in the central machine and 16 toggle switch registers which contain the program sequence executed whenever the START button on the operator's console is pushed. The contents of the toggle switch registers can be used as instructions or operands, but naturally cannot be altered by a program. The six 36-bit registers *A*, *B*, *C*, *D*, *E*, and *F* are also part of the *V* Memory but their contents can be used only as operands during the execution of an instruction. The programmer has, in a limited sense, a two address instruction machine when he refers to these registers in load and store type instructions. The other two flip-flop registers in the *V* Memory are a 60-counts-per-second clock and a random-number register.

When an instruction calls for the storing of an operand in memory, the operand memory cycle can be extended up to two microseconds. The extension occurs between the time that the memory register is read and the time that it is rewritten. During this extension time the memory register transfers in the central computer take place, the parity of the word read from memory is checked, and the parity of the new memory word computed. Because the extended cycle is less than the two complete cycles traditionally used for word-modifying instructions, an increase in computing efficiency is realized.

The *P* Register in the Program Element specifies the location of an instruction in memory and the *N* Register in the Control Element holds the instruction after it has been read from memory. The two leftmost digits of *P* select the memory system from which the instruction word is to be obtained; the right 16 digits address the word within the memory. Similarly, the *Q* Register locates the operand in one of the memory systems, the operand being placed in *M*.

Control and Indexing

An instruction word read into *N* has the structure shown in Fig. 3. The first two digits of the word specify information to the In-Out Element, and the four *cf* digits specify the computer configuration. The interpretation of the *b* and *d* digits is not discussed here.[2] The *cf* digits will be discussed later.

The operation code for the instruction is specified by

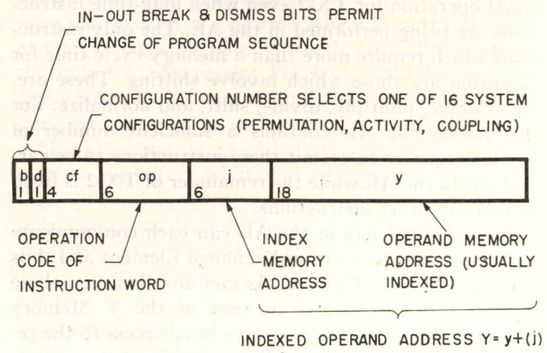


Fig. 3—TX-2 instruction word layout.

[2] *Ibid.*

the six *op* digits. On simple load and store type instructions these six digits are further subdivided into two groups of three. The first group determines the operation and the second specifies the register in the central computer which is being loaded or whose contents are being stored.

The base address for the operand, formed by the 18 y digits, is usually modified by the contents of the index register selected by the six j digits. The index registers form a unique 64-register, parity-checked core memory which has a 1 microsecond access time. The contents of the specified index register is read into the X Register of the Program Element via the paths indicated in Fig. 4. The base address and the index are fed into a full adder circuit which produces the sum, $Y=y+(j)$, in about 1 microsecond. The over-all complexity of the Program Element was reduced by having the adder produce both the sum, Y, and the unmodified base address, y; either of these quantities can be directed to the operand memory address register Q. Whenever the zeroth index register is chosen, the adder produces only the unmodified base address. The effect is the same as having the index register contain zero, so the programmer can avoid index modification altogether.

The instruction memory address register P normally is indexed by one as each instruction is executed, but jump instructions may cause the output of the index adder to be directed to P. The adder also provides a communication path for index jump instructions from the X Memory to the Memory Element by way of the Exchange Element.

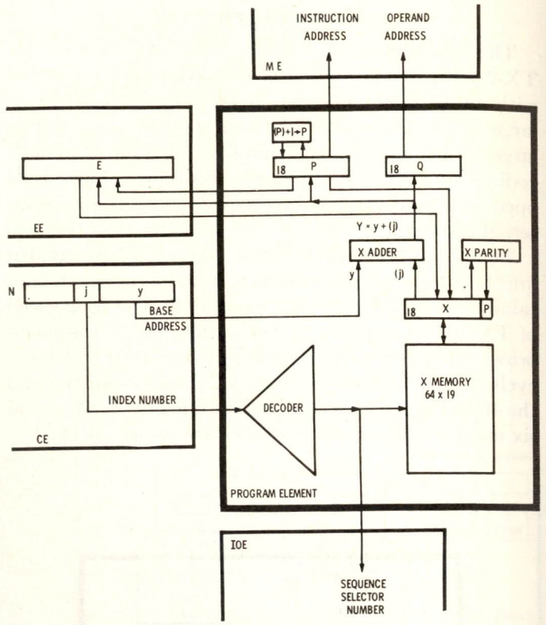


Fig. 4—TX-2 Program Element, determining the instruction and operand memory addresses, performing X memory operations.

Arithmetic Element

The registers and sufficient basic operations in the Arithmetic Element (AE) to implement addition, multiplication, division, shift, and various logical operations are shown in Fig. 5. Operation timing for most of the TX-2 instructions is also performed in the AE.

The design of the AE reflects the desire to attain high-speed operation for TX-2 even when long-time instructions are being performed in the AE. The only instructions which require more than a memory-cycle time for execution are those which involve shifting. These are, for example, multiply, divide, shift, and normalize. For this reason the AE contains a sufficient number of storage registers to permit these instructions to be carried out in the AE while the remainder of TX-2 is freed to perform other instructions.

The four registers in the AE can each communicate with the E Register in the Exchange Element and thus with the Memory Element. As mentioned earlier, these registers are addressable as part of the V Memory System. Therefore, programmers have access to the results in any register of an AE computation.

The AE registers, designated by A, B, C, and D, are described on the next page.

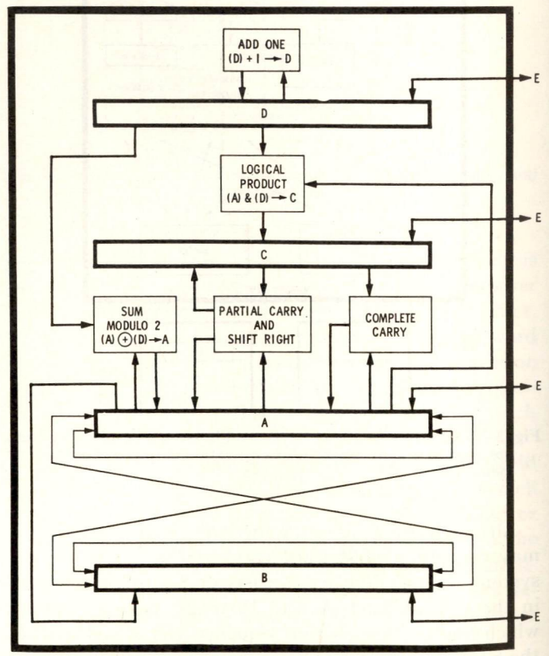


Fig. 5—TX-2 Arithmetic Element, showing the circuits and transfer paths for AE operations.

The A Register accumulates the results of all the arithmetic operations except division for which it holds the remainder. It holds one of the operands and accumulates the results of the three logical operations (AND, INCLUSIVE OR, EXCLUSIVE OR) which, it should be noted, are bit-wise operations. The information in the A Register can also be shifted (*i.e.*, multiplied by some positive or negative power of two) or cycled (*i.e.*, shifted, without preserving the special significance of the sign bit, as in a closed ring).

The B Register serves as an extension of A during multiplication, certain shifts and cycles, and, in a sense, during division when the least significant digits of the double-length dividend are stored in B. The resulting quotient then appears in B. Moreover, the information in B can be shifted or cycled independently of A. In multiplication, the multiplier originally in A is transferred via parallel paths directly into B (where the least significant digit then controls the operation).

The C Register stores the partial carries during arithmetic operations, most important during multiplication as described later. Since these partial carries are actually bit-wise logical products (AND), C is also used to accumulate logical products.

The D Register holds the multiplicands, divisors, addends, and one of the operands for the logical operations. It also holds the numbers which control the shifting and cycling of A and B, namely the number of places, up to 72, and the direction, right or left. The facility of D to count is used also in accumulating the results of the normalizing of A and counting ones in A.

Besides the above mentioned facilities, each of the AE registers can be complemented, which allows subtractions to be done.

AE Circuits

There are four *Add One* circuits on D, so that different parts of A and B can be controlled separately and simultaneously. For simplicity, just one *Add One* circuit is shown in Fig. 5. These *Add One* circuits use the simultaneous carry principle, permitting one count to occur every 0.4 microsecond. Each can count up to 127.

The *Logical Product* circuit of A and D into C and the *Sum Modulo 2* (EXCLUSIVE OR) circuit of A and D into A when used at the same time are called a *Partial Add*. When the *Complete Carry* circuit is activated after a Partial Add, the result is a full addition of D and A into A. The *Complete Carry* circuit uses the high-speed carry principle and takes about 1.5 microseconds for 36 bits.

The *Partial Carry and Shift Right* circuit is also known as "multiply step" and was, we believe, first used on Whirlwind I. As used in multiplication, this circuit obviates the need for a full addition for each "one" in the multiplier. Carries are propagated only one stage during each step except the last when a complete carry is executed. This iterative process takes about 16 microseconds in the worst case for a full 36-digit multiplication. The iterative process for division, on the other hand, requires a complete addition at each step and consequently takes about 72 microseconds in the worst case.

Two features of the AE control ought to be mentioned here. A 7-bit step counter, like the Add One circuit on D, is used to control multiplication and division and to limit the shifting in normalizing and the cycling in counting "ones." A flip-flop signifying overflow during addition and division is also used to remember the sign of the product during multiplication and the sign of the quotient during division. If a division overflow occurs, the sign is replaced by the overflow state and the quotient is lost.

Control of the Arithmetic Element is independent of the rest of the machine, thus providing the time-saving device of continuing to execute non-AE instructions while AE is performing one of the longer shift operations or a division.

System Timing

In part, the high speed of TX-2 is attained by overlapping the operation of as many components as is logically possible without incorporating large amounts of circuitry. The time-consuming cyclic operations in an indexed single-address computer are the *instruction-memory cycle*, the *index-memory cycle*, the *index-addition time*, the *operand-memory cycle*, and the *operation timing*. These cycles occur in the mentioned sequence during the execution of ordinary instructions.

Several asynchronous "clocks" which use a 5-megacycle pulse source control the various cycles. The instruction and operand memory cycles can be overlapped if they take place in different memory systems.

The overlap of these cycle times for a sequence of load type instructions is illustrated in Fig. 6(a). Here different instruction and operand memories with roughly equal cycle times are assumed. If a sequence of store type instructions is executed which requires extended memory cycles for the operand, then the situation shown in Fig. 6(b) results. Fig. 6(c) shows the time used when both the instruction and the operand are in the same memory.

"Peak" operating speed for the computer is attained only in Fig. 6(a); additional circuitry could improve Fig. 6(b) and Fig. 6(c), but only at considerable cost. It is interesting to note that if the computer is to run at peak speeds, the address of the operand used by the current instruction must be available before the earliest moment at which the next instruction memory cycle could begin. If the total accumulated time from the beginning of an instruction memory cycle until the time that the address of the operand is known is greater than the instruction memory cycle time, then the computer

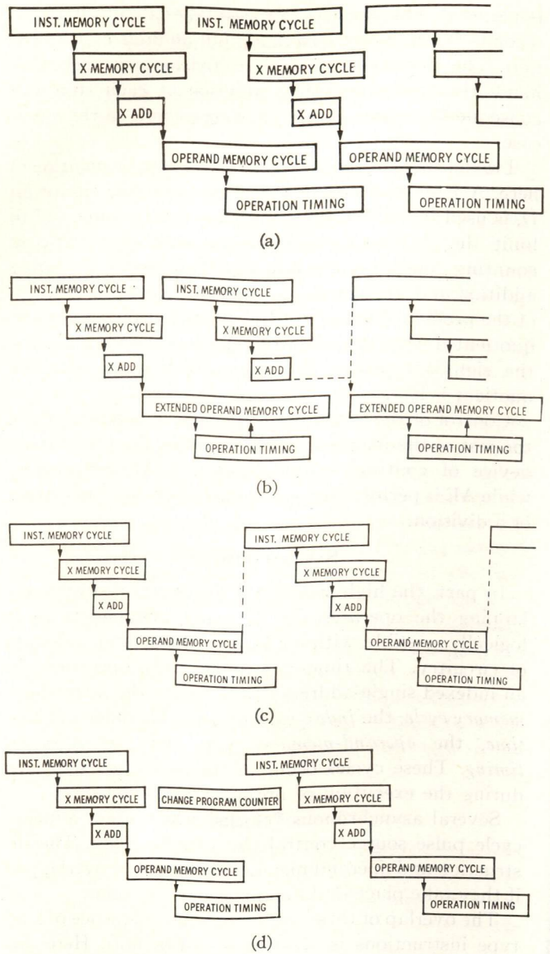


Fig. 6—TX-2 timing schematic, showing overlapped execution of memory and operation cycles. (a) Consecutive load type instructions—instructions and operands in different memories. (b) Consecutive store type instructions. (c) Instruction and operand in same memory. (d) Change sequence.

cannot run in the ideal manner shown in Fig. 6(a). This means that the access time of all memories, and the index add time must be kept as short as possible.

Fig. 6(d) depicts the timing of events when the In-Out Element causes a change in program sequence by changing the contents of the P register. The additional X Memory cycles which must be performed in carrying this out produce a timing situation similar to that of the X Memory load and store type instructions.

The operation timing for an instruction is executed when the operand is available from memory. Only the Arithmetic Element step counter instructions, multiply, divide, shift, etc., require an operating timing cycle longer than a memory cycle. Since only the Arithmetic Element is tied up when these instructions occur, the Control Element permits any non-Arithmetic-Element instruction to be executed while the AE is busy. Division takes up to 75 microseconds, so the programmer can write as many as 14 non-AE instructions following a divide, all of which can be executed before the division is completed.

Configuration

The design of a general purpose computer must necessarily reflect the contradictory demands for both short and long word lengths, floating and fixed point arithmetic operations, and a multitude of logical and decision instructions. The computer should be able to process information at an optimum rate in a variety of problems without the need for intricately coded programs. This ability should be achieved without excessively complex and costly circuitry.

The full 36-digit word in TX-2 represents a reasonable length for operands in some numerical computations, notably scientific and engineering computations. Though floating point arithmetic operations are not included in the instruction code, both they and multiple-precision operations can be easily synthesized by means of the existing instructions. The logical instructions in the code facilitate operations on individual digits, but also, a configuration which the programmer specifies anew with each instruction permits him to perform arithmetic operations on operands which are less than 36 digits long. When such is the case, several shorter operands can be manipulated simultaneously.

The four *cf* digits in an instruction word (see Fig. 3) are decoded as shown schematically in Fig. 7. The contents of the selected one of 16 9-digit configuration words are placed in a flip-flop register whose output levels determine a static configuration for the entire computer during the execution of the instruction. The contents of the first twelve registers are specified by a notation whose meaning will be clarified in the following discussion.

The full 36-digit word length is always maintained for instruction words, but during operation timing, every 36-digit register in the Memory, Exchange, and Arithmetic Elements is considered broken into four 9-digit *quarters* [numbered from 1 to 4, from right to left as in Fig. 8(a)]. While the instruction is being executed, these quarters are recombined on the basis of the configuration.

Parallel register transfers are the usual means for moving information about in the machine. The *EE permutation* digits select one of the four permutations P0, P1, P2, or P3 as defined in Fig. 8(b). The chosen permutation effects the corresponding cross-communication paths between the quarters of the E and M registers of the Exchange Element. As operands are transmitted through the EE, the quarters of the word follow the set of paths determined by the selected permutation. The result is that the operand is shifted $9n$ places to the left as it moves from M to E or $9n$ places

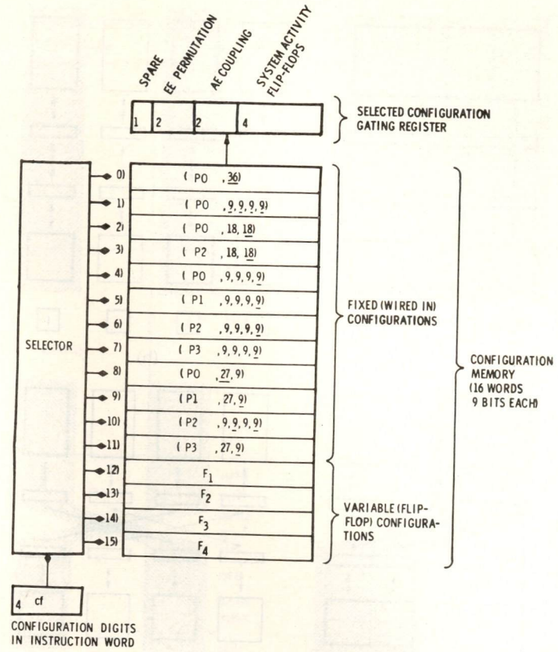


Fig. 7—TX-2 configuration selection. The *cf* digits select a configuration for the computer for use during the execution of the instruction.

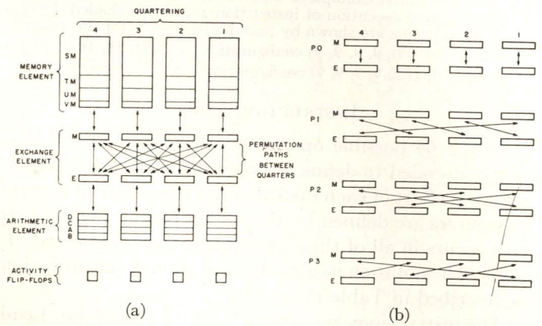


Fig. 8—TX-2 configuration. (a) Quartering, permutation paths, and activity flip-flops. (b) The four sets of permutation paths available, one of which is used during the execution of an instruction.

to the right as it moves from E to M, $n=0$, 1, 2, or 3. Thus the programmer can have any quarter of the AE communicate with any quarter of the ME.

This communication ability is focused more sharply by having the configuration specify a *system activity*. All operation timing events in a given quarter of the AE and EE and the quarter of the ME connected via the selected permutation path in the EE are controlled by the activity flip-flop of that quarter. If the activity flip-flop of a given quarter holds a "one," as specified by the configuration, then the operation timing events of the instruction occur in that quarter. If the activity flip-flop holds a "zero," then nothing happens.

During the execution of arithmetic operations, the *AE coupling* bits further specify the connections of the lateral information paths between quarters in the AE. Information flows laterally only through the shift and the carry circuits, and the connection of these circuits alone determines the word length of the numerical quantities manipulated in the AE.

In Fig. 9(a) (next page) every quarter of the AE has coupling units at each end which receive the shift and carry information entering the quarter. The general type of connections among several quarters is shown in Fig. 9(b). The digit *length* of operands during add and shift operations is determined by the number of quarters coupled together. In TX-2 from one to four quarters can be coupled together to permit arithmetic operations on 9, 18, 27, or 36-digit operands. The various combinations of coupling unit connections actually chosen by the AE coupling are symbolized in Fig. 9(c). Since A-register, B-register, and AB-register shifts are permitted in the Arithmetic Element, the programmer can obtain 18, 36, 54, or 72-digit shifts. All the possible shift (and cycle) configurations are shown in Fig. 9(d).

Only those inputs to the coupling units which would yield useful arithmetic element structures are realized by the AE coupling. It should be emphasized that the programmer can realize several arithmetic elements simultaneously. The coupling (36) gives only one 36-bit AE, but the coupling (18, 18) gives two complete, independent 18-bit arithmetic elements which are separately but simultaneously controlled by the instruction being executed. Two arithmetic elements are again available with the coupling (27, 9), one 27 bits and the other 9 bits long, and the (9, 9, 9, 9) case gives four 9-bit arithmetic elements. The permutation paths in the Exchange Element permit each arithmetic element to communicate with any quarter of a memory word and the activity flip-flops can specify just which of the realized arithmetic elements will actually be active and in active communication with the connected part of memory.

In Fig. 10, several examples are given of the different configurations which can be realized in TX-2. The most straightforward configuration has one 36-digit arithmetic element and communicates directly with memory. The notation (P0, 36) signifies the permutation (no shift) and the form of the arithmetic element (one 36-digit). The underlining indicates that the whole system is active. Slightly more varied is the (P0, 9, 9, 9, 9) configuration which specifies four 9-digit arithmetic elements communicating directly with memory, but with only two of them active. The (P2, 9, 9, 9, 9) configuration has the same arithmetic elements but with the associated memories interchanged. The (P2, 18, 18) configuration illustrates an 18-digit arithmetic element which uses the "other" half of memory.

One of the 9 configuration digits is at the moment unused, but will probably be used to control the ex-

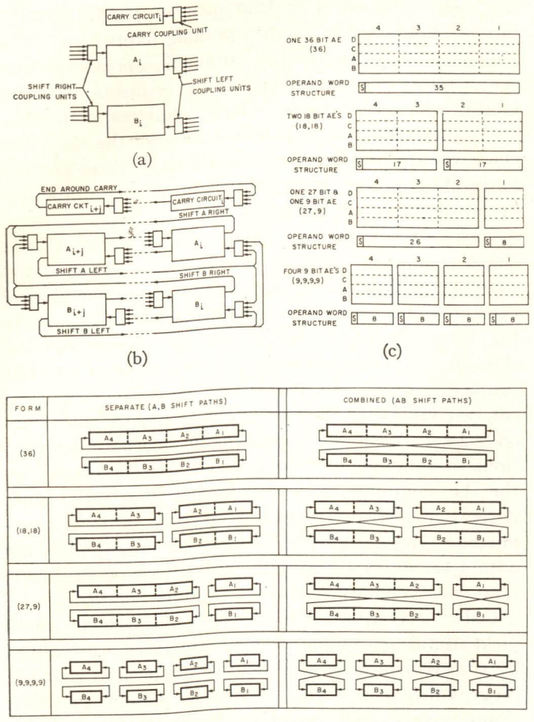


Fig. 9—TX-2 arithmetic element coupling units. (a) i-th quarter coupling units. The coupling units receive information moving laterally into the i-th quarter of the AE, $i=1, 2, 3, 4$. (b) Coupling unit connections between a contiguous group of quarters which realize a 9-bit ($j=0$), 18-bit ($j=1$), 27-bit ($j=2$) or 36-bit ($j=3$) "arithmetic element." (c) Arithmetic element and operand word structures. The four forms the arithmetic element can assume with associated operand word structure. (d) The possible shift path arrangements realized with the configurations.

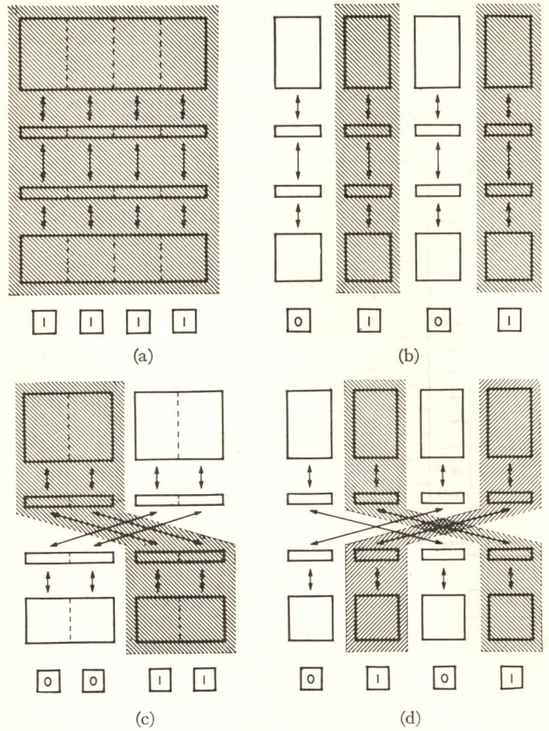


Fig. 10—Illustrative example of different configurations. Areas of activity during execution of instruction are shown shaded. Effects of AE coupling are shown by juxtaposition. (a) (P0, 36) configuration. (b) (P0, 9, 9, 9, 9) configuration. (c) (P2, 18, $\underline{18}$) configuration. (d) (P2. 9, $\underline{9}$, 9, $\underline{9}$) configuration.

tension of the sign of numbers as they pass through the EE on the way from the ME to the AE. The scheme presently under consideration would permit programmers to add, for example, a 9-digit memory operand to an 18-digit arithmetic element. This scheme would permit closer packing of operands in memory and significantly increase the speed of solving some real-time problems, where short data words need to be extended so higher precision can be maintained during computations. Working details of the scheme have yet to be fixed.

The configuration memory from which the programmer chooses a configuration for use with each instruction was shown in Fig. 7. Twelve of the configuration memory registers are fixed circuitry whose contents cannot be changed without changing the wiring of the computer. These configurations are assumed to be ones which will be useful to most programmers. The last four registers in the memory consist of the 36 digits of the F register. As will be seen the programmer can quite simply alter the contents of this register and thereby obtain any of the (less than 2^9) possible configurations.

Instruction Code

Of the 64 possible operation codes, only 51 are currently decoded to define instructions. In Table I (opposite) the effect of each instruction is described. If several computers are defined by the configuration, then the effect occurs in all of them simultaneously and independently. The notation used in the definition of the operation is described in Table II (p. 154).

The instructions are grouped according to type. Load and store type instructions simply effect an operand transfer between the selected register and memory. The load complement instructions are variants which load the one's complement into the specified registers. Exchange simply interchanges the contents of A and the indicated memory register. The insert instruction allows any set of bits in A, as specified by the bits in B, to be stored in memory. In the index memory load and store instructions, the j bits select the index register involved so the operand address is not modified.

All of the add and step-counter instructions can also be classed as load type instructions in so far as the operand memory cycle is concerned. The multiply instruction forms the full product in the A and B registers. Division is the inverse of multiplication, the double

TABLE I

Type	Mnemonic Code	Operation	Name
Load	lda	$(Y)\rightarrow A$	Load into A
	ldb	$(Y)\rightarrow B$	Load into B
	ldc	$(Y)\rightarrow C$	Load into C
	ldd	$(Y)\rightarrow D$	Load into D
	lde	$(Y)\rightarrow E$	Load into E
	ldf	$(Y)\rightarrow F$	Load into F
	lca	$\overline{(Y)}\rightarrow A$	Load complement into A
	lcb	$\overline{(Y)}\rightarrow B$	Load complement into B
	ldx	$(y)\rightarrow j$	Load into index
Store	sta	$(A)\rightarrow Y$	Store A
	stb	$(B)\rightarrow Y$	Store B
	stc	$(C)\rightarrow Y$	Store C
	std	$(D)\rightarrow Y$	Store D
	ste	$(E)\rightarrow Y$	Store E
	stf	$(F)\rightarrow Y$	Store F
	exa	$(Y)\rightarrow A$; $(A)\rightarrow Y$	Exchange A
	ins	(B) & (A) v $\overline{(B)}$ & $(Y)\rightarrow Y$	Insert digits of A
	stx	$(j)\rightarrow y$	Store index
Add	add	$(A) + (Y)\rightarrow A$	Add
	sub	$(A) + \overline{(Y)}\rightarrow A$	Subtract
	dma	$\lvert(A)\rvert + \lvert\overline{(Y)}\rvert\rightarrow A$	Difference of magnitude
	and	(A) & $(Y)\rightarrow A$	Logical and
	ori	(A) v $(Y)\rightarrow A$	Logical or—inclusive
	ore	$(A) \oplus (Y)\rightarrow A$; (A) & (Y) v $(C)\rightarrow C$	Logical or—exclusive (and accumulate product)
	axm	$(j)+(y)\rightarrow y$	Add index to memory
	amx	$(j)+(y)\rightarrow j$	Add memory to index
Set bit	sbo	$1\rightarrow Y_j$	Set j-th bit one
	sbz	$0\rightarrow Y_j$	Set j-th bit zero
	sbc	$\overline{(Y_j)}\rightarrow Y_j$	Set j-th bit complement
Step-Count	mul	$(A)\times(Y)\rightarrow AB$	Multiply
	div	$(AB)\div(Y)\rightarrow A$ (remainder), B (quotient)	Divide
	sha	$(A)\cdot 2^{(Y)}\rightarrow A$	Shift A
	sab	$(AB)\cdot 2^{(Y)}\rightarrow AB$	Shift AB together
	shb	$(B)\cdot 2^{(Y)}\rightarrow B$	Shift B
	cya	(A) cyc $(Y)\rightarrow A$	Cycle A
	cab	(AB) cyc $(Y)\rightarrow AB$	Cycle AB together
	cyb	(B) cyc $(Y)\rightarrow B$	Cycle B
	nab	$(AB)\times 2^{nf}\rightarrow AB$; $(Y)-nf\rightarrow D$	Normalize AB
	coa	$(Y)+no\rightarrow D$	Count ones in A
In-out	rds	$(Y)\rightarrow IO$; $IO\rightarrow Y$	Read and shift
	rdn	$(Y)\rightarrow IO$; $IO\rightarrow Y$	Read without shift
Jump	jpe	If $(E_i)=1$, then $y\rightarrow P$	Jump if j-th bit of E is a one
	jpp	If any $(A)>0$, then $Y\rightarrow P$	Jump if the contents of any A is positive
	jpn	If any $(A)\leq 0$, then $Y\rightarrow P$	Jump if the contents of any A is negative
	jpz	If any $(A)=0$, then $Y\rightarrow P$	Jump if the contents of any A is zero
	jpo	If any (A) overflowed, then $Y\rightarrow P$	Jump if the contents of any A has overflowed
	jxp	If $(j)\geq 0$, then $(j)-cf\rightarrow j$, $y\rightarrow P$	Jump if index positive and decrease index
	jxn	If $(j)<0$, then $(j)+cf\rightarrow j$, $y\rightarrow P$	Jump if index negative and increase index
	jpu	If $cf=1$, 3, then $(P)+1\rightarrow j$; If $cf=0$, 1, then $y\rightarrow P$; If $cf=2$, 3, then $Y\rightarrow P$	Jump unconditionally
Misc.	ios		In-out select
	opr		Operate

length dividend in A and B being divided by the memory operand. The remainder is left in A and the quotient in B. Normalize shifts the contents of A and B left until the magnitude of the number in A is between one-half and one. The number of shifts to do this, the normalizing factor, is subtracted from the memory operand in D. The shift and cycle instructions use the memory operand, rather than the address section of the instruction, to specify the number of places to shift. This is necessary since more than 18 bits are required to specify all the possible shifts for the ($\underline{9}$, $\underline{9}$, $\underline{9}$, $\underline{9}$) configuration. The count ones instruction adds the number of bits in A

TABLE II

Notation	Meaning
$\rightarrow$	goes into
(x)	contents of x
$Y=y+(j)$	indexed memory address
$\lvert(x)\rvert$	magnitude of (x)
$\overline{(x)}$	one's complement of (x)
&	logical and operation
v	inclusive or operation
$\oplus$	exclusive or operation
+	one's complement addition
nf	number of shifts to normalize
no	number of ones
Y_j	j-th digit of register Y

which are ones to the memory operand in D. This provides a simple means for determining bit density in areas of storage, since the one's count for several words can be accumulated in D.

The two replace add instructions, using the index memory, facilitate instruction and index modification. Both require two memory cycle times for execution.

The two in-out read instructions transmit information between the memory and the selected in-out unit. The details of these and the in-out select instruction are given in another paper.

Single bits in memory can be manipulated with the three bit-setting instructions. The bit-sensing instruction facilitates the use of single bits in memory as operands.

The variety of jump instructions available simplifies the coding of logical decision functions. The two-index jump instructions permit indexed program loops to refer successively in either the forwards or backwards direction to operands in a data block. The unconditional jump instruction uses the cf digits to specify whether the selected index register will be used to remember the previous contents of P. These contents are always transmitted to the E register whenever a jump occurs.

Arithmetic overflows can be caused by addition, subtraction, and division instructions. Such overflows as do occur are remembered in overflow flip-flops in the arithmetic element. The overflow condition can be detected by a jump instruction, or by the in-out element in a manner described in another paper. If an overflow is anticipated, however, it can be shifted into the A register by executing a normalize instruction. A normalize usually shifts AB left, but if an overflow exists AB is shifted right one place, and the overflow placed in the most significant digit position of A to the right of the sign digit. The memory operand is increased by one in the D register, when this occurs, rather than decreased. This interpretation of an overflow permits floating-point operations to be programmed quite simply in the arithmetic element. The in-out select and operate instructions differ from all the others in the sense that the y digits are used to specify different operations. In-out select chooses the mode in which an in-out unit will run. The operate instruction will control individual useful commands, as for example, round-off.

Instruction Times

The average execution time for instructions depends upon whether one memory or two different overlapped memories are used for instructions and operands. In the latter case the average time is the longer of the instruction memory and the operand memory cycle times, and in the first case the sum of the two cycle times. It should be remembered that any instruction which involves storing an operand in memory has the normal operand memory cycle time extended by from one to two microseconds. Instructions which alter or transfer the contents of index memory registers, require approximately two normal memory cycles even when instruction and operand memory cycles are overlapped.

Successive step counter instructions require a time which depends upon the length of the longest active arithmetic element. In the case of multiply, divide, and count ones, this time is a function of the operand word length only, but the shift, cycle, and normalize times depend upon the number of places actually shifted. Divide requires about 2 microseconds per digit and all other step counter instructions 0.4 microsecond per digit. These shift times become significant only when they exceed the one or two memory cycles already required. In the worst 36-digit case about 75 microseconds is required for division and 19 microseconds for multiplication. A 72 place shift would take 32 microseconds. These are the times required for these instructions when they are written in sequence. If the operand word length is shorter, then these times become proportionally less, down to the minimum memory times required.

Conclusion

The organization of TX-2 permits a programmer to pay considerable attention to coding details and receive a worthwhile reward in the form of increased efficiency of operation. The operating speed can be doubled when instructions and operands are stored in different memories. Further increases result by the sequencing of instructions so that non-Arithmetic-Element instructions are executed concurrently with AE step-counter instructions. And the ability to choose a configuration with each instruction means not only that some instructions take less time, but also that many of them can be eliminated from a program altogether.

However, this versatility and efficiency is not accompanied by a disastrous loss in simplicity. The system organization is such that details can be easily ignored by the naive programmer, without the details having even subtly obtrusive effects. If all the digits in an instruction word are zero except for the operation code and the base address, then TX-2 appears as a simple single address 36-bit operand word computer with a single, uniformly addressed 70,000 word memory.

If the j bits are used, then the machine is enlarged to become an indexed single-address 36-bit operand word computer for which the entire instruction code is meaningful. When the b and d bits are used, then the programmer can control the manner in which several in-out units running concurrently can cause program sequence changes. And by selecting various configurations the programmer can perform more operations simultaneously with each instruction.

The different facilities for indexing, memory overlap, instruction overlap, multiple-sequencing, and configuration can be ignored or used as the programmer desires. Ignoring them would seem to permit straightforward coding; using them actually permits much shorter and faster codes for a given function. Each facility is easily represented by a clear conceptual picture of what the facility permits, the only real difficulty being the greater number of simultaneous actions possible with each instruction. However, higher speeds and greater system capacity are obtained by shorter cycle times, increased bit storage, and greater simultaneity of events. In TX-2 all three aspects are emphasized.

Discussion

C. H. Richards (Convair-Astronautics): What is the accumulator length of TX-2, and where is the binary point located?

Mr. Frankovich: This is a 36-bit word accumulator, in a ones-complement machine. The binary point really exists only by virtue of what happens during multiplication or division type instructions. The left digit is the sign digit of whatever configuration you have, and the remaining digit is a numeric digit; and ordinarily during multiplication you can consider the binary point to be between the sign digit and the first significant digit on the remainder of the operand. During division, however, we have a different interpretation, so we cannot really say that this is a fractional machine. During addition it makes no difference where you put the binary point. During division the quotient is generated in a different register than the accumulator, so we cannot say that it is a fractional machine during that operation.

Chairman Pfister: When you multiply two 36-bit words, together you have a 72-bit product; where does the product go when you have an accumulator with only 36 bits?

Mr. Frankovich: There are 4 registers in the arithmetical element; and another register which acts as the right-hand extension of this accumulator—a B register. During multiplication the full 72-digit product is generated in the accumulator in the B register. The binary point is at the left-hand of the accumulator during the entire process. The other two registers are used, one to hold the partial carry during addition operations; another is used to carry out division, thereby enabling the arithmetical element to be completely selfcontained during such a long period of instruction.

D. L. Shell (General Electric): What happens on overflow?

Mr. Frankovich: We have four overflow indicators in the arithmetical element. If we have a full 36-bit operand for an instruction, then we use only the left-most overflow indicator, and associate one overflow indicator with each quarter; none of the other overflow indicators are affected at all. On the other hand, if we have four 9-bit operands, then we use all four overflow indicators to indicate overflow for any one of them.

I might also mention that during the jump on overflow instruction you can specify it by means of configuration control, in a very straightforward manner. There are further techniques for handling such situations which are devised to make programming easier.

Mr. Groelinger (Ramo-Wooldridge): Can the exchange element be used to store ac cumulator content in *several* places in memory?

G. G. Chapin (Remington Rand UNIVAC): Can you read from one memory element into more than one arithmetical element?

Mr. Frankovich: This can be done in two ways. As far as the one and one instruction in each transfer: if you want to store one-half of the arithmetical element in several places in the memory, be it in the left half, or the right half, wherever your location might be, then you give instructions to each transfer, unless the transfer were to be done in the same register. If you are loading the arithmetic element, you can load either half of the accumulator from a given memory register, but again this takes two instructions.

G. G. Chapin (Remington Rand UNIVAC): Can jump instructions be conditioned on more than one 9-bit section of the accumulator simultaneously?

Mr. Frankovich: Yes. The configuration control device is used universally and homogeneously upon all arithmetical instructions. If you have four 9-bit operands, and you want to jump on the basis of two of them, the jump instruction is interpreted to be "jump on either the first-quarter or the third-quarter."

A Computer-Integrated Rapid-Access Magnetic Tape System with Fixed Address

R. L. BEST
NONMEMBER AIEE

T. C. STOCKEBRAND
NONMEMBER AIEE

THIS paper describes the internal tape library system planned for the *TX-2* computer at Lincoln Laboratory, Massachusetts Institute of Technology (MIT). One hundred magnetic tape transports will be under the control of a central electronic system; the system will have a storage capacity of 10^{10} bits and an access time of about 30 seconds. It is particularly well suited for use with a computer of large random-access storage capacity such as *TX-2*, which has a core memory of $2^1/_2$ million bits. A simple tape transport having a high-speed search mode with redundant information transfer will make the ultimate library system for a computer, reliable and relatively inexpensive.

The tape transports are controlled by electronic circuits closely integrated with the computer. A permanent, constant-density timing track on the tape provides the speed reference for the control circuits and makes possible fixed position address-

R. L. BEST and T. C. STOCKEBRAND are with the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass.

The research in this document was supported jointly by the United States Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology.

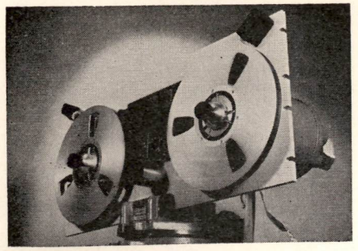


Fig. 1. Tape transport mechanism

ing and a variable rate of information transfer. The transport does not employ a constant speed capstan: two conventional 3-phase motors pull the tape in either direction. The control system varies the speed by modifying the torque on the driving motors after comparing the timing track bit rate with the desired bit rate. By this means, the computer can select an appropriate speed for recording data in real time over prolonged periods. A separate channel of block marks enables the computer to locate information blocks at any speed, including high-speed search at 920 inches per second (ips). Read and write speeds are 30 to 100 ips and acceleration time to these speeds is 1/2 to $1^1/_2$ seconds.

High reliability in the transfer of information to and from the tape is gained by making five channels out of ten redundantly paired tracks. Thus in the ten-track head assembly there are three information channels, one timing channel, and one block mark channel. Appropriate head shielding reduces crosstalk and permits reading of the timing channel when other channels are being used for writing. Since the amplitude of the signal from the tape varies greatly with speed, a system of recording is used that allows the polarity of the flux change, rather than its amplitude, to be sensed in reading. High-gain amplifiers may then be used in which the output stages are normally saturated and the gain need not be closely controlled.

Computer/Tape Library Relationship

The primary objective, a large library of quickly accessible information, is provided by a large number of tape transport mechanisms controlled by a small number of circuits which are closely integrated with the computer. One feature of *TX-2*[1,2] is its multiple sequence control; that is, it can share its attention between equipments which are operating simultaneously in real time. The computer commands the selected tape drive to attain some mode of operation, then turns its attention elsewhere until the tape control tells the computer that the desired mode has been attained. Multiple sequence control also enables the computer to vary the speed of the tape during information transfer: this feature would be used, for example, if the computation itself depended upon external, real-time events and the information transfer had to be synchronized with the results of the computation.

Another feature of *TX-2* is its large, random-access core memory[3] which can store large blocks of information at one time and which can be used as a buffer while the tape is accelerating. The control element for the tape library accepts the computer's commands rapidly but does not require instant response by the computer to events in the tape system. Generally the control takes care of the simple local problems and leaves the complex problems of operation to the computer, and therefore the programmer.

The tape mechanism, which has a wide variety of speeds, can search through a reel of tape much faster than the computer can accept information. For this reason, blocks of information on the tape are tagged with block marks that can be read at speeds up to the maximum, 920 ips. The read and write instructions command only a single 9-bit transfer between memory and tape. A series of such instructions is necessary to transfer large blocks of information; the instructions must occur at an average rate determined by the tape's speed.

Transport Design and Motion Control

Mechanical Design

The tape transports used in this system were made as simple and fool-proof as possible: they consist of a read-write head assembly, two reels, two drive motors, and a tape guide. The drive mechanism has no capstan. Thus a good deal of mechanical complexity is eliminated and a wide range of tape speeds is made possible. Fast starts and stops are precluded, however: 1/2 second and $7^1/_2$ inches of tape are required to reach 30 ips.

Figs. 1 and 2 show the transport mechanism. The motors are flange mounted, 1,800 rpm, 3-phase induction motors of the conventional type which have roughly constant torque characteristics when operating well below synchronous speed. The horsepower (hp) rating, and therefore the torque, is as high as possible, limited by the tensile strength of the tape. One-eighth-hp motors, each driven by a magnetic amplifier, provide the proper torque to operate 10-inch reels mounted directly on the motor shaft and loaded with polyester tape, 0.001-inch thick and 1/2-inch wide. Maximum tape speed is about 920 ips when the driving reel is full.

The head assembly and guide are shown in the insert, Fig. 2. The relatively large, constant radius of the guide reduces the pressure between tape and guide: At speeds above 20 ips the tape floats on an air cushion and is thus easy to edge guide. Skew, caused by nonuniform tape tension across the width of the tape and by variations in tape width,

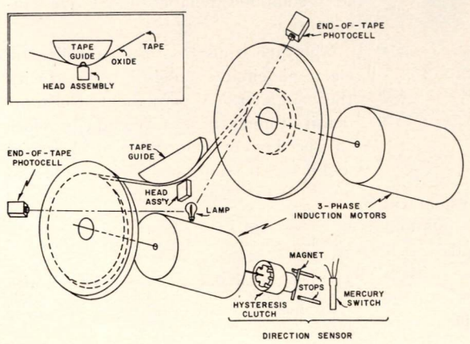


Fig. 2 (left). Tape transport mechanism

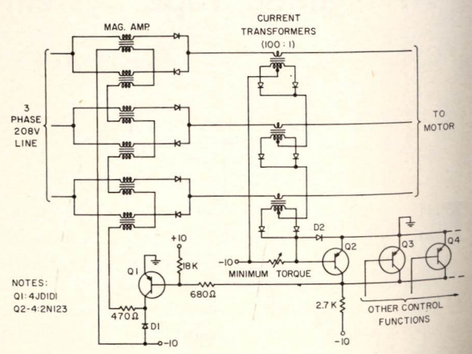


Fig. 3 (right). Motor control

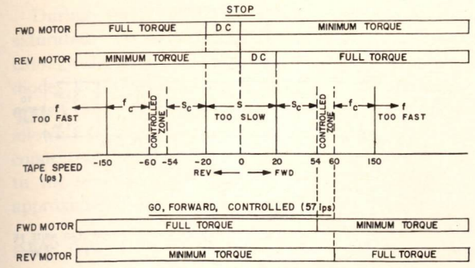


Fig. 4 (left). Motor torque versus tape speed for several command states

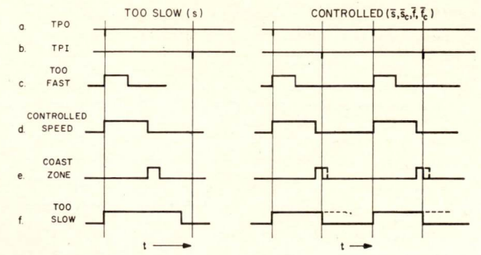


Fig. 6 (right). Speed - sensing waveforms

is minimized. There is no wrap around the head. Variations in tape tension, which are large in this transport, do not, therefore, cause excessive pressures on the head and wear is reduced. Because only short wave lengths (0.0025 and 0.005 inch) are used in the system, the area of tape-head contact need not be large.

The direction in which the transport is moving is determined by a sensing device mounted on the rear shaft extension of one motor. The sensor consists of an iron cup dragged against one of a pair of stops by hysteresis from a star-shaped permanent magnet on the motor shaft. The cup operates a mercury switch by rotating an attached magnet. This scheme gives positive direction information even at the slowest tape speed. A mercury wetted contact switch provides computer-level signals to the control without contact bounce and with good reliability.

Motion Control

Each motor can generate torque in only one direction to pull the tape from one reel to the other. The control of the motors is therefore simpler than if torque had to be reversed. Since tension is limited by tape strength, acceleration is relatively slow. A sudden change of torque, which might allow a loop to form, is prevented by a long time constant in the control windings of the motor magnetic amplifier.

To stop the tape, full torque is first applied by the trailing motor until the tape speed falls below 20 ips: at that point d-c is applied to the trailing motor to bring the tape to a smooth stop. The direction sensor indicates which motor is trailing. With d-c in the motor field winding the rotor will resist applied torque even at zero velocity due to the hysteresis in the rotor. Voltage is never completely removed from either motor in order that some tape tension always be maintained. The end of the tape is sensed by a photoelectric cell which receives light through transparent leaders at each end of the tape. The timing track is continued on the edges of the 100-foot transparent strips so that the control element will know when the tape has fallen below 20 ips as previously described.

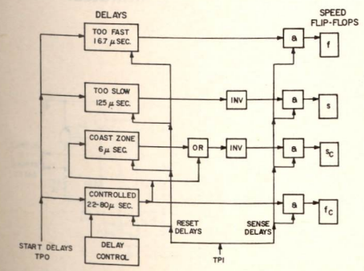


Fig. 5. Speed-sensing logic

The control circuit for one of the motors is shown in Fig. 3. The transistor, *Q*-1, regulates the current through the control windings of the 3-phase magnetic amplifier, and it switches between saturation and cutoff at various duty cycles. When *Q*-1 is cut off, *D*-1 conducts, so that the control winding-time constant is determined solely by its own inductance and the 470-ohm resistor. This time constant is made long enough to prevent abrupt torque changes and to average the control current.

Feedback was included to provide close control of minimum torque. Too much minimum torque will either allow the tape to creep or require excessive d-c holding currents in the trailing motor. Too little torque will fail to overcome static friction, and allow a loop of tape to form. The feedback prevents large variations in the output current of the magnetic amplifier which would be caused by unbalanced line voltage or small variations in reactor control current, especially when the amplifiers are nearly cut off. The feedback signal is derived from the sum of currents in all three motor leads. The diode *D*-2 limits the sum output voltage of the current transformers to 10 volts and thus keeps the voltage drop across their primaries negligible under high current conditions. To generate full torque, one of the other transistors such as *Q*-3 is saturated, cutting off the magnetic amplifier control current independent of the feedback and allowing full current to flow to the motor.

The direct current which is applied to the trailing motor when the transport is slowing to a stop is switched to one lead of the motor by a relay contact (not shown on Fig. 3). The d-c flows into that motor lead and out the other two, back through the magnetic amplifiers. Although the magnetic amplifiers are biased into a low torque condition they will pass the d-c (approximately 1 ampere) since the average voltage across any one reactor must be zero.

Digital Speed Control

To determine which motor should receive full torque, minimum torque, or d-c, the desired condition of the transport is compared with the existing one. As shown in Fig. 4, the motion control is based on a group of speed domains: too fast (f), faster than controlled (f_c), slower than controlled (s_c), and too slow (s). Various torque commands are shown as a function of speed and direction for several desired conditions.

The speed sensing logic is diagrammed in Fig. 5. The speed is detected by comparing the interval between timing pulses from the tape with the delays of delay units, as shown in Fig. 6. Two pulses are generated from a tape timing channel as it travels over the head, Fig. 6(A) and (B). The first is used to fire three delay units, two of which, Fig. 6(C) and (F), establish the boundaries between the area of usable speeds and too fast or too slow. The third delay unit, Fig. 6(D), can be set by the computer to any one of several delay times representing speeds in the useful range and provides close speed control at preset and selectable tape speeds. It in turn drives a fourth delay unit, Fig. 6(E), to provide a controlled zone. In this condition the transport coasts. The second timing pulse occurs at a time determined by the speed of the tape. It is used to sense the condition of

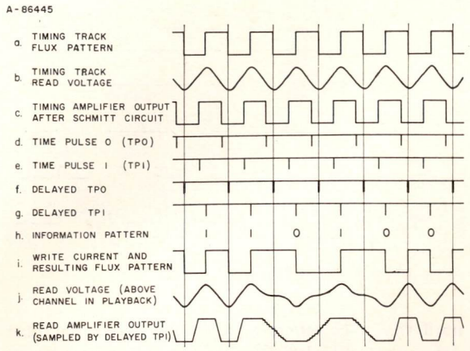


Fig. 7 (left). Phase-modulated nonreturn-to-zero waveforms

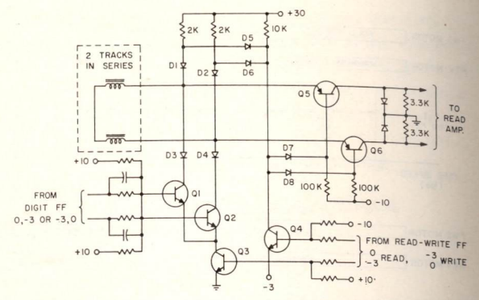


Fig. 8. Read-write switch and write circuit

the delay units and to set flip-flops to define the speed domain. It is also used to reset the units so they have time to recover before the next "set" pulse. The dotted waveforms in Fig. 6(E) and (F) indicate this resetting when the tape is in the controlled-speed zone. The delay units must be resettable, and the delay time of one must be capable of electronic variation.

Head Assembly and Read-Write Method

Head Assembly Design

The 10-track head assembly contains five channels: three information, one timing, and one block mark. Each channel consists of two redundantly paired tracks; the tracks in each pair are nonadjacent to minimize the effect of a speck of dirt lifting a portion of the tape. The timing channel occupies the two outside tracks which are heavily shielded from the interior ones in order that the timing channel may be read while the others are being written.

The timing channel controls tape speed, information density, and the fixed address feature. It assures constant information density regardless of tape speed and makes possible the changing of a single word in a message. Its density must therefore be known and constant. It is permanently recorded, either with a constant-speed capstan temporarily attached or on a separate constant-speed machine.

Read and Write Principles

Since the amplitude of the signal from the tape varies greatly with speed, a system of recording is used that allows the polarity of the flux change to be significant rather than its amplitude. High-gain amplifiers may then be used in which the gain need not be constant. Fig. 7 shows the flux pattern and other waveforms. The idealized timing-track flux pattern consists of 200 complete cycles of flux per inch, or 400-flux reversals per inch, see Fig. 7(A). The timing track read voltage, Fig. 7(B), is the expected derivative waveform from a 0.0005-inch gap looking at a signal of this density. The signal is amplified and squared in a Schmitt circuit, Fig. 7(C); the finite hysteresis of the Schmitt circuit delays the signal slightly as shown. Time pulses are generated from the transitions of the Schmitt circuit; time pulse 0 (TP0) from the negative transitions and time pulse 1 (TP1) from the positive transitions, 7(D) and 7(E). The time pulses must then be slightly delayed, 7(F) and 7(G), so that the information flux pattern may be written in phase with the timing flux pattern. The delay is a function of tape speed and is varied by an analog voltage fed to the delay circuit. The analog voltage is in turn derived from a circuit whose output is a predetermined function of the average frequency of the time pulses fed to it. The flux is laid on the tape in phase with the timing flux so that information may be read or written while the tape is moving in either direction.

The delayed time pulses control the transfer of information to the writing flip-flops. Delayed TP0 transfers the bit to be written to the flip-flop which is controlling write current; delayed TP1 complements the flip-flop. Thus a flux change is written in the center of each line corresponding to the bit to be written; there may or may not be flux changes between the lines. A typical information pattern and resulting ideal flux pattern are shown in Fig. 7(H) and 7(I). The voltage which would be read from this channel during read time is shown in Fig. 7(J). Notice that there is a saturation signal at the center of each line, whereas, in between lines there is sometimes a signal and sometimes not. The signal is then amplified more than necessary, Fig. 7(K). The amplifier has enough gain so that one of the redundant tracks may be completely separated from the head by a speck of dirt while a half-amplitude signal is being received from the other track; a saturation signal will still be delivered by the amplifier at the center of the line. The amplifier is strobed by delayed TP1, so that the logic doesn't know what the amplifier output looks like at any other time. The saturation output received at the center of each line with phase-modulated nonreturn-to-zero recording also allows the tape to be read correctly with plenty of amplifier gain margin over a wide range of tape speeds.

Read and Write Circuits

The read-write switch and write circuit for one digit are shown in Fig. 8. During "write," $Q4$ is cut off and $Q3$ is saturated. With $Q4$ cut off, its $10K$ collector resistor lifts the bases of $Q5$ and $Q6$ towards $+30$, leaving them cut off and the read amplifier disconnected. The silicon diodes at the amplifier input prevent any large voltage excursions from reaching the amplifier. With $Q3$ saturated, the digit flip-flop will cause either $Q1$ or $Q2$ to also be saturated. With the circuit values shown, 15 milliamperes (ma) will flow through the two series-connected tracks and 30 ma through the saturated transistor ($Q1$ or $Q2$). The direction of current flow through the tracks is determined by the flip-flop state; i.e., whether $Q1$ or $Q2$ is saturated.

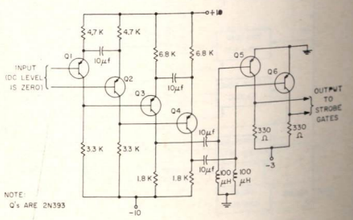


Fig. 9. Read amplifier

During "read," $Q3$ is cut off and $Q4$ is saturated. $Q4$ takes the full write current through diodes $D5$ and $D6$, back biasing diodes $D1$ through $D4$. With $Q4$ saturated diodes $D7$ and $D8$ are back biased, allowing $Q5$ and $Q6$ to be saturated, thus connecting the two series-connected tracks to the read amplifier at a d-c level of approximately zero.

Read Amplifier

The read amplifier, Fig. 9, has two difference-amplifier stages and one output stage with more than enough gain to give a saturation output signal at a tape speed of 20 ips. In the first two stages, the common mode gain per stage is less than unity while the difference signal gain is approximately beta. The low common mode gain insures that power supply noise will not be amplified. Each transistor ($Q1$-$Q4$) is biased to a constant d-c operating point of approximately 3.8-smitter-collector volts and 1.9-ma collector current. The capacitors shown must only be large enough to have negligible signal attenuation at 20 ips, the lowest tape speed of interest. The lowest frequency signal will be at 20 ips and 100 cycles per inch [alternate ones and zeros; see Fig. 7(J)] for a frequency of 2 kc. The highest frequency will be at 920 ips and 200 cycles per inch (all ones or all zeros) for a frequency of 184 kc. The micro-alloy *2N393* transistors have more than enough bandwidth for this application. The signal amplitude at the input to $Q5$ and $Q6$ is large enough so that, most of the time, one of these transistors is saturated. The output signals are of a computer-type amplitude (0 or -3) and are sampled by TP1 using conventional *TX-2* logical circuits.[4]

References

1. Functional Description of TX-2 Computer, **J. M. Frankovich, H. P. Peterson.** *Proceedings of the Western Joint Computer Conference*, Institute of Radio Engineers, New York, N. Y., 1957, pp. 146–50.

2. The Lincoln TX-2 Input-Output System, **J. W. Forgie.** *Ibid.*, pp. 156–60.

3. Memory Units in the Lincoln TX-2, **R. L. Best.** *Ibid.*, pp. 160–66.

4. Transistor Circuitry in the Lincoln TX-2, **K. H. Olsen.** *Ibid.*, pp. 167–71.

Reprinted from Special Publication T-107
Proceedings of the
Western Joint Computer Conference
1958

COMPARISON OF FERRITE AND HYPERSIL TRANSFORMERS IN GATE TUBE CIRCUIT

$N_1/N_2 = 27/9$ (FERRITE)

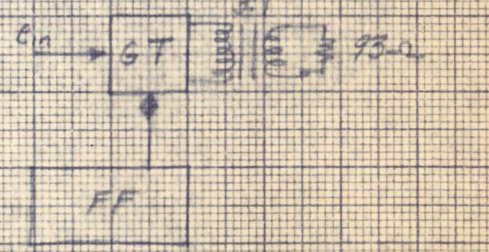
LEGEND
- #1 - HYPERSIL (193-7)
- #2 - FERRITE - BUNCH WOUND
- #3 - FERRITE - LAYER WOUND - MINIMUM COUPLING
- #4 - FERRITE - LAYER WOUND ON SAME AREA

#2 to #4 not potted

OUTPUT PULSE AMPLITUDE

INPUT PULSE AMPLITUDE

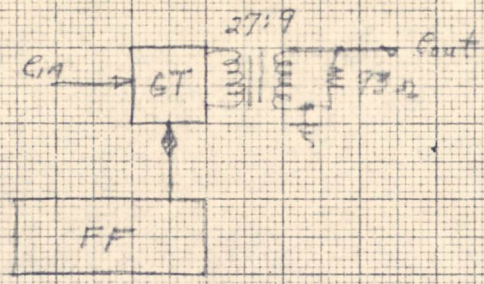
UNITY GAIN



5 HJP P3 71
22 JUNE 1953

EFFECT OF POTTING ON FERRITE TRANSFORMER PERFORMANCE

BOTH WINDINGS LAYER WOUND ONE UPON THE OTHER



GATE TUBE
E_{bb} = +250
E_{c2} = 90 v
I_{bb} = 37.0 ma
FF clamped T_c = 10 v

HYPERSIL

FERRITE { NOT POTTED / POTTED

OUTPUT PULSE AMPLITUDE

0 4 8 12 16 20 24 28 32 36 40

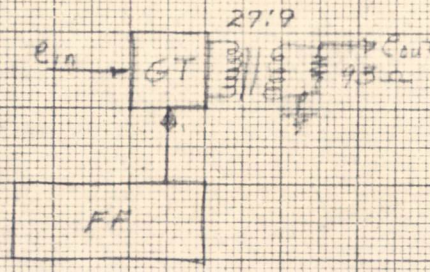
0 5 10 15 20 25 30 35 40

INPUT PULSE AMPLITUDE

5 HJP Fig 11
29 JUNE 1953

EFFECT OF POTTING ON FERRITE TRANSFORMER PERFORMANCE

SINGLE LAYER WINDINGS 180° APART



GATE TUBE:
E_{bb} = 250v
E_{c2} = 90v
I_{bb} = 370 ma
FF clamped to +10v

HYPERSIL
FERRITE { NOT POTTED / POTTED

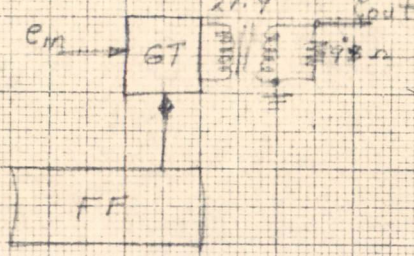
OUTPUT PULSE AMPLITUDE

INPUT PULSE AMPLITUDE

5 HJP P3 71
29 JUNE 1963

EFFECT OF POTTING ON FERRITE TRANSFORMER PERFORMANCE

BOTH WINDINGS WOUND BUNCHED UP IN SAME AREA



GATE TUBE

E_{bb} = 250 v

E_{c2} = 90 v

I_{bb} = 37.0 ma

FF clamped to +10 v

FERRITE NOT POTTED

HYPERSIL

FERRITE POTTED

OUTPUT PULSE AMPLITUDE

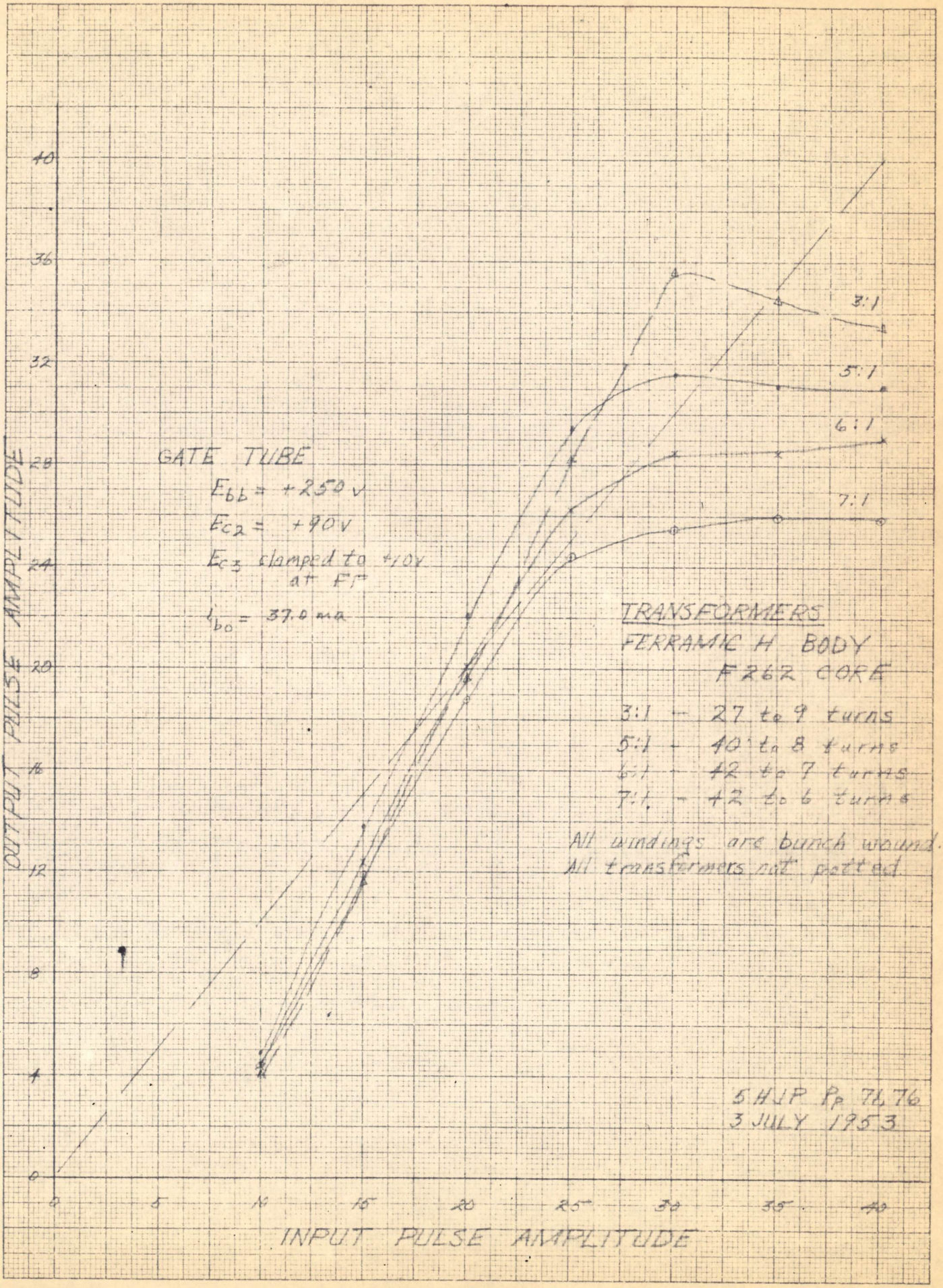
0 4 8 12 16 20 24 28 32 36 40

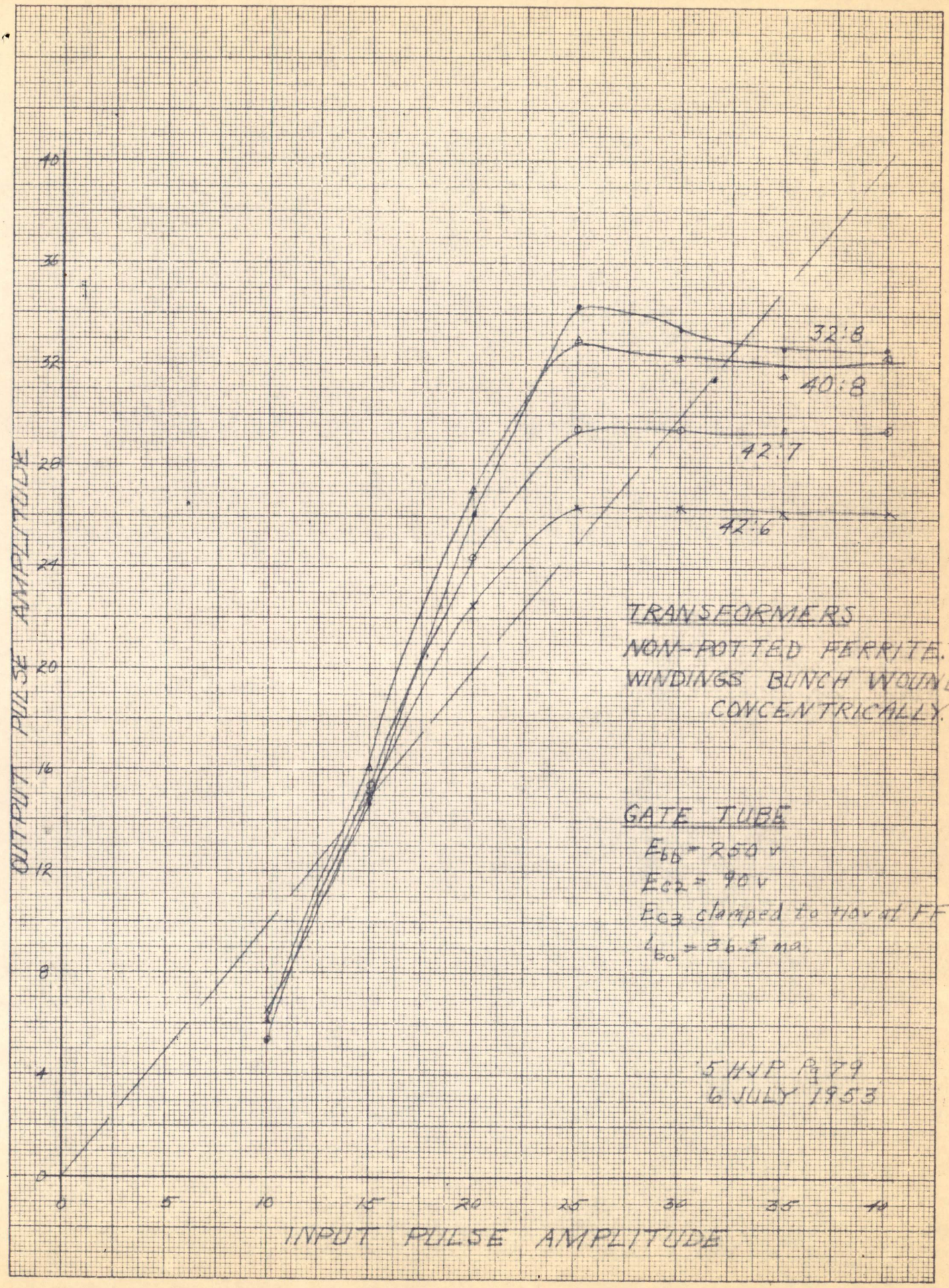
INPUT PULSE AMPLITUDE

0 5 10 15 20 25 30 35 40

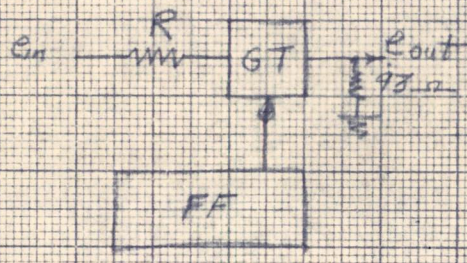
S.H.J.P. Fig 71

29 JUNE 1953





EFFECT OF SERIES RESISTANCE IN CONTROL GRID ON GATE TUBE



R = 200 OHMS

R = 100 OHMS

R = 0 OHMS

R = 470 OHMS

R = 1000 OHMS

UNITY GAIN

E_{bb} = 250 V
E_{c2} = 90 V
FF OUTPUT CLAMPED TO +10 V.
I_{bb} = 37.0 ma.

5HJP Pg. 74
29 JUNE 1953

OUTPUT PULSE AMPLITUDE

0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 40

0 5 10 15 20 25 30 35 40

INPUT PULSE AMPLITUDE

MIT REPRINTS

A Computer-Integrated Rapid-Access Magnetic Tape System with Fixed Address
R. L. BEST T. C. STOCKEBRAND

A Functional Description of the Lincoln TX-2 Computer
J. M. FRANKOVICH H. P. PETERSON

The Lincoln TX-2 Computer Development
WELSLEY A. CLARK

The Lincoln TX-2 Input-Output System
JAMES W. FORGIE

Transistor Circuitry in the Lincoln TX-2
KENNETH H. OLSEN

TX-0, A Transistor Computer with a 256 by 256 Memory
J. L. MITCHELL K. H. OLSEN

THE LINCOLN TX-2 INPUT-OUTPUT SYSTEM

BY

JAMES W. FORGIE

Reprinted from the PROCEEDINGS OF THE WESTERN JOINT COMPUTER CONFERENCE
Los Angeles, California, February 1957

PRINTED IN THE U.S.A.

The Lincoln TX-2 Input-Output System*

JAMES W. FORGIE†

Introduction

THE input-output system of the Lincoln TX-2 computer contains a variety of input-output devices suitable for general research and control applications. The system is designed in such a way that several input-output devices may be operated simultaneously. Since the computer is experimental in nature, and changes in the complement of input-output devices are anticipated, the modular scheme used will facilitate expansion and modification. The experimental nature of the computer also requires that the input-output system provide a maximum of flexibility in operating and programming for its input-output devices.

The input-output devices, currently scheduled for connection to TX-2, include magnetic-tape units for auxiliary storage; photoelectric paper-tape readers for program input; a high-speed printer, cathode-ray-tube displays, and Flexowriters for direct output; analog-to-digital conversion equipment; data links with other computers; and miscellaneous special-purpose equipment. This paper will not be concerned with the details of these devices, but will limit itself to a discussion of the logical incorporation of them into the system.

In describing the TX-2 input-output system, reference will be made to certain design aspects of other parts of the TX-2 as set forth in the previous paper.

The Multiple-Sequence Program Technique

Of the various organizational schemes which permit the simultaneous operation of many devices, we have chosen the "multiple-sequence program technique" for incorporation in TX-2. A multiple-sequence computer is one that has several program (instruction) counters. If the program sequences associated with these program counters are arranged to time-share the hardware of the central computer, a machine can be obtained which will behave as if it were a number of logically separate computers. We call these logical computers *sequences* and therefore refer to TX-2 as a *multiple-sequence* computer. By associating each input-output device with such a sequence, we effectively obtain an input-output computer for each device.

Since the one physical computer in which these sequences operate is capable of performing only one instruction at a time, it is necessary to interleave the sequences if they are to operate simultaneously. This interleaving process can take place aperiodically to suit the needs of and under the control of, whatever individual input-output devices are operating. The number of sequences which can operate simultaneously, and the complexity of the individual sequences, is limited by the peak and average data-handling rate of the central computer hardware.

In a multiple-sequence computer, the main body of the computation can be carried out in any sequence, but if maximum efficiency of input-output operation is to be achieved, the bulk of arithmetic operations must be confined to a few special sequences, called *main* sequences, which have no associated input-output devices. The input-output sequences may then be kept short, and a large number of them can be executed at once.

Multiple-Sequence Operation in TX-2

In TX-2, one-half of the index-register memory has been made available for storing program counters. Thus, a total of 32 sequences may be operated in the machine. (Actually an additional sequence of special characteristics is obtained by using index register number 0 as a program counter. This special sequence will be discussed later.) Some of these sequences are associated with input-output devices. Others perform functions, such as interpreting arithmetic overflows, that are called into action by conditions arising within the central computer. Finally, there are the main sequences which are intended to carry out the bulk of the arithmetic computations performed by the machine.

A priority scheme is used to determine which sequence will control the computer at a given time. If more than one sequence requires attention at the same time, control of the machine will go to the sequence having the highest priority, and instructions addressed by its program counter will be executed.

Table I is a list of the sequences currently planned for inclusion in TX-2. They are listed in approximate order of priority with the highest at the top. Asterisks mark sequences which are not associated with any particular in-out device. A special sequence (number 0) has first priority and will be used to start any of the other sequences at arbitrary addresses. The next two sequences interpret alarms (under program control). These three sequences have the highest priorities, since they must be capable of interrupting the activities of other sequences. The input-output devices follow, with high-speed, free-running units carrying next highest priorities. The main sequences (we anticipate three) are at the bottom of the list. The priority of any sequence may be easily changed, but such changes are not under program control. Priorities are intended to remain fixed under normal operating conditions. The list totals about 25 sequences, leaving eight spaces for future expansion.

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with Mass. Inst. Tech.

† Lincoln Lab., M.I.T., Lexington, Mass.

TABLE I

TX-2 Sequence Assignments in the Order of Their Priority

*Start-Over (special index register number 0 sequence)
*In-out alarms
*Arithmetic alarms (overflows, etc.)
Magnetic tape units (several sequences)
High-speed printer
Analog-to-digital converter
Photoelectric paper tape readers (several sequences)
Light pen (photoelectric pick-up device)
Display (several sequences)
MTC (Memory Test Computer)
TX-0
Digital-to-analog converter
Paper tape punch
Flexowriters (several sequences)
*Main sequences (three)

* The sequences have no input-output device.

Switching between sequences is under the control of both the input-output devices (generalized to include alarms, etc.) and the programmed instructions within the sequence.

Once a sequence is selected and its instructions are controlling the computer, further switching is under control of the programmed instructions. Program control of sequence switching is maintained through two bits, called the *break* and *dismiss* bits, in each instruction. The *break* bit governs changes to higher-priority sequences. When the *break* bit permits a change, and some higher-priority sequence requests attention, a change will be made. The *dismiss* bit indicates that the sequence has completed its operation (for the moment, at least) and that lower-priority sequences may receive attention. The interpretation of the *break* and *dismiss* bits will be discussed in more detail.

The TX-2 Input-Output Element

The TX-2 input-output element is shown schematically in Fig. 1. It consists of a number of input-output devices, associated buffers, and a sequence selector. Each device has enough control circuitry to permit it to operate in some selected mode once it has been placed in that mode by signals from the central computer. Associated with each device is a buffer storage of appropriate size. This buffer may be large or small, to suit individual data-rate requirements, but the buffers used in TX-2 will generally be the smallest possible. For the most part, buffering for only one line of data from the device (*e.g.*, 6 bits for a paper-tape reader) will be provided. Each input-output device is associated with one stage of the sequence selector. The sequence selector provides the control information necessary for proper interleaving of the program sequences. When it is desired to add a new input-output device to the computer, the three packages, in-out unit, buffer, and sequence-selector stage, must be provided.

As shown in Fig. 1, data is transferred between the input-output element and the central computer by way of the exchange element. Fig. 1 indicates two-way paths between the E register and all in-out buffers. Actually,

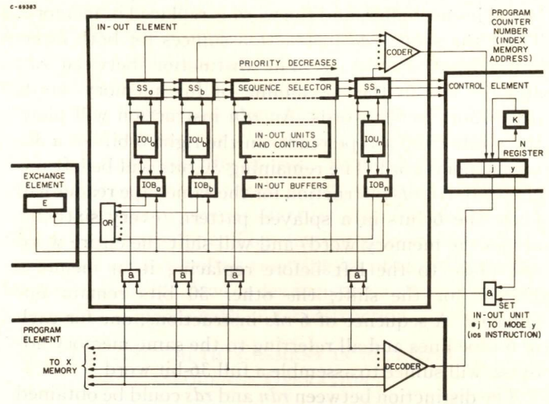


Fig. 1—Block diagram of TX-2 in-out element.

most devices are either readers or recorders, but not both, and therefore require one-way paths only. Only the necessary paths are provided; the drawing simply shows the most general case.

Signals from the sequence selector connect the appropriate buffer register to the E register to transfer data. When a sequence is selected (*i.e.*, its program counter is supplying instruction locations), the associated buffer is connected to the E register, and all other buffers are disconnected. A *read* instruction will effect a transfer of information between the buffer and the E register. A particular buffer is thus accessible only to *read* instructions in the sequence associated with the buffer's in-out unit.

Fig. 1 shows paths from the sequence selector to a coder which provides an output called the program-counter number. These paths are used in the process of changing sequences to be described in a later section.

Fig. 1 also shows paths for mode selection in the in-out element. The use of these paths is described in the next section under *ios*.

Input-Output Instructions

In addition to the break and dismiss bits on all instructions, the programmer has three computer instructions for operating the input-output system. There are two *read* instructions, *rdn* and *rds*, which transfer data between the in-out devices and the central computer memory. The third instruction, *ios*, selects the mode of operation of the in-out devices.

rdn and rds

Both of the *read* instructions obtain a word from memory. If the in-out device associated with the sequence in which the read instruction occurs is in a reading (input) mode, appropriate bits of the memory word are altered, and the modified word is replaced in memory. If the in-out device is in a recording (output) mode, appropriate bits of the memory word are fed to the se-

lected in-out buffer, and the word is replaced in memory. Thus, the same *read* instruction suffices for both input and output operations. The distinction between *rdn* and *rds* lies in the assembling of full memory words from short buffer words. An *rdn* instruction will place the 6 bits from a tape reader in the right 6 bits of a 36-bit memory word. The remaining 30 bits will be left unchanged. An *rds* instruction for the same tape reader will place the 6 bits in a splayed pattern (every sixth bit across the memory word) and will shift the entire word one place to the left before replacing it in memory. Except for the shift, the other 30 bits remain unchanged. A sequence of 6 *rds* instructions, one for each of 6 tape lines and all referring to the same memory address, will suffice to assemble a full 36-bit word.

The distinction between *rdn* and *rds* could be obtained from mode information in the in-out device, but the inclusion of both instructions in the order code allows the programmer to interchange the two types freely to suit his needs. The *rdn* instruction makes use of the permutation aspect of the TX-2 configuration control and is, therefore, particularly convenient for dealing with alphanumeric Flexowriter characters. Configuration is not applicable to the *rds* instruction.

ios

The *ios* instruction serves to put a particular in-out device into a desired mode of operation. The j bits of the instruction word, normally the index register number, in this case specify the unit number of the in-out device. This number is the same as the program counter number for the associated sequence, although the correspondence is not necessary. The y bits of the instruction word specify the mode of operation in which the unit is to be placed. Two of the y bits are sent directly to the jth sequence selector stage and serve to control the sequence, regardless of the mode of its associated in-out device. These two bits allow *ios* instructions to arbitrarily dismiss or request attention for any sequence in the machine. By means of these instructions, one sequence can start or stop all others in the machine. A third y bit determines whether the mode of the in-out device is to change as a result of the instruction. If it is to change, the remaining 15 bits specify the new mode. An *ios* instruction occurring in any sequence can thus start or stop any sequence and/or change the mode of its in-out device.

A further property of the *ios* instruction is that it leaves in the E register a map of the state of the specified in-out control prior to any changes resulting from the instruction itself; *ios* instructions may, therefore, be used to sense the state of the in-out system without altering it in any way.

Sequence-Changing and Operation of the Sequence-Selector

At some point just before the completion of the instruction memory cycle in TX-2, the Control must decide whether the next instruction would be taken from the current sequence or from some new sequence. The information on which this decision must be based comes from the *break* and *dismiss* bits of the instruction word currently in use and from the sequence selector. Fig. 2 is a detailed drawing of one stage of the sequence selector. All stages, except that with the highest-priority, are identical. The lowest-priority stage returns the final three control signals to the control element.

Each stage of the sequence selector retains two pieces of information concerning its associated sequence. One flip-flop (ss j.1) remembers whether or not the sequence is selected (*i.e.*, whether or not it is receiving attention). The priority signal (labeled *no higher priority sequence requests attention*) passes from higher to lower priority stages until it encounters a stage which requests, but is not receiving attention. Such a stage is said to have priority at the moment, and its output to the program-counter-number coder prepares the number of the new program counter in anticipation of a sequence change.

The process of changing sequences involves storing the program counter for the old sequence and obtaining the counter for the new. Actually, to speed up the overall process, the new program counter is obtained first, so that it may be used while the old is being stored. Using the paths shown in Fig. 1, the new program counter number is placed in the j bits of the N register. The new program counter is then obtained from the X memory and interchanged with the old program counter contents which have been in the P register.[1] The K register, which has been holding the old program counter number since the last sequence change, is now interchanged with the j bits, and the old counter is stored at the proper location in the X memory. The state of the sequence selector is changed, to conform to the change of sequence, by sending a *select new sequence* command from Control. This command clears the ss j.2 flip-flop in the old-sequence stage and sets the ss j.2 flip-flop to a ONE in the new-sequence stage.[2]

Interpretation of the Break Bit

The programmer uses the *break* bit of an instruction word to indicate whether or not change to a higher priority sequence may occur at the completion of the instruction. The fact that a programmer permits a *break* does not mean that the sequence has completed its current task, but merely that no harm will be done if a change to some higher-priority sequence is made. *Breaks* should be permitted at every opportunity if a number of in-out devices are operating. The sort of situation in which a *break* cannot be permitted occurs when the E register is left containing information which the program requires at a later step. If a change occurred in this case, the contents of the E register would be destroyed and lost to the program.

[1] The P register is shown in Fig. 4 of Frankovich and Peterson, this issue, p. 148.

[2] The relative timing of the central computer actions during the change process is shown in Fig. 6(d) of Frankovich and Peterson, this issue, p. 150.

When a *break* is permitted by the current instruction, a sequence change will actually take place only if some higher-priority sequence requests attention. A signal from the sequence selector to the control element provides this information (Fig. 2). When a *break* type of sequence change is made, the ss j.1 flip-flop in the sequence selector remains unchanged, and the sequence which was abandoned in favor of one of a higher-priority continues to request attention.

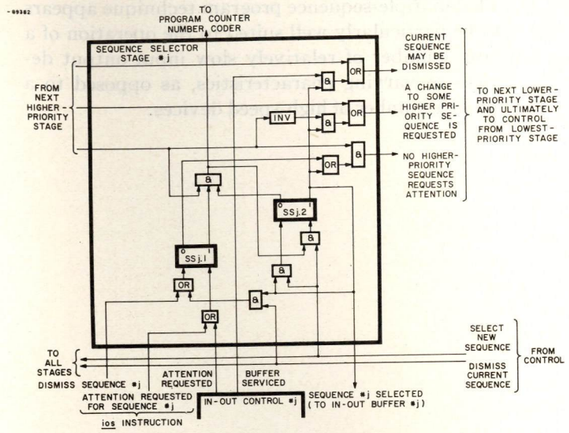


Fig. 2—Block diagram of TX-2 sequence selector stage.

Interpretation of the Dismiss Bit

The *dismiss* bit is used by the programmer to indicate that the sequence presently in use has completed its task. To provide synchronization in the in-out system, *dismiss* bits must be programmed between attention requests from the in-out devices. In this case, the *dismiss* operation guarantees that the computer will wait for the next signal from the in-out device before proceeding with the associated program sequence.

The *dismiss* bit is also used to accomplish the halt function in TX-2. A multiple-sequence computer *halts* when all sequences have been dismissed and all in-out units turned off. The priority signal from the sequence selector to the control element provides the information as to whether or not any sequence in the machine requests attention. When none request attention, the control stops all activity in the machine as soon as a *dismiss* bit appears on an instruction in the sequence being used. Activity is resumed in the machine as soon as some in-out device or push button requests attention.

The sequence change which results from a *dismiss* bit is identical with that resulting from a *break* except that a *dismiss current sequence* command accompanies the *select new sequence* command from Control to the Sequence Selector (Fig. 2).

Starting a Multiple-Sequence Computer

In a single-sequence computer the starting process involves resetting the program counter to some arbitrary value and starting the control. In a multiple-sequence computer, the program counter for a particular sequence must be reset and the sequence started. In TX-2 a special sequence (number 0) has the highest priority and is used to facilitate starting. This sequence has the special feature that its program counter always starts at an initial memory location specified by a set of toggle switches. Attention for the sequence is requested by pushing a button on the console. By executing a short program stored in the toggle-switch registers of the V memory, this sequence can start (or stop) any other sequence in the machine. The starting process for an arbitrary sequence involves resetting its program counter by means of an *ldx* (load index register) instruction, and starting its sequence with an *ios* instruction.

The Arithmetic Element in Multiple-Sequence Operation

While efficient operation requires that the bulk of arithmetic operations be carried out in a main sequence, the arithmetic element in TX-2 is available to all sequences. Since once a change has been made to a higher-priority sequence, control cannot return to a lower-priority sequence until the higher-priority one has been dismissed, a simple rule allows the arithmetic element to be used in any sequence without confusion. If, whenever a higher-priority sequence requires the arithmetic element, it stores the contents of any registers it will need (A, B, C, D, or F) and reloads them before dismissing, all lower-priority sequences will find the registers as they left them. This storing and loading operation requires time and, therefore, lowers the total data-handling capacity, but the flexibility obtained may well be worth the loss in capacity.

The step-counter class of arithmetic element instructions is a special problem. These instructions can require many microseconds to complete, and while TX-2 is designed to allow in-out and program element instructions to take place while the arithmetic element is busy, the case can arise in which an arithmetic element instruction (load, store, etc.) appears before the AE is finished with a step-counter class instruction. The machine would normally wait in an inactive state until the operation is complete, but since there is a chance that some higher-priority sequence may request attention in the interim and have instructions which can be carried out, provision is made to keep trying changes to higher-priority sequences as they request attention. The machine thus waits in an inactive state only when no higher-priority sequences have instructions which can be performed. This provision allows the programmer to ignore the arithmetic element in considerations of peak- and average-peak rate calculations when he desires to operate a maximum number of in-out devices.

Conclusion

Multiple-sequence operation of input-output devices, as realized in TX-2, has a number of significant characteristics. Among them are:

1) A number of in-out devices may be operated concurrently with a minimum of buffering storage.

2) Machine time is used efficiently, since no time need be lost waiting for input-output devices to complete their operation. Other machine activity may proceed meanwhile.
3) Each input-output device may be treated separately for programming purposes. Efficiency of operation is obtained automatically when several separately programmed devices are operated simultaneously, although average- and peak-rate limitations must be considered.
4) Maximum flexibility in programming for input-output devices is obtained. The full power of the central machine may be used by each input-output sequence if desired. Routines for each device may be as long or as short as the particular situation requires.
5) The modular organization of the input-output equipment simplifies additions and modifications to the complement of in-out devices.
6) The organization of buffering storage allows the amount and kind of such storage to be tailored to the needs of the individual devices and the data-handling requirements to be met by the system.
7) The multiple-sequence program technique appears to be particularly well suited to the operation of a large number of relatively slow input-output devices of varying characteristics, as opposed to a smaller number of high-speed devices.

THE LINCOLN TX-2 INPUT-OUTPUT SYSTEM

BY

JAMES W. FORGIE

Reprinted from the PROCEEDINGS OF THE WESTERN JOINT COMPUTER CONFERENCE
Los Angeles, California, February 1957

PRINTED IN THE U.S.A.

The Lincoln TX-2 Input-Output System*

JAMES W. FORGIE†

Introduction

THE input-output system of the Lincoln TX-2 computer contains a variety of input-output devices suitable for general research and control applications. The system is designed in such a way that several input-output devices may be operated simultaneously. Since the computer is experimental in nature, and changes in the complement of input-output devices are anticipated, the modular scheme used will facilitate expansion and modification. The experimental nature of the computer also requires that the input-output system provide a maximum of flexibility in operating and programming for its input-output devices.

The input-output devices, currently scheduled for connection to TX-2, include magnetic-tape units for auxiliary storage; photoelectric paper-tape readers for program input; a high-speed printer, cathode-ray-tube displays, and Flexowriters for direct output; analog-to-digital conversion equipment; data links with other computers; and miscellaneous special-purpose equipment. This paper will not be concerned with the details of these devices, but will limit itself to a discussion of the logical incorporation of them into the system.

In describing the TX-2 input-output system, reference will be made to certain design aspects of other parts of the TX-2 as set forth in the previous paper.

The Multiple-Sequence Program Technique

Of the various organizational schemes which permit the simultaneous operation of many devices, we have chosen the "multiple-sequence program technique" for incorporation in TX-2. A multiple-sequence computer is one that has several program (instruction) counters. If the program sequences associated with these program counters are arranged to time-share the hardware of the central computer, a machine can be obtained which will behave as if it were a number of logically separate computers. We call these logical computers *sequences* and therefore refer to TX-2 as a *multiple-sequence* computer. By associating each input-output device with such a sequence, we effectively obtain an input-output computer for each device.

Since the one physical computer in which these sequences operate is capable of performing only one instruction at a time, it is necessary to interleave the sequences if they are to operate simultaneously. This interleaving process can take place aperiodically to suit the needs of and under the control of, whatever individual input-output devices are operating. The number of sequences which can operate simultaneously, and the complexity of the individual sequences, is limited by the peak and average data-handling rate of the central computer hardware.

In a multiple-sequence computer, the main body of the computation can be carried out in any sequence, but if maximum efficiency of input-output operation is to be achieved, the bulk of arithmetic operations must be confined to a few special sequences, called *main* sequences, which have no associated input-output devices. The input-output sequences may then be kept short, and a large number of them can be executed at once.

Multiple-Sequence Operation in TX-2

In TX-2, one-half of the index-register memory has been made available for storing program counters. Thus, a total of 32 sequences may be operated in the machine. (Actually an additional sequence of special characteristics is obtained by using index register number 0 as a program counter. This special sequence will be discussed later.) Some of these sequences are associated with input-output devices. Others perform functions, such as interpreting arithmetic overflows, that are called into action by conditions arising within the central computer. Finally, there are the main sequences which are intended to carry out the bulk of the arithmetic computations performed by the machine.

A priority scheme is used to determine which sequence will control the computer at a given time. If more than one sequence requires attention at the same time, control of the machine will go to the sequence having the highest priority, and instructions addressed by its program counter will be executed.

Table I is a list of the sequences currently planned for inclusion in TX-2. They are listed in approximate order of priority with the highest at the top. Asterisks mark sequences which are not associated with any particular in-out device. A special sequence (number 0) has first priority and will be used to start any of the other sequences at arbitrary addresses. The next two sequences interpret alarms (under program control). These three sequences have the highest priorities, since they must be capable of interrupting the activities of other sequences. The input-output devices follow, with high-speed, free-running units carrying next highest priorities. The main sequences (we anticipate three) are at the bottom of the list. The priority of any sequence may be easily changed, but such changes are not under program control. Priorities are intended to remain fixed under normal operating conditions. The list totals about 25 sequences, leaving eight spaces for future expansion.

* The research in this document was supported jointly by the Army, Navy, and Air Force under contract with Mass. Inst. Tech.
† Lincoln Lab., M.I.T., Lexington, Mass.

TABLE I

TX-2 Sequence Assignments in the Order of Their Priority

*Start-Over (special index register number 0 sequence)
*In-out alarms
*Arithmetic alarms (overflows, etc.)
Magnetic tape units (several sequences)
High-speed printer
Analog-to-digital converter
Photoelectric paper tape readers (several sequences)
Light pen (photoelectric pick-up device)
Display (several sequences)
MTC (Memory Test Computer)
TX-0
Digital-to-analog converter
Paper tape punch
Flexowriters (several sequences)
*Main sequences (three)

* The sequences have no input-output device.

Switching between sequences is under the control of both the input-output devices (generalized to include alarms, etc.) and the programmed instructions within the sequence.

Once a sequence is selected and its instructions are controlling the computer, further switching is under control of the programmed instructions. Program control of sequence switching is maintained through two bits, called the *break* and *dismiss* bits, in each instruction. The *break* bit governs changes to higher-priority sequences. When the *break* bit permits a change, and some higher-priority sequence requests attention, a change will be made. The *dismiss* bit indicates that the sequence has completed its operation (for the moment, at least) and that lower-priority sequences may receive attention. The interpretation of the *break* and *dismiss* bits will be discussed in more detail.

The TX-2 Input-Output Element

The TX-2 input-output element is shown schematically in Fig. 1. It consists of a number of input-output devices, associated buffers, and a sequence selector. Each device has enough control circuitry to permit it to operate in some selected mode once it has been placed in that mode by signals from the central computer. Associated with each device is a buffer storage of appropriate size. This buffer may be large or small, to suit individual data-rate requirements, but the buffers used in TX-2 will generally be the smallest possible. For the most part, buffering for only one line of data from the device (*e.g.*, 6 bits for a paper-tape reader) will be provided. Each input-output device is associated with one stage of the sequence selector. The sequence selector provides the control information necessary for proper interleaving of the program sequences. When it is desired to add a new input-output device to the computer, the three packages, in-out unit, buffer, and sequence-selector stage, must be provided.

As shown in Fig. 1, data is transferred between the input-output element and the central computer by way of the exchange element. Fig. 1 indicates two-way paths between the E register and all in-out buffers. Actually,

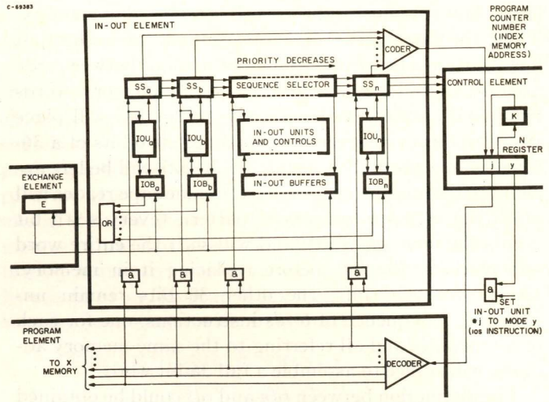


Fig. 1—Block diagram of TX-2 in-out element.

most devices are either readers or recorders, but not both, and therefore require one-way paths only. Only the necessary paths are provided; the drawing simply shows the most general case.

Signals from the sequence selector connect the appropriate buffer register to the E register to transfer data. When a sequence is selected (*i.e.*, its program counter is supplying instruction locations), the associated buffer is connected to the E register, and all other buffers are disconnected. A *read* instruction will effect a transfer of information between the buffer and the E register. A particular buffer is thus accessible only to *read* instructions in the sequence associated with the buffer's in-out unit.

Fig. 1 shows paths from the sequence selector to a coder which provides an output called the program-counter number. These paths are used in the process of changing sequences to be described in a later section.

Fig. 1 also shows paths for mode selection in the in-out element. The use of these paths is described in the next section under *ios*.

Input-Output Instructions

In addition to the break and dismiss bits on all instructions, the programmer has three computer instructions for operating the input-output system. There are two *read* instructions, *rdn* and *rds*, which transfer data between the in-out devices and the central computer memory. The third instruction, *ios*, selects the mode of operation of the in-out devices.

rdn and rds

Both of the *read* instructions obtain a word from memory. If the in-out device associated with the sequence in which the read instruction occurs is in a reading (input) mode, appropriate bits of the memory word are altered, and the modified word is replaced in memory. If the in-out device is in a recording (output) mode, appropriate bits of the memory word are fed to the se-

lected in-out buffer, and the word is replaced in memory. Thus, the same *read* instruction suffices for both input and output operations. The distinction between *rdn* and *rds* lies in the assembling of full memory words from short buffer words. An *rdn* instruction will place the 6 bits from a tape reader in the right 6 bits of a 36-bit memory word. The remaining 30 bits will be left unchanged. An *rds* instruction for the same tape reader will place the 6 bits in a splayed pattern (every sixth bit across the memory word) and will shift the entire word one place to the left before replacing it in memory. Except for the shift, the other 30 bits remain unchanged. A sequence of 6 *rds* instructions, one for each of 6 tape lines and all referring to the same memory address, will suffice to assemble a full 36-bit word.

The distinction between *rdn* and *rds* could be obtained from mode information in the in-out device, but the inclusion of both instructions in the order code allows the programmer to interchange the two types freely to suit his needs. The *rdn* instruction makes use of the permutation aspect of the TX-2 configuration control and is, therefore, particularly convenient for dealing with alphanumeric Flexowriter characters. Configuration is not applicable to the *rds* instruction.

ios

The *ios* instruction serves to put a particular in-out device into a desired mode of operation. The j bits of the instruction word, normally the index register number, in this case specify the unit number of the in-out device. This number is the same as the program counter number for the associated sequence, although the correspondence is not necessary. The y bits of the instruction word specify the mode of operation in which the unit is to be placed. Two of the y bits are sent directly to the jth sequence selector stage and serve to control the sequence, regardless of the mode of its associated in-out device. These two bits allow *ios* instructions to arbitrarily dismiss or request attention for any sequence in the machine. By means of these instructions, one sequence can start or stop all others in the machine. A third y bit determines whether the mode of the in-out device is to change as a result of the instruction. If it is to change, the remaining 15 bits specify the new mode. An *ios* instruction occurring in any sequence can thus start or stop any sequence and/or change the mode of its in-out device.

A further property of the *ios* instruction is that it leaves in the E register a map of the state of the specified in-out control prior to any changes resulting from the instruction itself; *ios* instructions may, therefore, be used to sense the state of the in-out system without altering it in any way.

Sequence-Changing and Operation of the Sequence-Selector

At some point just before the completion of the instruction memory cycle in TX-2, the Control must decide whether the next instruction would be taken from the current sequence or from some new sequence. The information on which this decision must be based comes from the *break* and *dismiss* bits of the instruction word currently in use and from the sequence selector. Fig. 2 is a detailed drawing of one stage of the sequence selector. All stages, except that with the highest-priority, are identical. The lowest-priority stage returns the final three control signals to the control element.

Each stage of the sequence selector retains two pieces of information concerning its associated sequence. One flip-flop (ss j.1) remembers whether or not the sequence is selected (*i.e.*, whether or not it is receiving attention). The priority signal (labeled *no higher priority sequence requests attention*) passes from higher to lower priority stages until it encounters a stage which requests, but is not receiving attention. Such a stage is said to have priority at the moment, and its output to the program-counter-number coder prepares the number of the new program counter in anticipation of a sequence change.

The process of changing sequences involves storing the program counter for the old sequence and obtaining the counter for the new. Actually, to speed up the overall process, the new program counter is obtained first, so that it may be used while the old is being stored. Using the paths shown in Fig. 1, the new program counter number is placed in the j bits of the N register. The new program counter is then obtained from the X memory and interchanged with the old program counter contents which have been in the P register.[1] The K register, which has been holding the old program counter number since the last sequence change, is now interchanged with the j bits, and the old counter is stored at the proper location in the X memory. The state of the sequence selector is changed, to conform to the change of sequence, by sending a *select new sequence* command from Control. This command clears the ss j.2 flip-flop in the old-sequence stage and sets the ss j.2 flip-flop to a ONE in the new-sequence stage.[2]

Interpretation of the Break Bit

The programmer uses the *break* bit of an instruction word to indicate whether or not change to a higher priority sequence may occur at the completion of the instruction. The fact that a programmer permits a *break* does not mean that the sequence has completed its current task, but merely that no harm will be done if a change to some higher-priority sequence is made. *Breaks* should be permitted at every opportunity if a number of in-out devices are operating. The sort of situation in which a *break* cannot be permitted occurs when the E register is left containing information which the program requires at a later step. If a change occurred in this case, the contents of the E register would be destroyed and lost to the program.

[1] The P register is shown in Fig. 4 of Frankovich and Peterson, this issue, p. 148.

[2] The relative timing of the central computer actions during the change process is shown in Fig. 6(d) of Frankovich and Peterson, this issue, p. 150.

When a *break* is permitted by the current instruction, a sequence change will actually take place only if some higher-priority sequence requests attention. A signal from the sequence selector to the control element provides this information (Fig. 2). When a *break* type of sequence change is made, the ss j.1 flip-flop in the sequence selector remains unchanged, and the sequence which was abandoned in favor of one of a higher-priority continues to request attention.

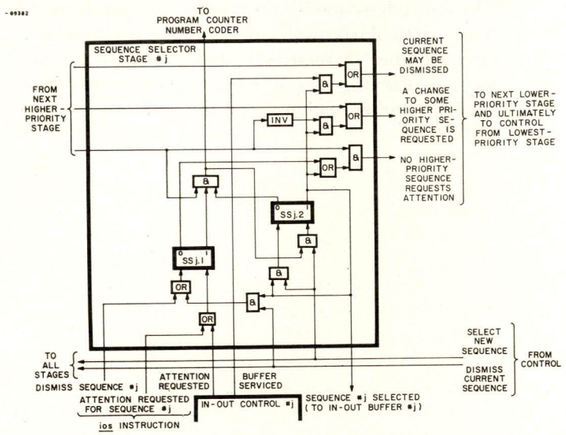


Fig. 2—Block diagram of TX-2 sequence selector stage.

Interpretation of the Dismiss Bit

The *dismiss* bit is used by the programmer to indicate that the sequence presently in use has completed its task. To provide synchronization in the in-out system, *dismiss* bits must be programmed between attention requests from the in-out devices. In this case, the *dismiss* operation guarantees that the computer will wait for the next signal from the in-out device before proceeding with the associated program sequence.

The *dismiss* bit is also used to accomplish the halt function in TX-2. A multiple-sequence computer *halts* when all sequences have been dismissed and all in-out units turned off. The priority signal from the sequence selector to the control element provides the information as to whether or not any sequence in the machine requests attention. When none request attention, the control stops all activity in the machine as soon as a *dismiss* bit appears on an instruction in the sequence being used. Activity is resumed in the machine as soon as some in-out device or push button requests attention.

The sequence change which results from a *dismiss* bit is identical with that resulting from a *break* except that a *dismiss current sequence* command accompanies the *select new sequence* command from Control to the Sequence Selector (Fig. 2).

Starting a Multiple-Sequence Computer

In a single-sequence computer the starting process involves resetting the program counter to some arbitrary value and starting the control. In a multiple-sequence computer, the program counter for a particular sequence must be reset and the sequence started. In TX-2 a special sequence (number 0) has the highest priority and is used to facilitate starting. This sequence has the special feature that its program counter always starts at an initial memory location specified by a set of toggle switches. Attention for the sequence is requested by pushing a button on the console. By executing a short program stored in the toggle-switch registers of the V memory, this sequence can start (or stop) any other sequence in the machine. The starting process for an arbitrary sequence involves resetting its program counter by means of an *ldx* (load index register) instruction, and starting its sequence with an *ios* instruction.

The Arithmetic Element in Multiple-Sequence Operation

While efficient operation requires that the bulk of arithmetic operations be carried out in a main sequence, the arithmetic element in TX-2 is available to all sequences. Since once a change has been made to a higher-priority sequence, control cannot return to a lower-priority sequence until the higher-priority one has been dismissed, a simple rule allows the arithmetic element to be used in any sequence without confusion. If, whenever a higher-priority sequence requires the arithmetic element, it stores the contents of any registers it will need (A, B, C, D, or F) and reloads them before dismissing, all lower-priority sequences will find the registers as they left them. This storing and loading operation requires time and, therefore, lowers the total data-handling capacity, but the flexibility obtained may well be worth the loss in capacity.

The step-counter class of arithmetic element instructions is a special problem. These instructions can require many microseconds to complete, and while TX-2 is designed to allow in-out and program element instructions to take place while the arithmetic element is busy, the case can arise in which an arithmetic element instruction (load, store, etc.) appears before the AE is finished with a step-counter class instruction. The machine would normally wait in an inactive state until the operation is complete, but since there is a chance that some higher-priority sequence may request attention in the interim and have instructions which can be carried out, provision is made to keep trying changes to higher-priority sequences as they request attention. The machine thus waits in an inactive state only when no higher-priority sequences have instructions which can be performed. This provision allows the programmer to ignore the arithmetic element in considerations of peak- and average-peak rate calculations when he desires to operate a maximum number of in-out devices.

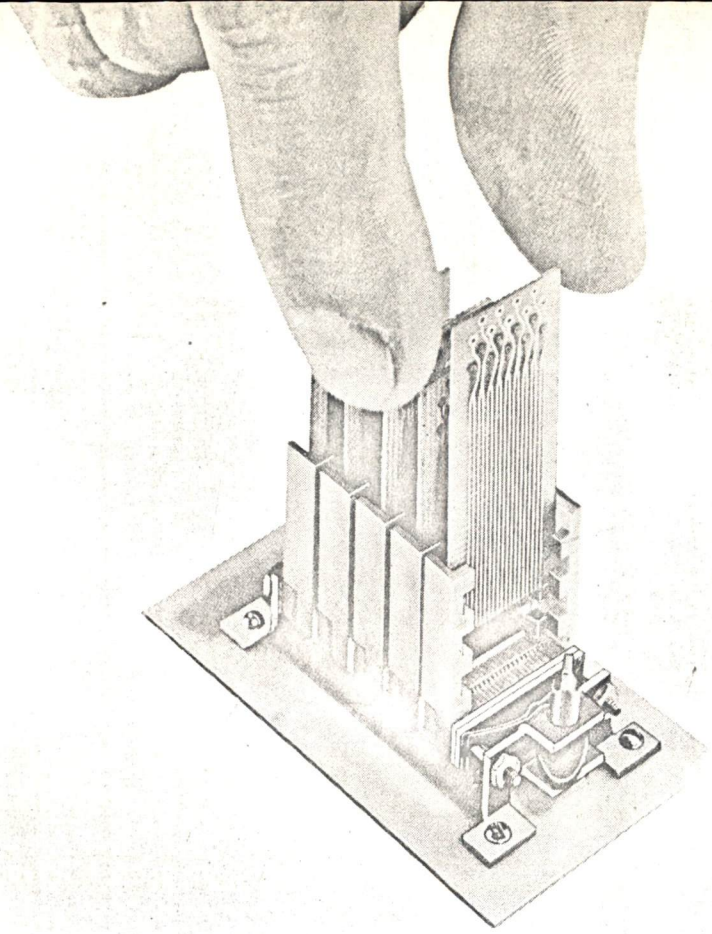
Conclusion

Multiple-sequence operation of input-output devices, as realized in TX-2, has a number of significant characteristics. Among them are:

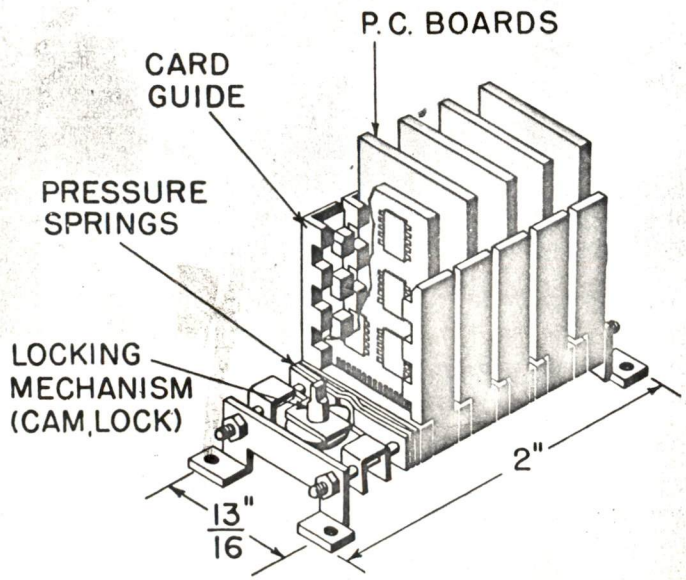
1) A number of in-out devices may be operated concurrently with a minimum of buffering storage.

2) Machine time is used efficiently, since no time need be lost waiting for input-output devices to complete their operation. Other machine activity may proceed meanwhile.
3) Each input-output device may be treated separately for programming purposes. Efficiency of operation is obtained automatically when several separately programmed devices are operated simultaneously, although average- and peak-rate limitations must be considered.
4) Maximum flexibility in programming for input-output devices is obtained. The full power of the central machine may be used by each input-output sequence if desired. Routines for each device may be as long or as short as the particular situation requires.
5) The modular organization of the input-output equipment simplifies additions and modifications to the complement of in-out devices.
6) The organization of buffering storage allows the amount and kind of such storage to be tailored to the needs of the individual devices and the data-handling requirements to be met by the system.
7) The multiple-sequence program technique appears to be particularly well suited to the operation of a large number of relatively slow input-output devices of varying characteristics, as opposed to a smaller number of high-speed devices.

External Clamping Assembly Offers Many Advantages in High Density Packaging



Typical PRESKAM five-board assembly with cam lock in open position. Board is being dropped into place between card guides. Actual contacts can be seen directly below the PC board. Contacts are on 0.025" centers.



Drawing illustrates the PRESKAM system for five PC boards. Flat circuit packs on boards are shown for size comparison. The five boards are released simultaneously when the cam lock is turned 90° in either direction. The card guides hold the boards in position when pressure is released. Cards have connector density of 80 connections per lineal inch (0.025" centers)

An entirely new concept in electron interconnection devices has been announced by Cinch Mfg. Co. of Chicago, Ill. Called PRESKAM, it permits substantially increased contact density and reliability. The PRESKAM system was designed primarily for use in the interconnection of microelectronic circuit boards. However, the principle involved is applicable to any size or thickness of circuit board.

As shown in the accompanying photos, the system utilizes a cam-actuated insulating bar to provide contact with the printed circuit boards by external clamping—after the board is inserted into the device. A resilient material recessed in the insulating bar, behind the connector contacts, applies pressure over the entire length of the contact surface.

Contact forces in conventional PC board connectors are said to be limited by insertion and withdrawal forces, as well as by the amount of contact "drag" on the board terminals. According to Cinch, these limitations do not exist in their PRESKAM approach, because the contact pressure is not applied until after the board has been inserted in the connector. Thus, higher contact pressures can be achieved with a resultant increase in connection reliability.

Reliability of PRESKAM is also increased by a substantial increase in the mated contact area. Conventional spring-type contacts provide electrical contact only over a portion of their length, whereas the PRESKAM contact surface is parallel to the circuit board and electrical contact is made over its entire length. PRESKAM offers, according to the company, generally about 50-70% more surface area in the same contact length than is provided by a conventional spring type connector.

The PRESKAM assembly is currently being produced with 0.025" centers. It is anticipated that even closer spacings can be developed, should they be required with the

Interconnecting Micro PC Boards

PRESKAM approach, contact spacing is no longer limited by connector design, but by printed circuit board technology.

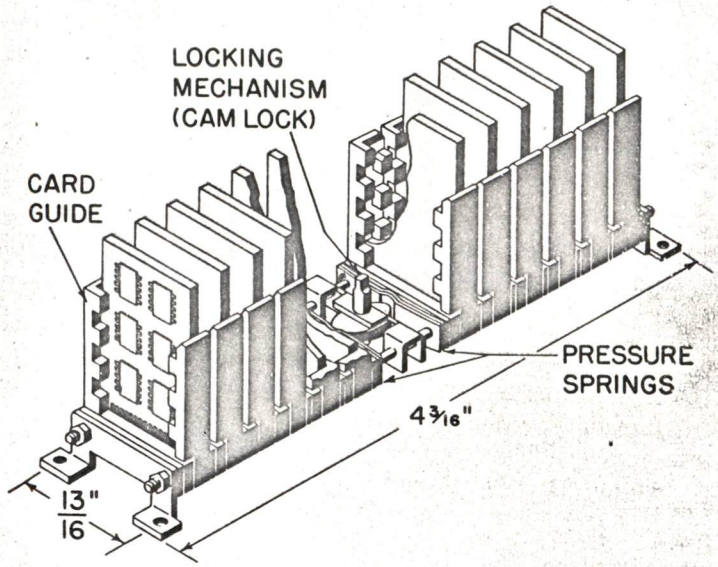
Standard, proved materials are used in the PRESKAM so that environmental reliability can be readily established. Contacts are of gold-plated, copper-base alloy. Insulating members are currently made of Lexan, and neoprene is used for the pressure materials. Other materials can be utilized to meet specific environmental conditions.

According to Roy Witte, vice president of engineering at Cinch, the new technique offers many advantages over conventional connectors in high density packaging applications. Among these are: greatly increased resistance to shock and vibration due to the increased contact forces of the PRESKAM; use of glass or ceramic substrates, since insertion and withdrawal forces are no longer a factor. High densities can be achieved in equipment packaging design. Not only does PRESKAM permit closer spacing of the contacts, but also closer spacing of the boards. The boards have been spaced with as little as 0.100″ clearance between them.

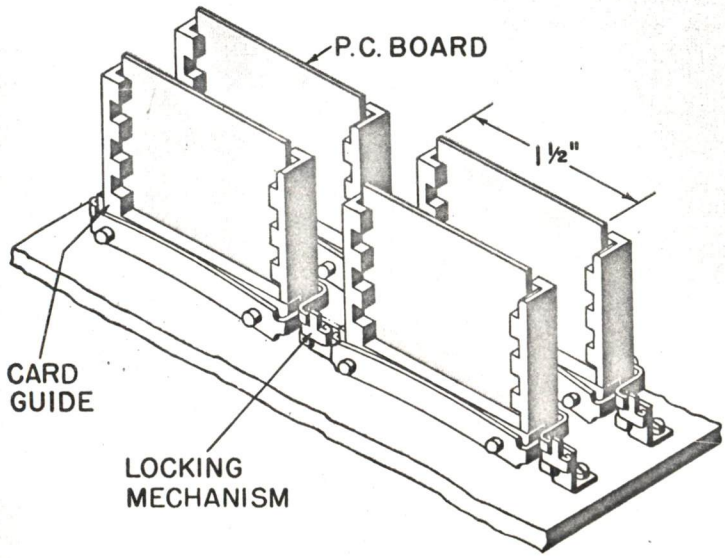
Equipment maintenance is simplified and a reduction of component damage in the removal and replacement of circuit boards can be anticipated. A board can be lifted out and a new board dropped into place without the pressures normally involved in pulling or pushing the board.

PRESKAM can be used singly or in multiple stacks controlled by a single locking device. Assemblies containing as many as ten connectors have been designed. The company reports that individual connectors or multiple stacks show no evidence of deterioration when subjected to tests for contact resistance, moisture resistance, temperature cycling, vibration, and salt spray per MIL-STD-202B.

Circle No. 110 on Inquiry Card



PRESKAM system for 12 PC boards. System illustrated provides 432 connections.



Drawing illustrates flexibility of the PRESKAM system. Packaging approach uses four single-board assemblies. Assemblies can be controlled by individual clamping devices as illustrated, or by a single clamp that would control all four (or more) simultaneously. Board size is approximately 1¼″ x 1½″. Each board has 104 connections to mother board.

Nanoamp Leakage Measurements with the Type 575

The following article, by Pete Sylvan of GE Semiconductor, Syracuse, describes a simple, efficient method of making nanoamp measurements of reverse leakage currents in diodes, using the 575.

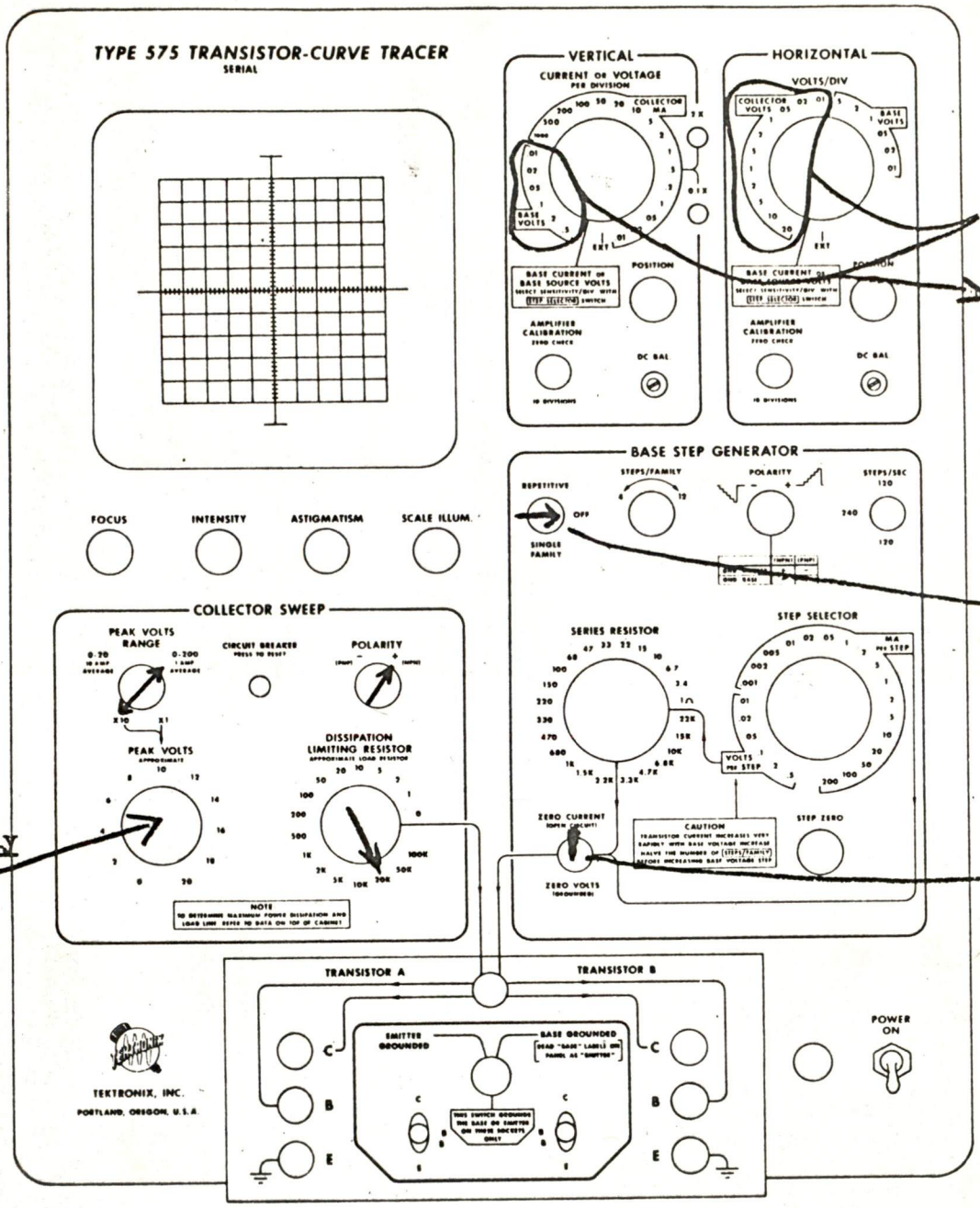
Briefly, Pete filters the collector supply so that DC is available from the collector terminal. By using 20 k series resistance in the collector supply and tying a 1 mfd capacitor from the collector terminal to ground, the normal 120-cycle collector sweep is converted to essentially DC. The base step generator is turned off and disconnected from the base terminal by the method described.. ...or by holding the lever switch in the UP position with a rubber band from a knob above. A 10 meg, 1% resistor is tied from the base terminal to ground and used as a current-sampling resistor. The diode can now be inserted between the collector and base terminal, and its leakage current sampled and read out on the CRT.

Settings and additional circuitry are shown in the illustration that follows Pete's presentation, below:

"The introduction of passivated silicon diodes and transistors has made necessary the measurement of leakage currents in the nanoampere (10^{-9} ampere) range. Present techniques require the use of a special test set up using a regulated power supply and a sensitive current meter. This ties up expensive test equipment for use only in leakage current tests.

"A technique and test jig has been developed for measuring the leakage currents of transistors and diodes in the nanoampere range using a standard Tektronix transistor curve tracer. This measurement technique requires no auxiliary equipment, provides a quick, accurate measurement of leakage currents down to 0.2 nanoamperes with the convenience of a X-Y presentation of the relationship between leakage current and voltage. The jig, which plugs into the binding posts on the front of the scope, is shown schematically in Figure 1. The jig construction is not critical but should provide shielding for the components, binding posts and device under test so as to minimize 60 cps pickup. The base source switch, which must be kept in the open position, is a spring return switch, but can be converted to a locking type switch by bending the leaf spring with pliers. The 1 μfd capacitor together with the collector limiting resistor provides a d-c voltage across the diode while the collector limiting resistor protects the rectifiers inside the scope against the high reverse voltages.

"To make a measurement, all controls are set as shown below and the zero setting is made with no diode or transistor in the test socket (the regular zero adjust control is not accurate for this test). The diode or transistor is inserted in the test clips, and the collector voltage control is moved slowly up and down to sweep the spot over the desired voltage range."


TYPE 575 TRANSISTOR-CURVE TRACER
SERIAL
VERTICAL
CURRENT OR VOLTAGE PER DIVISION
HORIZONTAL
VOLTS/DIV
Use any of circled settings.
.01 v/div = 1na/div
.02 v/div = 2na/div
.05 v/div = 5na/div
etc.
FOCUS
INTENSITY
ASTIGMATISM
SCALE ILLUM.
BASE STEP GENERATOR
REPETITIVE
SINGLE FAMILY
STEPS/FAMILY
POLARITY
STEPS/SEC
Off
COLLECTOR SWEEP
PEAK VOLTS RANGE
CIRCUIT BREAKER
POLARITY
PEAK VOLTS
DISSIPATION LIMITING RESISTOR
SERIES RESISTOR
STEP SELECTOR
ZERO CURRENT
ZERO VOLTS
CAUTION
STEP ZERO
Rotate slowly to make test.
Open (hold with rubber band)
TRANSISTOR A
TRANSISTOR B
EMITTER GROUNDED
BASE GROUNDED
POWER ON
TEKTRONIX, INC.
PORTLAND, OREGON, U.S.A.

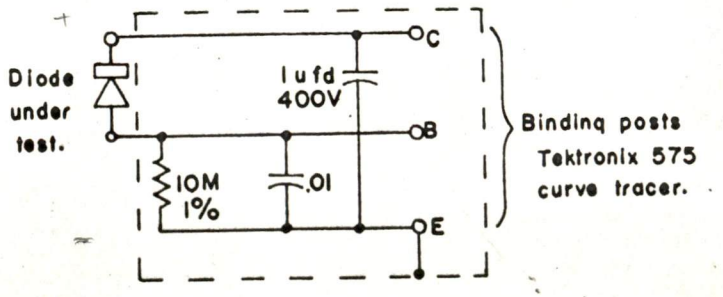

Diode under test.
1 ufd 400V
10M 1%
.01
C
B
E
Binding posts Tektronix 575 curve tracer.

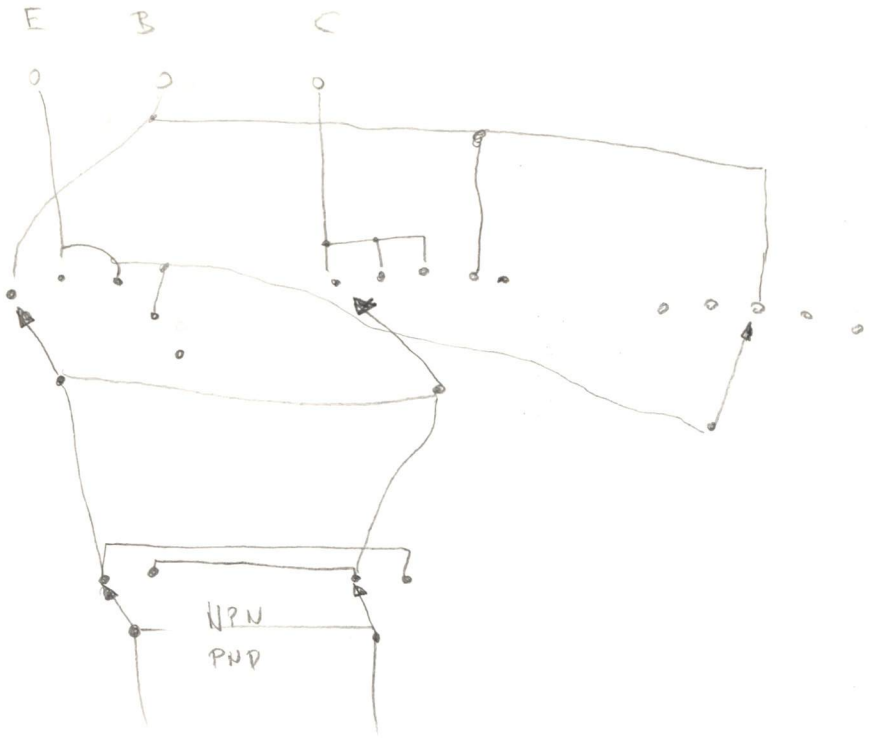
Figure 1

CBO

EBO

CES

EBO



Nanoamp Leakage Measurements with the Type 575

The following article, by Pete Sylvan of GE Semiconductor, Syracuse, describes a simple, efficient method of making nanoamp measurements of reverse leakage currents in diodes, using the 575.

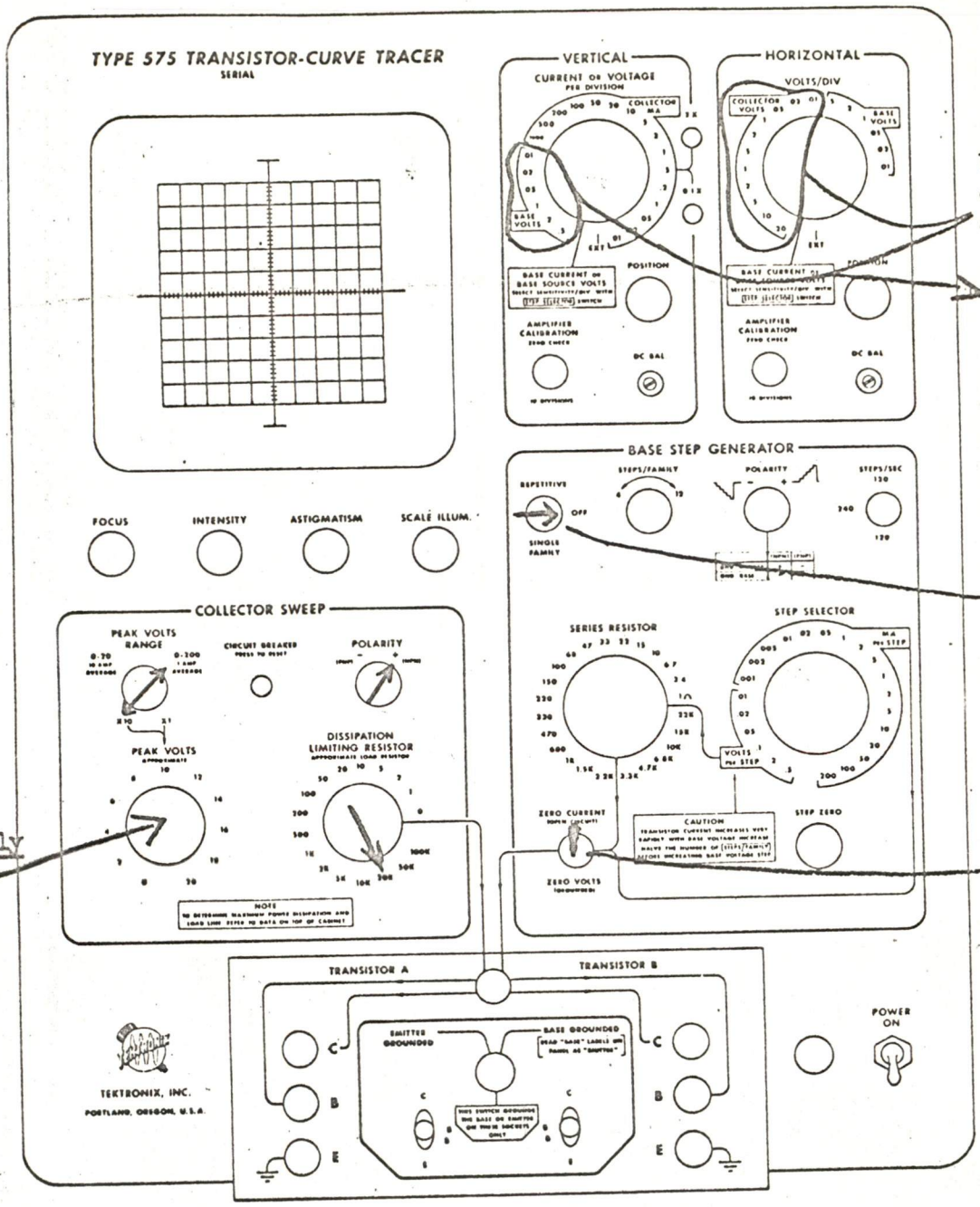
Briefly, Pete filters the collector supply so that DC is available from the collector terminal. By using 20 k series resistance in the collector supply and tying a 1 mfd capacitor from the collector terminal to ground, the normal 120-cycle collector sweep is converted to essentially DC. The base step generator is turned off and disconnected from the base terminal by the method described.. ...or by holding the lever switch in the UP position with a rubber band from a knob above. A 10 meg, 1% resistor is tied from the base terminal to ground and used as a current-sampling resistor. The diode can now be inserted between the collector and base terminal, and its leakage current sampled and read out on the CRT.

Settings and additional circuitry are shown in the illustration that follows Pete's presentation, below:

"The introduction of passivated silicon diodes and transistors has made necessary the measurement of leakage currents in the nanoampere (10^{-9} ampere) range. Present techniques require the use of a special test set up using a regulated power supply and a sensitive current meter. This ties up expensive test equipment for use only in leakage current tests.

"A technique and test jig has been developed for measuring the leakage currents of transistors and diodes in the nanoampere range using a standard Tektronix transistor curve tracer. This measurement technique requires no auxiliary equipment, provides a quick, accurate measurement of leakage currents down to 0.2 nanoamperes with the convenience of a X-Y presentation of the relationship between leakage current and voltage. The jig, which plugs into the binding posts on the front of the scope, is shown schematically in Figure 1. The jig construction is not critical but should provide shielding for the components, binding posts and device under test so as to minimize 60 cps pickup. The base source switch, which must be kept in the open position, is a spring return switch, but can be converted to a locking type switch by bending the leaf spring with pliers. The 1 μfd capacitor together with the collector limiting resistor provides a d-c voltage across the diode while the collector limiting resistor protects the rectifiers inside the scope against the high reverse voltages.

"To make a measurement, all controls are set as shown below and the zero setting is made with no diode or transistor in the test socket (the regular zero adjust control is not accurate for this test). The diode or transistor is inserted in the test clips, and the collector voltage control is moved slowly up and down to sweep the spot over the desired voltage range."



Use any of circled settings.

.01 v/div = 1na/div
.02 v/div = 2na/div
.05 v/div = 3na/div

etc.

Off

Rotate slowly to make test.

Open (hold with rubber band)

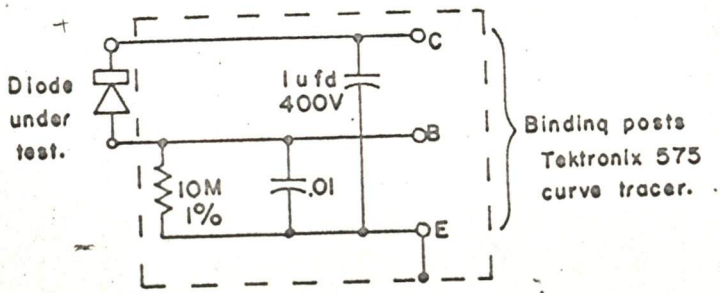


Figure 1

Nanoamp Leakage Measurements with the Type 575

The following article, by Pete Sylvan of GE Semiconductor, Syracuse, describes a simple, efficient method of making nanoamp measurements of reverse leakage currents in diodes, using the 575.

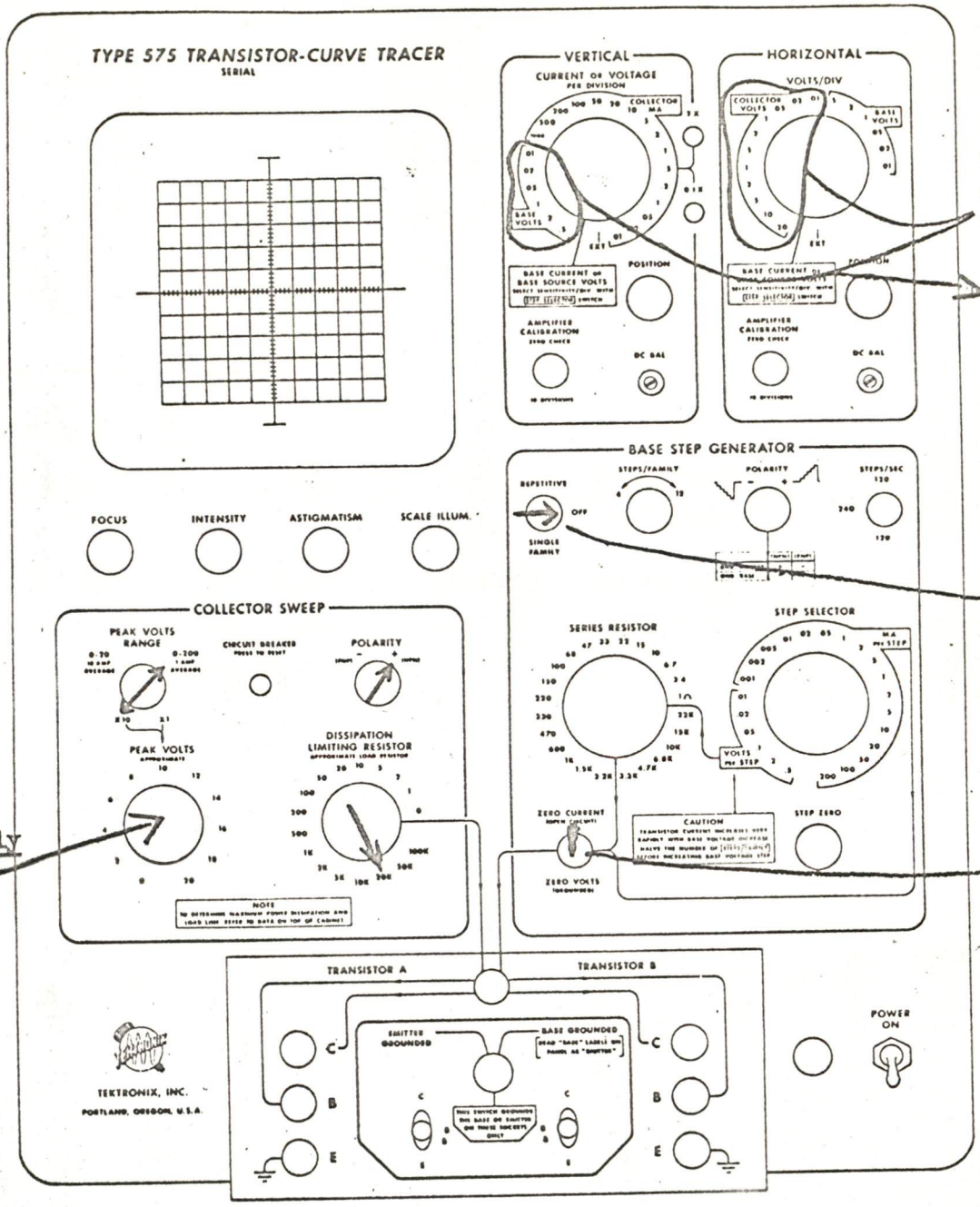
Briefly, Pete filters the collector supply so that DC is available from the collector terminal. By using 20 k series resistance in the collector supply and tying a 1 mfd capacitor from the collector terminal to ground, the normal 120-cycle collector sweep is converted to essentially DC. The base step generator is turned off and disconnected from the base terminal by the method described.. ...or by holding the lever switch in the UP position with a rubber band from a knob above. A 10 meg, 1% resistor is tied from the base terminal to ground and used as a current-sampling resistor. The diode can now be inserted between the collector and base terminal, and its leakage current sampled and read out on the CRT.

Settings and additional circuitry are shown in the illustration that follows Pete's presentation, below:

"The introduction of passivated silicon diodes and transistors has made necessary the measurement of leakage currents in the nanoampere (10^{-9} ampere) range. Present techniques require the use of a special test set up using a regulated power supply and a sensitive current meter. This ties up expensive test equipment for use only in leakage current tests.

"A technique and test jig has been developed for measuring the leakage currents of transistors and diodes in the nanoampere range using a standard Tektronix transistor curve tracer. This measurement technique requires no auxiliary equipment, provides a quick, accurate measurement of leakage currents down to 0.2 nanoamperes with the convenience of a X-Y presentation of the relationship between leakage current and voltage. The jig, which plugs into the binding posts on the front of the scope, is shown schematically in Figure 1. The jig construction is not critical but should provide shielding for the components, binding posts and device under test so as to minimize 60 cps pickup. The base source switch, which must be kept in the open position, is a spring return switch, but can be converted to a locking type switch by bending the leaf spring with pliers. The 1 µfd capacitor together with the collector limiting resistor provides a d-c voltage across the diode while the collector limiting resistor protects the rectifiers inside the scope against the high reverse voltages.

"To make a measurement, all controls are set as shown below and the zero setting is made with no diode or transistor in the test socket (the regular zero adjust control is not accurate for this test). The diode or transistor is inserted in the test clips, and the collector voltage control is moved slowly up and down to sweep the spot over the desired voltage range."


TYPE 575 TRANSISTOR-CURVE TRACER
SERIAL
VERTICAL
CURRENT OR VOLTAGE PER DIVISION
HORIZONTAL
VOLTS/DIV
POSITION
AMPLIFIED CALIBRATION
DC BAL
FOCUS
INTENSITY
ASTIGMATISM
SCALE ILLUM.
BASE STEP GENERATOR
REPETITIVE
SINGLE FAMILY
STEPS/FAMILY
POLARITY
STEPS/SEC
COLLECTOR SWEEP
PEAK VOLTS RANGE
CIRCUIT BREAKER
POLARITY
PEAK VOLTS
DISSIPATION LIMITING RESISTOR
SERIES RESISTOR
STEP SELECTOR
ZERO CURRENT
ZERO VOLTS
CAUTION
STEP ZERO
TRANSISTOR A
TRANSISTOR B
EMITTER GROUNDED
BASE GROUNDED
POWER ON
TEKTRONIX, INC.
PORTLAND, OREGON, U.S.A.
Use any of circled settings.
.01 v/div = 1ma/div
.02 v/div = 2ma/div
.05 v/div = 5ma/div
etc.
Off
Rotate slowly to make test.
Open (hold with rubber band)

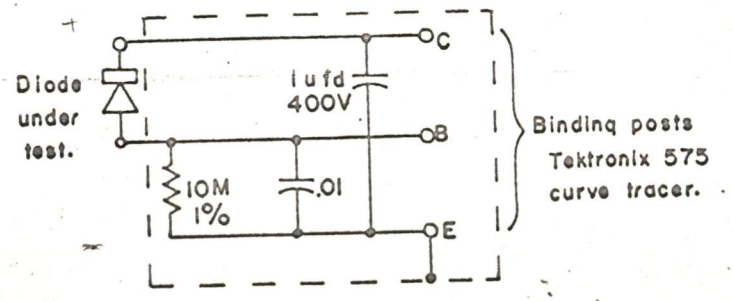

Diode under test.
1 ufd 400V
10M 1%
.01
C
B
E
Binding posts Tektronix 575 curve tracer.

Figure 1

Infinite Impedance From a Constant-Current Source

R. W. Thorpe
Design Specialist
General Dynamics/Pomona, Pomona, Calif.

The trick is to sample the voltage across the load as well as the current through it. The positive feedback developed in this manner can raise the output impedance to infinity for a given range of load conditions.

HOW CAN a conventional constant-current supply be modified for infinite output impedance? One answer is, by adding a regulating loop that senses and compensates for the circuit variable causing the departure from ideal behavior.

The culprit is the voltage developed across the load. This voltage increases in proportion to the size of the load; with large loads it can become big enough to divert some of the current through the supply's internal shunt resistance. If the presence of this voltage can be sensed and used to introduce a compensating current, the supply can be made to act as if it had infinite output impedance.

A circuit incorporating this load voltage compensation, in addition to the regular current-sensing control, is shown in Fig. 1. The circuit is simple enough, but its evolution will give designers an appreciation of the concepts involved.

The problems in designing a high-performance current source are illustrated by the differences between a perfect and a practical constant-current source. The perfect current source, Fig. 2a, is by definition one that delivers a specified current into any load to which it may be connected. The specified current may be a steady dc level or any desired function of time or other variable. The voltage across the load may be any value from plus infinity to minus infinity and will be determined solely by the current delivered and the value of the load impedance. The compliance of the source, defined as its output voltage capability, is, therefore, infinite.

The practical current source, on the other hand, only approximates a perfect source. In its linear range it may be described, as in Fig. 2b, as a perfect current source in parallel with a practical impedance. For short-circuited conditions, $Z_L = 0$, there is no problem. But as the load Z_L is increased, a load voltage V_L is developed, which diverts some of the current through the source impedance Z_L. Unless Z_s can be made very high with respect to the highest expected Z_L, the load current will not be constant. But if Z_s can be made infinite, none of the source current will be diverted through the source impedance and the load current will always equal the source current.

Conventional Circuit Is Base for Modifications

A simplified schematic of a conventional current regulator is shown in Fig. 3. Here, the load current I_L flows through a sensing resistor, R_2, to develop a voltage that is proportional to I_L. This voltage is compared to a reference voltage developed across a Zener diode, *D1*. Any difference is amplified by *Q1*, which acts as a differential amplifier. The amplified signal drives the regulator or "pass" transistor *Q2*, which controls the current delivered by battery E to the load. Resistor R_1 is needed only to provide current for the Zener diode.

If the load impedance decreases, the load current increases, along with the voltage across R_2, driving the emitter of *Q1* more positive with respect to its base and tending

to cut off *Q2*. This counteracts the original increase of I_L.

The regulator acts to maintain the load current at a value that makes $I_L R_2$ approximately equal to the reference voltage across the Zener diode. The output current can be changed through adjustment of R_2.

The performance of conventional current-source supplies is adequate for low and moderate load impedances, but higher load impedances tend to upset the regulating action.

The clue to the desired modifications is found in Fig. 2b. As long as $Z_L = 0$, all the source current flows through the load.

It is the voltage developed across a non-zero Z_L that causes the current to flow in Z_s, upsetting the current regulation. But if this voltage could be sampled and used to increase the current delivered by the "perfect" current source, this effect could be counteracted. If this increase could be made just equal to the current drawn by Z_s, there would be no change in the current through Z_L and the desired infinite Z_o will have been achieved.

The first of the series of modification to realize this objective is shown in Fig. 4a. Any voltage developed across the load is sampled at the arm of potentiometer R_4 and applied through *D1* to the base of *Q1*, tending to turn *Q1* further on. This in turn drives *Q2* further on, permitting more load current to flow. In effect, this modification provides a negative output impedance that is parallel with and cancels the positive output impedance of the unmodified source. The parallel combination can be made positive, negative or infinite by the setting of R_4.

Further inspection of Fig. 4a shows that the Zener current flows through the lower part of R_4, developing a dc voltage in series with the Zener voltage. Any changes in Zener current, such as would be caused by a variation in the dc supply voltage E, would change the effective reference voltage for the regulator and, correspondingly, the current delivered.

This difficulty can be circumvented by feeding the developed load voltage signal back to the amplifier base in parallel, rather than in series, with the reference diode. This technique is illustrated in Fig. 4b. Rheostat R_3 is varied to control the amount of feedback, and thus the regulator output impedance. Resistor R_5 limits the amount of feedback to protect the circuit.

The current through R_3 returns to the minus side of the dc supply by two paths. The first is through the low-impedance Zener diode and the second is through the relatively high-impedance path made up of R_2 and the base-emitter diode of *Q1*. Thus, it is apparent that only a small fraction of the feedback current reaches *Q1* and that this fraction is affected by the dynamic impedance of the Zener diode.

A more sensitive and more stable current injection scheme is shown in Fig. 4c. Resistor R_6 forces a larger fraction of the feedback current through R_3 into the transistor base and also stabilizes the total impedance into which the current flows. The effect of R_6 on the basic regulator performance is negligible. The value of R_3 now can be increased; there still will be enough feedback to produce the infinite impedance effect.

Positive Feedback May Cause Stability Problem

The two transistors in the conventional regulator shown in Fig. 3 were used as cascaded voltage amplifiers. If the negative side of the dc supply is considered as a reference point, it will be seen that *Q1* was operated in the common-base mode while *Q2* was operated in the common-emitter mode. A signal starting at the emitter of *Q1* would appear amplified without phase reversal at the collector of *Q1*, and from there would feed into the base of *Q2*. The load impedance on the collector of *Q1* is the base input impedance of transistor *Q2*.

The signal at the base of *Q2* appears amplified and inverted at the collector of *Q2*. The load for *Q2* is Z_L—the external load impedance—plus the parallel combination of R_2 and the input emitter impedance of *Q1*.

The signal at the collector of *Q2* is divided between Z_L and the combination of R_2 and *Q1* input impedance and appears at the emitter of *Q1* out of phase with the initiating signal. This feedback is, therefore, negative for frequencies at which extraneous phase shifts are negligible.

But with increasing frequencies, additional phase lags become more noticeable, and when the total lag reaches 360 degrees, the feedback becomes positive. If the loop gain is greater than unity when the lag reaches 360 degrees, self-oscillation occurs, and the advantages of the regulator are lost.

A series RC feedback from collector to base of *Q1*, as shown in Fig. 4d, helps combat this tendency by reducing the loop gain at higher frequencies without introducing additional phase shift. As the frequency increases from zero, the reactance of the capacitor C begins to shunt the collector load for *Q1* and the loop gain will decrease. Some phase lag will be introduced at this point.

As the frequency increases further, the reactance of C becomes negligible with respect to R_7 and the load for $Q1$ is nearly the value of R_7. Because of this resistive load, the extra phase shift disappears and because of the small size of R_7, the loop over-all gain is materially reduced. The bandwidth over which the constant current effect is produced depends on the bandwidth of the amplifier loop. This stabilization is at the expense of the supply's bandwidth.

The modification for infinite output impedance also may cause instability. The new loop that senses the presence of voltage across the load introduces a small amount of positive feedback. When the supply is exactly adjusted for infinite output impedance, it is on the verge of instability for the case of infinite load impedance. But the finite load impedance that would be encountered in any real application reduces the gain of this secondary loop to less than unity, ensuring stability.

The circuit of Fig. 1 uses the modified parallel-injection scheme of Fig. 4c. The transistors used are complementary germanium types chosen as representative of a broad class of low-cost, general-purpose and switching devices. Since $Q1$'s collector must supply the base current for $Q2$, $Q1$ must be able to operate with collector currents on the order of 10-100 μa.

The second Zener, $D2$, protects the regulator in case of an open-circuit load. Without $D2$, an open circuit-load would result in an excessive forward base-emitter bias on $Q1$. In that case, the $Q2$'s base-emitter diode, the saturated $Q1$ and R_2 would be in series across the 12-v supply. The large resultant current surge could easily damage the circuit. Thus, it is important to have $D2$ available to carry the entire specified load current in the event of an open-circuit load.

Experimental Setups to Help Designer Evaluate and Adjust Supply

The dc output impedance of the regulator can be measured as shown in Fig. 5a. Meter $M1$ reads the total load current, which for this circuit falls between 1 and 5 ma. Meter $M2$ registers the difference between the load current I_L and the test loop current I_2. If this difference is made approximately zero through adjustment of R_8, $M2$ may be a very sensitive meter capable of showing a change of a fraction of a microampere. Any change indicated by $M2$ results from a change in I_L, or:

$$\Delta I_L = \Delta I_{M2} \frac{R_8 + R_M}{R_8}$$

where:

ΔI_L = change in load current.
ΔI_{M2} = change in current through $M2$.
R_M = resistance of $M2$.

If the switch is moved to change the load voltage by an amount $\Delta E_L = E_2$, the regulator output impedande is given by:

$$Z_o = \frac{E_2}{\Delta I_L} = \frac{E_2 R_8}{\Delta I_{M2}(R_8 + R_M)}$$

Measurements on the unmodified source—that is, the circuit of Fig. 1 with $R_6 = 0$, $R_3 = R_5 = \infty$—will show an output impedance of 95 K at a load current of 5 ma, increasing nearly linearly to 400 K at 1 ma. The load-current output can be reduced by increasing R_2. This also increases the loop gain and the output impedance.

Feedback in the regulator can be adjusted so that ΔI_{M2} is zero for non-zero E_2, yielding infinite output impedance at dc. A further increase in the amount of feedback in the regulator results in an increase of I_L for an increase of E_2, demonstrating that a negative output impedance can exist.

It will be noted that there is a gradual, rather than immediate, change in I_L on the order of 1 μamp soon after the test circuit switch is thrown. The voltage drop across $Q2$ changes by an amount very nearly equal to E_2 and the internal dissipation changes accordingly.

The change in dissipation alters the device temperature, which in turn causes a slight change in the operating point. This accounts for the slow response observed in I_L.

Setup to Show Regulation As Function of Frequency

The output impedance as a function of frequency may be measured as shown in Fig. 5b. The current "i" is a small ac component superimposed on the dc load current I_L. Its value is given by:

$$i = \frac{e_2}{R_9}$$

and the output impedance is then:

$$Z_o = \frac{e_i}{i} - R_9 = R_9\left(\frac{e_1}{e_2} - 1\right)$$

As Z_o becomes large, e_1 far exceeds e_2 and this expression reduces to:

$$Z_o = R_9 \frac{e_1}{e_2}$$

As Z_o becomes infinite, i and e_2 become zero; e_2 being a particularly good indication for the adjustment for infinite Z_o.

If e_2 is read on a sensitive oscilloscope synchronized to the audio oscillator, it will be seen that the phase of e_2 reverses as the null is traversed. This shows that the output impedance of the current regulator actually becomes negative.

As a practical matter in the setup of this test circuit, the transformer used for signal injection must be able to operate without saturation over the frequency range of interest. It must be remembered that the transformer will at times be carrying dc as well as ac signals. In presenting the results of tests on this sort of supply, it is convenient to use the reciprocal of output impedance or the admittance. Fig. 6 shows the output admittance of the regulator as a function of frequency for any load current in the range of 1 to 5 ma.

If it is possible to omit the stabilizing network, the bandwidth can be extended, the amount of this increase being shown in the graph, Fig. 6.

The high output impedance at the higher frequencies can be maintained by using an inductor in series with the output lead. This is, of course, analogous to the use of a shunt capacitor in the more common voltage-regulated supplies. The disadvantage of this technique is that some of the output voltage capability of the current regulator may be lost across the inductor, thus reducing the amount available for the load. ■ ■

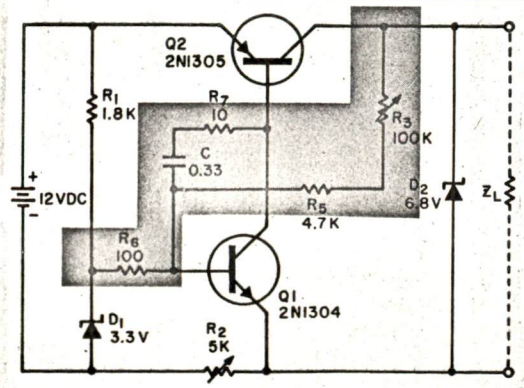


Fig. 1. The modified constant-current source achieves infinite output impedance by the positive feedback loop (shaded), which senses any voltage developed by the load.

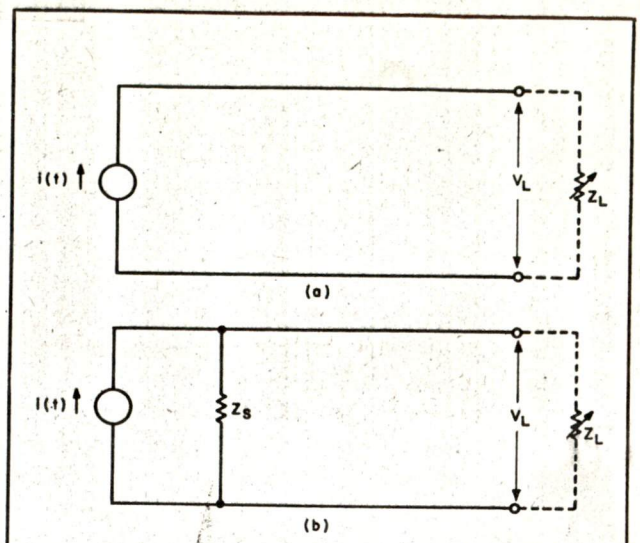


Fig. 2. Review of symbols for perfect (*a*) and practical (*b*) current sources: Though the ideal current source is by definition able to deliver a specific value of current regardless of the nature of the load, practical sources have internal shunts, Z_s, that make them sensitive to the nature of the load and the frequency of the source or any system disturbance. The ideal current source is shown as a function of time to indicate that the analysis is not necessarily limited to dc.

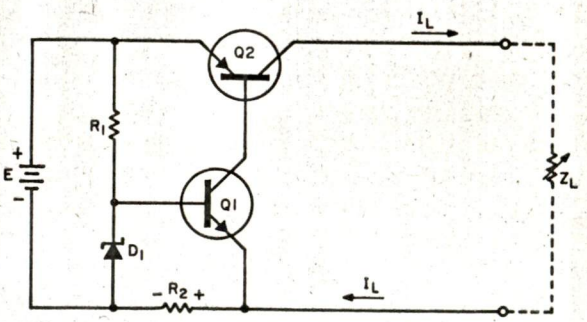


Fig. 3. Conventional constant-current sources depend entirely on sensing the load current. Changes in the load current alter the voltage across R_2 and the base-emitter voltage into *Q1*, which in turn alters the amount of current that Q2 passes.

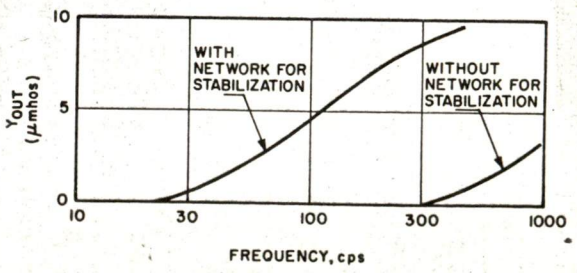
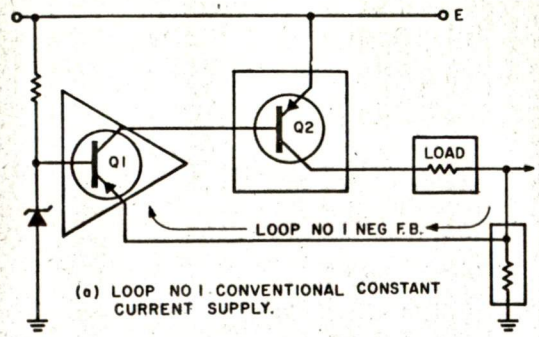
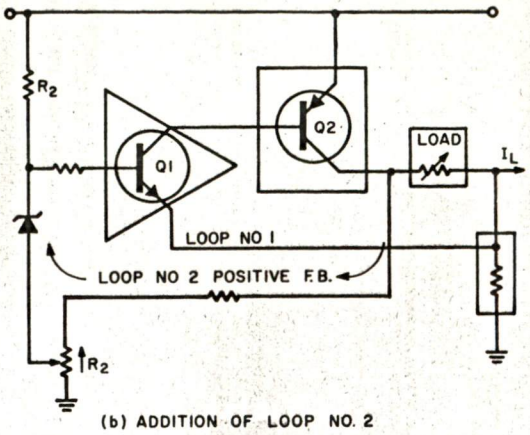


Fig. 6. Curves of admittance against frequency show the effect of the stabilization loop. It is possible to dispense with the stabilization under certain load-impedance conditions.

Purpose of Three Loops

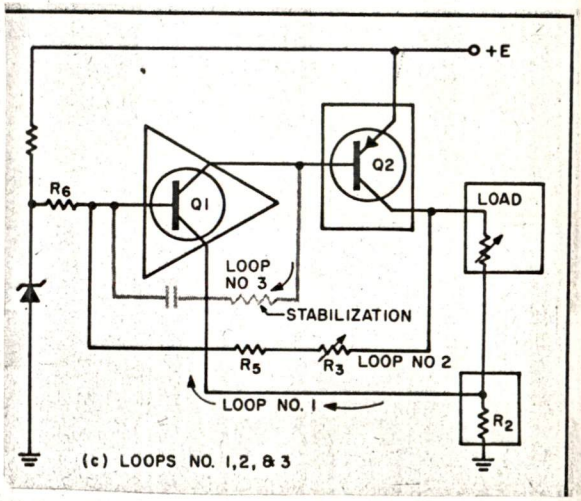


(a) LOOP NO 1 CONVENTIONAL CONSTANT CURRENT SUPPLY.



(b) ADDITION OF LOOP NO. 2

When the constant current circuit is redrawn in the classical servo format, the action of each of the three loops in the circuit is made more obvious. Loop No. 1 is the conventional current-sensing loop. Loop No. 2 is the positive feedback loop that senses the presence of voltage across the load, and loop No. 3 is the stabilizing loop that cuts down the gain at higher frequencies.



(c) LOOPS NO. 1, 2, & 3

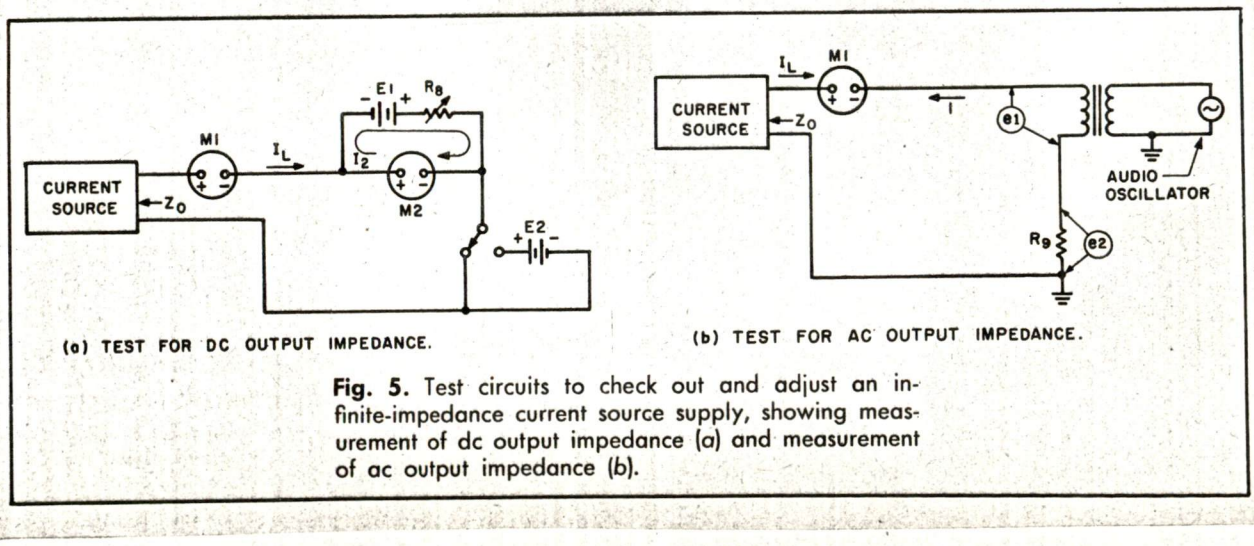


Fig. 5. Test circuits to check out and adjust an infinite-impedance current source supply, showing measurement of dc output impedance (a) and measurement of ac output impedance (b).

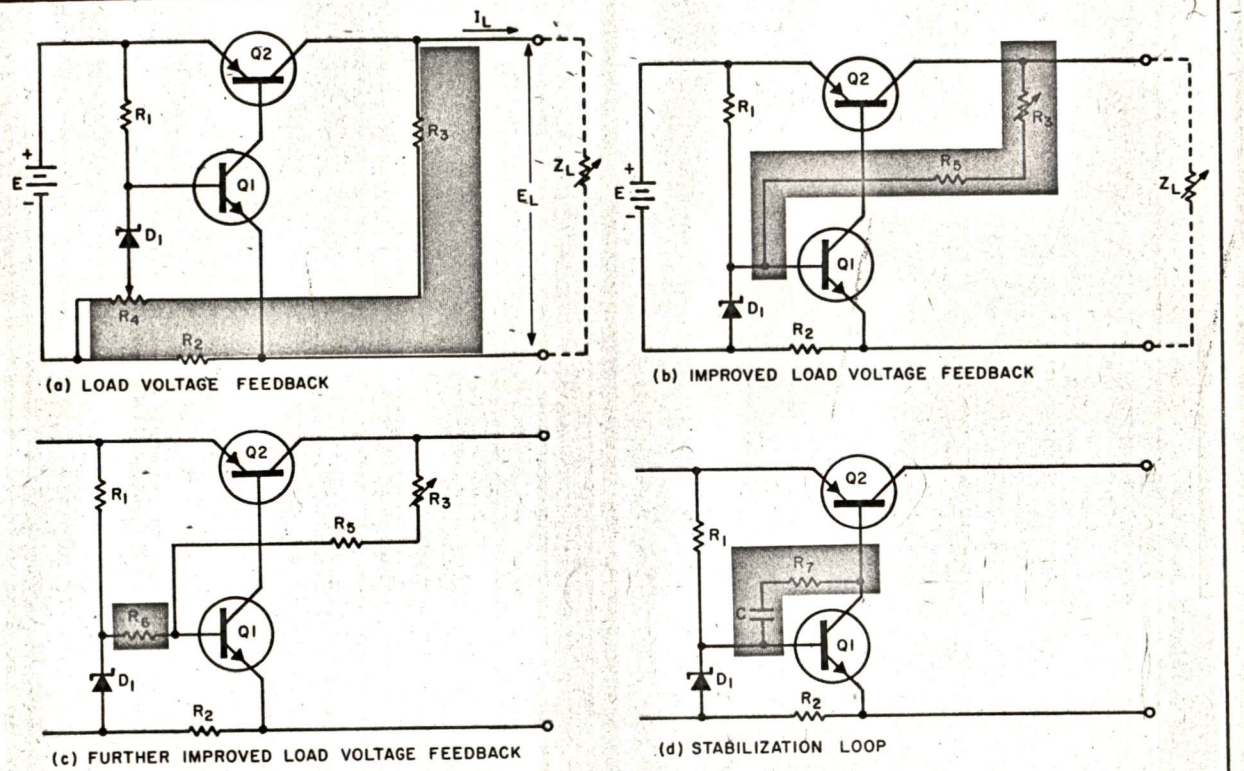


Fig. 4. The four stages of circuit development leading to the circuit in Fig. 1. (a) Voltage across the load is sensed and used to change the reference set by Zener *D1*. (b) The positive feedback can be introduced directly into the base of the transistor. (c) Resistor R_6 directs more of the feedback to the transistor. (d) Stabilization is added to offset the phase lags.

Infinite Impedance From a Constant-Current Source

R. W. Thorpe
Design Specialist
General Dynamics/Pomona, Pomona, Calif.

The trick is to sample the voltage across the load as well as the current through it. The positive feedback developed in this manner can raise the output impedance to infinity for a given range of load conditions.

HOW CAN a conventional constant-current supply be modified for infinite output impedance? One answer is, by adding a regulating loop that senses and compensates for the circuit variable causing the departure from ideal behavior.

The culprit is the voltage developed across the load. This voltage increases in proportion to the size of the load; with large loads it can become big enough to divert some of the current through the supply's internal shunt resistance. If the presence of this voltage can be sensed and used to introduce a compensating current, the supply can be made to act as if it had infinite output impedance.

A circuit incorporating this load voltage compensation, in addition to the regular current-sensing control, is shown in Fig. 1. The circuit is simple enough, but its evolution will give designers an appreciation of the concepts involved.

The problems in designing a high-performance current source are illustrated by the differences between a perfect and a practical constant-current source. The perfect current source, Fig. 2a, is by definition one that delivers a specified current into any load to which it may be connected. The specified current may be a steady dc level or any desired function of time or other variable. The voltage across the load may be any value from plus infinity to minus infinity and will be determined solely by the current delivered and the value of the load impedance. The compliance of the source, defined as its output voltage capability, is, therefore, infinite.

The practical current source, on the other hand, only approximates a perfect source. In its linear range it may be described, as in Fig. 2b, as a perfect current source in parallel with a practical impedance. For short-circuited conditions, $Z_L = 0$, there is no problem. But as the load Z_L is increased, a load voltage V_L is developed, which diverts some of the current through the source impedance Z_L. Unless Z_s can be made very high with respect to the highest expected Z_L, the load current will not be constant. But if Z_s can be made infinite, none of the source current will be diverted through the source impedance and the load current will always equal the source current.

Conventional Circuit Is Base for Modifications

A simplified schematic of a conventional current regulator is shown in Fig. 3. Here, the load current I_L flows through a sensing resistor, R_2, to develop a voltage that is proportional to I_L. This voltage is compared to a reference voltage developed across a Zener diode, *D1*. Any difference is amplified by *Q1*, which acts as a differential amplifier. The amplified signal drives the regulator or "pass" transistor *Q2*, which controls the current delivered by battery E to the load. Resistor R_1 is needed only to provide current for the Zener diode.

If the load impedance decreases, the load current increases, along with the voltage across R_2, driving the emitter of *Q1* more positive with respect to its base and tending

to cut off *Q2*. This counteracts the original increase of I_L.

The regulator acts to maintain the load current at a value that makes $I_L R_2$ approximately equal to the reference voltage across the Zener diode. The output current can be changed through adjustment of R_2.

The performance of conventional current-source supplies is adequate for low and moderate load impedances, but higher load impedances tend to upset the regulating action.

The clue to the desired modifications is found in Fig. 2b. As long as $Z_L = 0$, all the source current flows through the load.

It is the voltage developed across a non-zero Z_L that causes the current to flow in Z_s, upsetting the current regulation. But if this voltage could be sampled and used to increase the current delivered by the "perfect" current source, this effect could be counteracted. If this increase could be made just equal to the current drawn by Z_s, there would be no change in the current through Z_L and the desired infinite Z_o will have been achieved.

The first of the series of modification to realize this objective is shown in Fig. 4a. Any voltage developed across the load is sampled at the arm of potentiometer R_4 and applied through *D1* to the base of *Q1*, tending to turn *Q1* further on. This in turn drives *Q2* further on, permitting more load current to flow. In effect, this modification provides a negative output impedance that is parallel with and cancels the positive output impedance of the unmodified source. The parallel combination can be made positive, negative or infinite by the setting of R_4.

Further inspection of Fig. 4a shows that the Zener current flows through the lower part of R_4, developing a dc voltage in series with the Zener voltage. Any changes in Zener current, such as would be caused by a variation in the dc supply voltage E, would change the effective reference voltage for the regulator and, correspondingly, the current delivered.

This difficulty can be circumvented by feeding the developed load voltage signal back to the amplifier base in parallel, rather than in series, with the reference diode. This technique is illustrated in Fig. 4b. Rheostat R_3 is varied to control the amount of feedback, and thus the regulator output impedance. Resistor R_5 limits the amount of feedback to protect the circuit.

The current through R_3 returns to the minus side of the dc supply by two paths. The first is through the low-impedance Zener diode and the second is through the relatively high-impedance path made up of R_2 and the base-emitter diode of *Q1*. Thus, it is apparent that only a small fraction of the feedback current reaches *Q1* and that this fraction is affected by the dynamic impedance of the Zener diode.

A more sensitive and more stable current injection scheme is shown in Fig. 4c. Resistor R_6 forces a larger fraction of the feedback current through R_3 into the transistor base and also stabilizes the total impedance into which the current flows. The effect of R_6 on the basic regulator performance is negligible. The value of R_3 now can be increased; there still will be enough feedback to produce the infinite impedance effect.

Positive Feedback May Cause Stability Problem

The two transistors in the conventional regulator shown in Fig. 3 were used as cascaded voltage amplifiers. If the negative side of the dc supply is considered as a reference point, it will be seen that *Q1* was operated in the common-base mode while *Q2* was operated in the common-emitter mode. A signal starting at the emitter of *Q1* would appear amplified without phase reversal at the collector of *Q1*, and from there would feed into the base of *Q2*. The load impedance on the collector of *Q1* is the base input impedance of transistor *Q2*.

The signal at the base of *Q2* appears amplified and inverted at the collector of *Q2*. The load for *Q2* is Z_L—the external load impedance—plus the parallel combination of R_2 and the input emitter impedance of *Q1*.

The signal at the collector of *Q2* is divided between Z_L and the combination of R_2 and *Q1* input impedance and appears at the emitter of *Q1* out of phase with the initiating signal. This feedback is, therefore, negative for frequencies at which extraneous phase shifts are negligible.

But with increasing frequencies, additional phase lags become more noticeable, and when the total lag reaches 360 degrees, the feedback becomes positive. If the loop gain is greater than unity when the lag reaches 360 degrees, self-oscillation occurs, and the advantages of the regulator are lost.

A series RC feedback from collector to base of *Q1*, as shown in Fig. 4d, helps combat this tendency by reducing the loop gain at higher frequencies without introducing additional phase shift. As the frequency increases from zero, the reactance of the capacitor C begins to shunt the collector load for *Q1* and the loop gain will decrease. Some phase lag will be introduced at this point.

As the frequency increases further, the reactance of C becomes negligible with respect to R_7 and the load for $Q1$ is nearly the value of R_7. Because of this resistive load, the extra phase shift disappears and because of the small size of R_7, the loop over-all gain is materially reduced. The bandwidth over which the constant current effect is produced depends on the bandwidth of the amplifier loop. This stabilization is at the expense of the supply's bandwidth.

The modification for infinite output impedance also may cause instability. The new loop that senses the presence of voltage across the load introduces a small amount of positive feedback. When the supply is exactly adjusted for infinite output impedance, it is on the verge of instability for the case of infinite load impedance. But the finite load impedance that would be encountered in any real application reduces the gain of this secondary loop to less than unity, ensuring stability.

The circuit of Fig. 1 uses the modified parallel-injection scheme of Fig. 4c. The transistors used are complementary germanium types chosen as representative of a broad class of low-cost, general-purpose and switching devices. Since $Q1$'s collector must supply the base current for $Q2$, $Q1$ must be able to operate with collector currents on the order of 10-100 μa.

The second Zener, $D2$, protects the regulator in case of an open-circuit load. Without $D2$, an open circuit-load would result in an excessive forward base-emitter bias on $Q1$. In that case, the $Q2$'s base-emitter diode, the saturated $Q1$ and R_2 would be in series across the 12-v supply. The large resultant current surge could easily damage the circuit. Thus, it is important to have $D2$ available to carry the entire specified load current in the event of an open-circuit load.

Experimental Setups to Help Designer Evaluate and Adjust Supply

The dc output impedance of the regulator can be measured as shown in Fig. 5a. Meter $M1$ reads the total load current, which for this circuit falls between 1 and 5 ma. Meter $M2$ registers the difference between the load current I_L and the test loop current I_2. If this difference is made approximately zero through adjustment of R_8, $M2$ may be a very sensitive meter capable of showing a change of a fraction of a microampere. Any change indicated by $M2$ results from a change in I_L, or:

$$\Delta I_L = \Delta I_{M2} \frac{R_8 + R_M}{R_8}$$

where:

ΔI_L = change in load current.
ΔI_{M2} = change in current through $M2$.
R_M = resistance of $M2$.

If the switch is moved to change the load voltage by an amount $\Delta E_L = E_2$, the regulator output impedande is given by:

$$Z_o = \frac{E_2}{\Delta I_L} = \frac{E_2 R_8}{\Delta I_{M2}(R_8 + R_M)}$$

Measurements on the unmodified source—that is, the circuit of Fig. 1 with $R_6 = 0$, $R_3 = R_5 = \infty$—will show an output impedance of 95 K at a load current of 5 ma, increasing nearly linearly to 400 K at 1 ma. The load-current output can be reduced by increasing R_2. This also increases the loop gain and the output impedance.

Feedback in the regulator can be adjusted so that ΔI_{M2} is zero for non-zero E_2, yielding infinite output impedance at dc. A further increase in the amount of feedback in the regulator results in an increase of I_L for an increase of E_2, demonstrating that a negative output impedance can exist.

It will be noted that there is a gradual, rather than immediate, change in I_L on the order of 1 μamp soon after the test circuit switch is thrown. The voltage drop across $Q2$ changes by an amount very nearly equal to E_2 and the internal dissipation changes accordingly.

The change in dissipation alters the device temperature, which in turn causes a slight change in the operating point. This accounts for the slow response observed in I_L.

Setup to Show Regulation As Function of Frequency

The output impedance as a function of frequency may be measured as shown in Fig. 5b. The current "i" is a small ac component superimposed on the dc load current I_L. Its value is given by:

$$i = \frac{e_2}{R_9}$$

and the output impedance is then:

$$Z_o = \frac{e_i}{i} - R_9 = R_9\left(\frac{e_1}{e_2} - 1\right)$$

As Z_o becomes large, e_1 far exceeds e_2 and this expression reduces to:

$$Z_o = R_9 \frac{e_1}{e_2}$$

As Z_o becomes infinite, i and e_2 become zero; e_2 being a particularly good indication for the adjustment for infinite Z_o.

If e_2 is read on a sensitive oscilloscope synchronized to the audio oscillator, it will be seen that the phase of e_2 reverses as the null is traversed. This shows that the output impedance of the current regulator actually becomes negative.

As a practical matter in the setup of this test circuit, the transformer used for signal injection must be able to operate without saturation over the frequency range of interest. It must be remembered that the transformer will at times be carrying dc as well as ac signals. In presenting the results of tests on this sort of supply, it is convenient to use the reciprocal of output impedance or the admittance. Fig. 6 shows the output admittance of the regulator as a function of frequency for any load current in the range of 1 to 5 ma.

If it is possible to omit the stabilizing network, the bandwidth can be extended, the amount of this increase being shown in the graph, Fig. 6.

The high output impedance at the higher frequencies can be maintained by using an inductor in series with the output lead. This is, of course, analogous to the use of a shunt capacitor in the more common voltage-regulated supplies. The disadvantage of this technique is that some of the output voltage capability of the current regulator may be lost across the inductor, thus reducing the amount available for the load. ■ ■

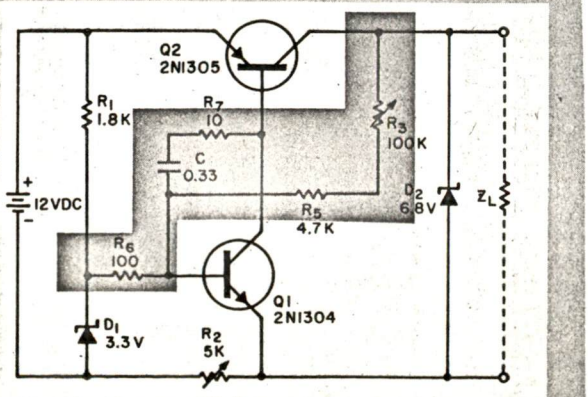


Fig. 1. The modified constant-current source achieves infinite output impedance by the positive feedback loop (shaded), which senses any voltage developed by the load.

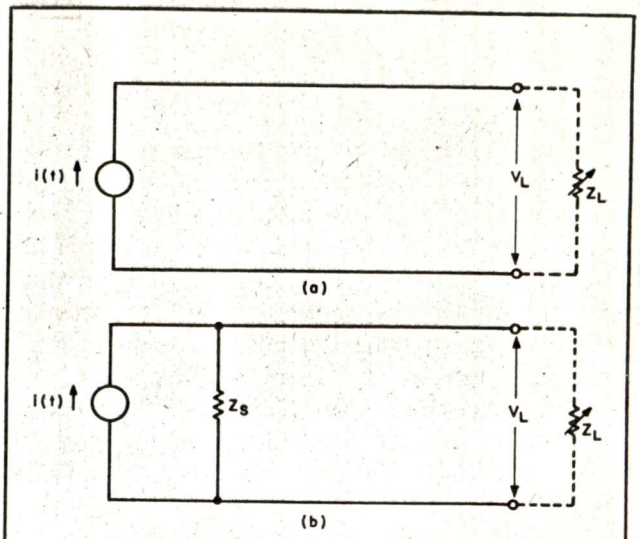


Fig. 2. Review of symbols for perfect (*a*) and practical (*b*) current sources: Though the ideal current source is by definition able to deliver a specific value of current regardless of the nature of the load, practical sources have internal shunts, Z_s, that make them sensitive to the nature of the load and the frequency of the source or any system disturbance. The ideal current source is shown as a function of time to indicate that the analysis is not necessarily limited to dc.

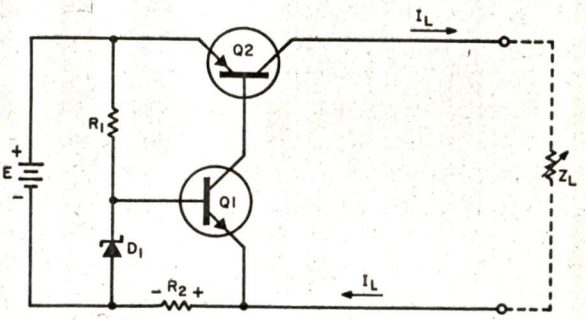


Fig. 3. Conventional constant-current sources depend entirely on sensing the load current. Changes in the load current alter the voltage across R_2 and the base-emitter voltage into *Q1*, which in turn alters the amount of current that Q2 passes.

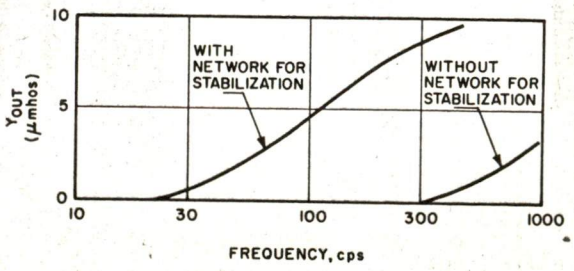
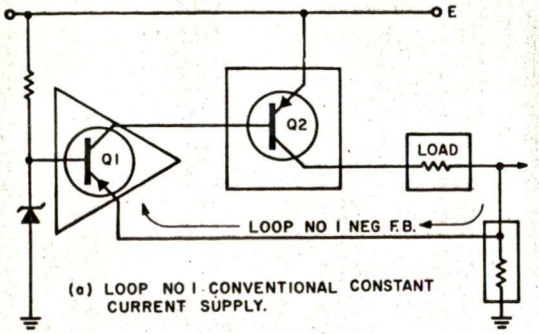
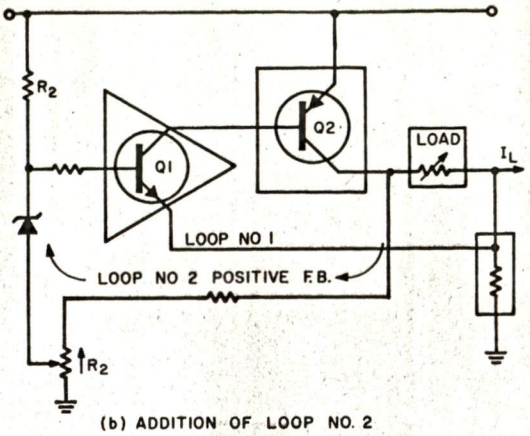


Fig. 6. Curves of admittance against frequency show the effect of the stabilization loop. It is possible to dispense with the stabilization under certain load-impedance conditions.

Purpose of Three Loops

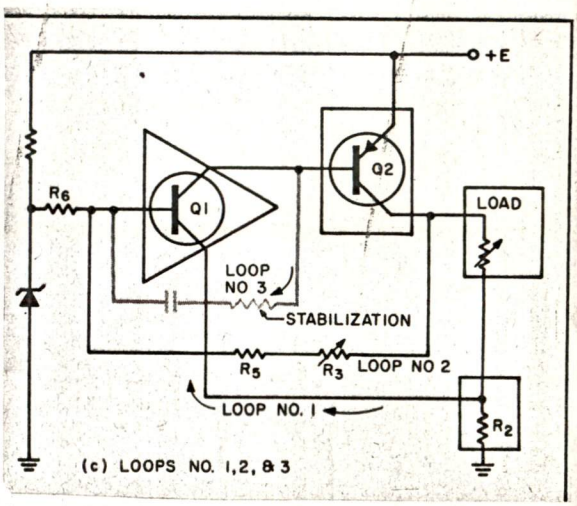


(a) LOOP NO 1 CONVENTIONAL CONSTANT CURRENT SUPPLY.

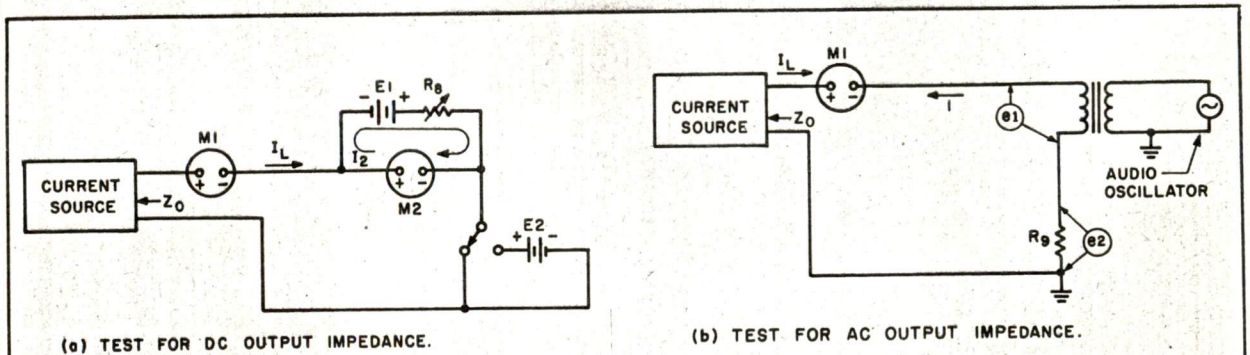


(b) ADDITION OF LOOP NO. 2

When the constant current circuit is redrawn in the classical servo format, the action of each of the three loops in the circuit is made more obvious. Loop No. 1 is the conventional current-sensing loop. Loop No. 2 is the positive feedback loop that senses the presence of voltage across the load, and loop No. 3 is the stabilizing loop that cuts down the gain at higher frequencies.



(c) LOOPS NO. 1, 2, & 3



(a) TEST FOR DC OUTPUT IMPEDANCE.

(b) TEST FOR AC OUTPUT IMPEDANCE.

Fig. 5. Test circuits to check out and adjust an infinite-impedance current source supply, showing measurement of dc output impedance (a) and measurement of ac output impedance (b).

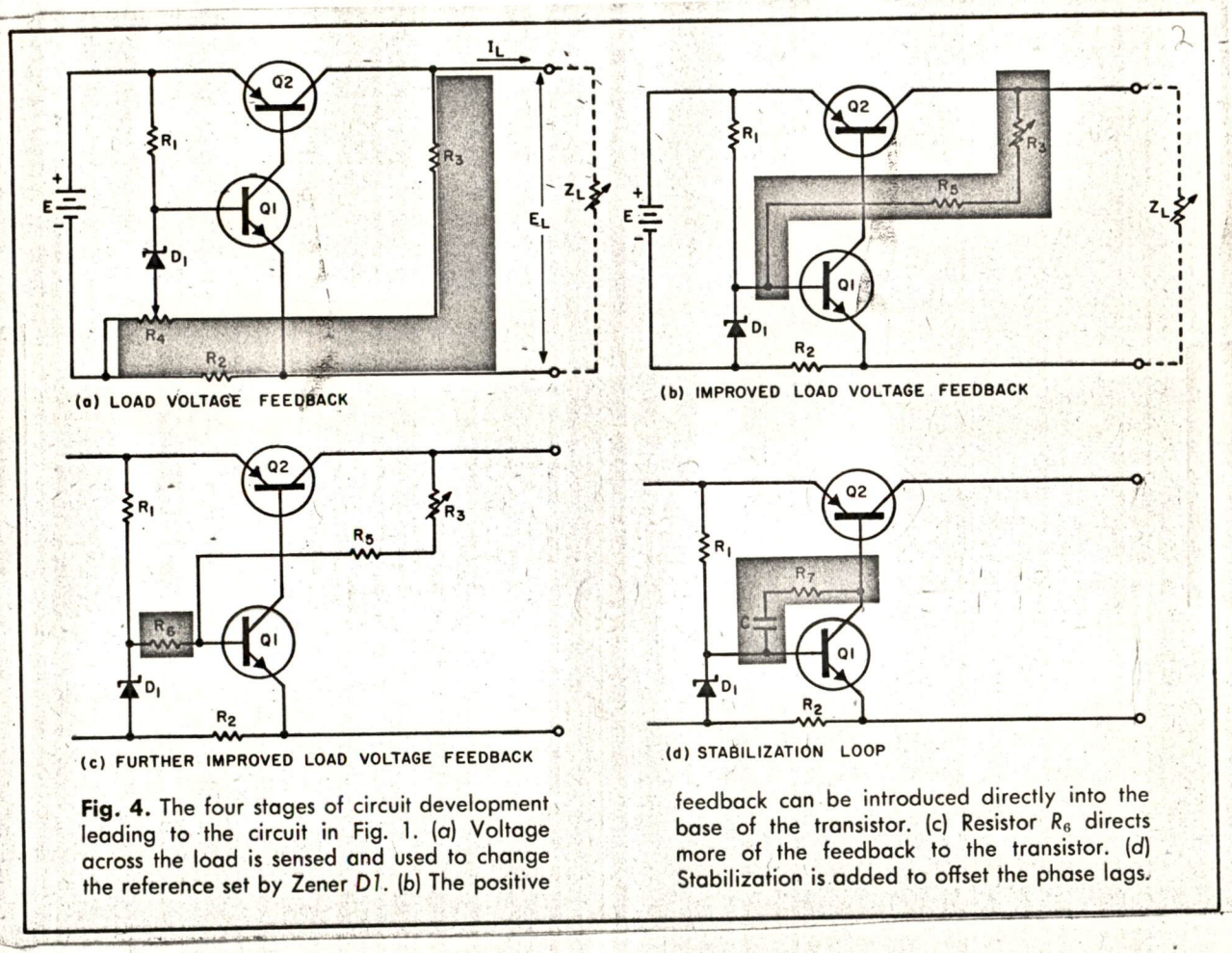


Fig. 4. The four stages of circuit development leading to the circuit in Fig. 1. (a) Voltage across the load is sensed and used to change the reference set by Zener *D1*. (b) The positive feedback can be introduced directly into the base of the transistor. (c) Resistor R_6 directs more of the feedback to the transistor. (d) Stabilization is added to offset the phase lags.

Infinite Impedance From a Constant-Current Source

R. W. Thorpe
Design Specialist
General Dynamics/Pomona, Pomona, Calif.

The trick is to sample the voltage across the load as well as the current through it. The positive feedback developed in this manner can raise the output impedance to infinity for a given range of load conditions.

HOW CAN a conventional constant-current supply be modified for infinite output impedance? One answer is, by adding a regulating loop that senses and compensates for the circuit variable causing the departure from ideal behavior.

The culprit is the voltage developed across the load. This voltage increases in proportion to the size of the load; with large loads it can become big enough to divert some of the current through the supply's internal shunt resistance. If the presence of this voltage can be sensed and used to introduce a compensating current, the supply can be made to act as if it had infinite output impedance.

A circuit incorporating this load voltage compensation, in addition to the regular current-sensing control, is shown in Fig. 1. The circuit is simple enough, but its evolution will give designers an appreciation of the concepts involved.

The problems in designing a high-performance current source are illustrated by the differences between a perfect and a practical constant-current source. The perfect current source, Fig. 2a, is by definition one that delivers a specified current into any load to which it may be connected. The specified current may be a steady dc level or any desired function of time or other variable. The voltage across the load may be any value from plus infinity to minus infinity and will be determined solely by the current delivered and the value of the load impedance. The compliance of the source, defined as its output voltage capability, is, therefore, infinite.

The practical current source, on the other hand, only approximates a perfect source. In its linear range it may be described, as in Fig. 2b, as a perfect current source in parallel with a practical impedance. For short-circuited conditions, $Z_L = 0$, there is no problem. But as the load Z_L is increased, a load voltage V_L is developed, which diverts some of the current through the source impedance Z_L. Unless Z_s can be made very high with respect to the highest expected Z_L, the load current will not be constant. But if Z_s can be made infinite, none of the source current will be diverted through the source impedance and the load current will always equal the source current.

Conventional Circuit Is Base for Modifications

A simplified schematic of a conventional current regulator is shown in Fig. 3. Here, the load current I_L flows through a sensing resistor, R_2, to develop a voltage that is proportional to I_L. This voltage is compared to a reference voltage developed across a Zener diode, *D1*. Any difference is amplified by *Q1*, which acts as a differential amplifier. The amplified signal drives the regulator or "pass" transistor *Q2*, which controls the current delivered by battery E to the load. Resistor R_1 is needed only to provide current for the Zener diode.

If the load impedance decreases, the load current increases, along with the voltage across R_2, driving the emitter of *Q1* more positive with respect to its base and tending

to cut off $Q2$. This counteracts the original increase of I_L.

The regulator acts to maintain the load current at a value that makes $I_L R_2$ approximately equal to the reference voltage across the Zener diode. The output current can be changed through adjustment of R_2.

The performance of conventional current-source supplies is adequate for low and moderate load impedances, but higher load impedances tend to upset the regulating action. The clue to the desired modifications is found in Fig. 2b. As long as $Z_L = 0$, all the source current flows through the load.

It is the voltage developed across a non-zero Z_L that causes the current to flow in Z_s, upsetting the current regulation. But if this voltage could be sampled and used to increase the current delivered by the "perfect" current source, this effect could be counteracted. If this increase could be made just equal to the current drawn by Z_s, there would be no change in the current through Z_L and the desired infinite Z_o will have been achieved.

The first of the series of modification to realize this objective is shown in Fig. 4a. Any voltage developed across the load is sampled at the arm of potentiometer R_4 and applied through $D1$ to the base of $Q1$, tending to turn $Q1$ further on. This in turn drives $Q2$ further on, permitting more load current to flow. In effect, this modification provides a negative output impedance that is parallel with and cancels the positive output impedance of the unmodified source. The parallel combination can be made positive, negative or infinite by the setting of R_4.

Further inspection of Fig. 4a shows that the Zener current flows through the lower part of R_4, developing a dc voltage in series with the Zener voltage. Any changes in Zener current, such as would be caused by a variation in the dc supply voltage E, would change the effective reference voltage for the regulator and, correspondingly, the current delivered.

This difficulty can be circumvented by feeding the developed load voltage signal back to the amplifier base in parallel, rather than in series, with the reference diode. This technique is illustrated in Fig. 4b. Rheostat R_3 is varied to control the amount of feedback, and thus the regulator output impedance. Resistor R_5 limits the amount of feedback to protect the circuit.

The current through R_3 returns to the minus side of the dc supply by two paths. The first is through the low-impedance Zener diode and the second is through the relatively high-impedance path made up of R_2 and the base-emitter diode of $Q1$. Thus, it is apparent that only a small fraction of the feedback current reaches $Q1$ and that this fraction is affected by the dynamic impedance of the Zener diode.

A more sensitive and more stable current injection scheme is shown in Fig. 4c. Resistor R_6 forces a larger fraction of the feedback current through R_3 into the transistor base and also stabilizes the total impedance into which the current flows. The effect of R_6 on the basic regulator performance is negligible. The value of R_3 now can be increased; there still will be enough feedback to produce the infinite impedance effect.

Positive Feedback May Cause Stability Problem

The two transistors in the conventional regulator shown in Fig. 3 were used as cascaded voltage amplifiers. If the negative side of the dc supply is considered as a reference point, it will be seen that $Q1$ was operated in the common-base mode while $Q2$ was operated in the common-emitter mode. A signal starting at the emitter of $Q1$ would appear amplified without phase reversal at the collector of $Q1$, and from there would feed into the base of $Q2$. The load impedance on the collector of $Q1$ is the base input impedance of transistor $Q2$.

The signal at the base of $Q2$ appears amplified and inverted at the collector of $Q2$. The load for $Q2$ is Z_L—the external load impedance—plus the parallel combination of R_2 and the input emitter impedance of $Q1$.

The signal at the collector of $Q2$ is divided between Z_L and the combination of R_2 and $Q1$ input impedance and appears at the emitter of $Q1$ out of phase with the initiating signal. This feedback is, therefore, negative for frequencies at which extraneous phase shifts are negligible.

But with increasing frequencies, additional phase lags become more noticeable, and when the total lag reaches 360 degrees, the feedback becomes positive. If the loop gain is greater than unity when the lag reaches 360 degrees, self-oscillation occurs, and the advantages of the regulator are lost.

A series RC feedback from collector to base of $Q1$, as shown in Fig. 4d, helps combat this tendency by reducing the loop gain at higher frequencies without introducing additional phase shift. As the frequency increases from zero, the reactance of the capacitor C begins to shunt the collector load for $Q1$ and the loop gain will decrease. Some phase lag will be introduced at this point.

As the frequency increases further, the reactance of C becomes negligible with respect to R_7 and the load for *Q1* is nearly the value of R_7. Because of this resistive load, the extra phase shift disappears and because of the small size of R_7, the loop over-all gain is materially reduced. The bandwidth over which the constant current effect is produced depends on the bandwidth of the amplifier loop. This stabilization is at the expense of the supply's bandwidth.

The modification for infinite output impedance also may cause instability. The new loop that senses the presence of voltage across the load introduces a small amount of positive feedback. When the supply is exactly adjusted for infinite output impedance, it is on the verge of instability for the case of infinite load impedance. But the finite load impedance that would be encountered in any real application reduces the gain of this secondary loop to less than unity, ensuring stability.

The circuit of Fig. 1 uses the modified parallel-injection scheme of Fig. 4c. The transistors used are complementary germanium types chosen as representative of a broad class of low-cost, general-purpose and switching devices. Since *Q1*'s collector must supply the base current for *Q2*, *Q1* must be able to operate with collector currents on the order of 10-100 μa.

The second Zener, *D2*, protects the regulator in case of an open-circuit load. Without *D2*, an open circuit-load would result in an excessive forward base-emitter bias on *Q1*. In that case, the *Q2*'s base-emitter diode, the saturated *Q1* and R_2 would be in series across the 12-v supply. The large resultant current surge could easily damage the circuit. Thus, it is important to have *D2* available to carry the entire specified load current in the event of an open-circuit load.

Experimental Setups to Help Designer Evaluate and Adjust Supply

The dc output impedance of the regulator can be measured as shown in Fig. 5a. Meter *M1* reads the total load current, which for this circuit falls between 1 and 5 ma. Meter *M2* registers the difference between the load current I_L and the test loop current I_2. If this difference is made approximately zero through adjustment of R_8, *M2* may be a very sensitive meter capable of showing a change of a fraction of a microampere. Any change indicated by *M2* results from a change in I_L, or:

$$\Delta I_L = \Delta I_{M2} \frac{R_8 + R_M}{R_8}$$

where:

ΔI_L = change in load current.
ΔI_{M2} = change in current through *M2*.
R_M = resistance of *M2*.

If the switch is moved to change the load voltage by an amount $\Delta E_L = E_2$, the regulator output impedande is given by:

$$Z_o = \frac{E_2}{\Delta I_L} = \frac{E_2 R_8}{\Delta I_{M2}(R_8 + R_M)}$$

Measurements on the unmodified source—that is, the circuit of Fig. 1 with $R_6 = 0$, $R_3 = R_5 = \infty$—will show an output impedance of 95 K at a load current of 5 ma, increasing nearly linearly to 400 K at 1 ma. The load-current output can be reduced by increasing R_2. This also increases the loop gain and the output impedance.

Feedback in the regulator can be adjusted so that ΔI_{M2} is zero for non-zero E_2, yielding infinite output impedance at dc. A further increase in the amount of feedback in the regulator results in an increase of I_L for an increase of E_2, demonstrating that a negative output impedance can exist.

It will be noted that there is a gradual, rather than immediate, change in I_L on the order of 1 μamp soon after the test circuit switch is thrown. The voltage drop across *Q2* changes by an amount very nearly equal to E_2 and the internal dissipation changes accordingly.

The change in dissipation alters the device temperature, which in turn causes a slight change in the operating point. This accounts for the slow response observed in I_L.

Setup to Show Regulation As Function of Frequency

The output impedance as a function of frequency may be measured as shown in Fig. 5b. The current "*i*" is a small ac component superimposed on the dc load current I_L. Its value is given by:

$$i = \frac{e_2}{R_9}$$

and the output impedance is then:

$$Z_o = \frac{e_i}{i} - R_9 = R_9\left(\frac{e_1}{e_2} - 1\right)$$

As Z_o becomes large, e_1 far exceeds e_2 and this expression reduces to:

$$Z_o = R_9 \frac{e_1}{e_2}$$

As Z_o becomes infinite, i and e_2 become zero; e_2 being a particularly good indication for the adjustment for infinite Z_o.

If e_2 is read on a sensitive oscilloscope synchronized to the audio oscillator, it will be seen that the phase of e_2 reverses as the null is traversed. This shows that the output impedance of the current regulator actually becomes negative.

As a practical matter in the setup of this test circuit, the transformer used for signal injection must be able to operate without saturation over the frequency range of interest. It must be remembered that the transformer will at times be carrying dc as well as ac signals. In presenting the results of tests on this sort of supply, it is convenient to use the reciprocal of output impedance or the admittance. Fig. 6 shows the output admittance of the regulator as a function of frequency for any load current in the range of 1 to 5 ma.

If it is possible to omit the stabilizing network, the bandwidth can be extended, the amount of this increase being shown in the graph, Fig. 6.

The high output impedance at the higher frequencies can be maintained by using an inductor in series with the output lead. This is, of course, analogous to the use of a shunt capacitor in the more common voltage-regulated supplies. The disadvantage of this technique is that some of the output voltage capability of the current regulator may be lost across the inductor, thus reducing the amount available for the load. ■ ■

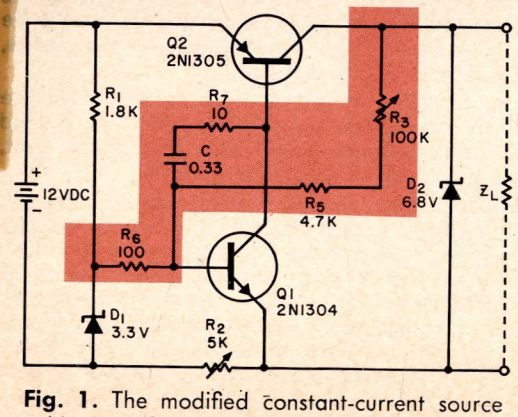


Fig. 1. The modified constant-current source achieves infinite output impedance by the positive feedback loop (shaded), which senses any voltage developed by the load.

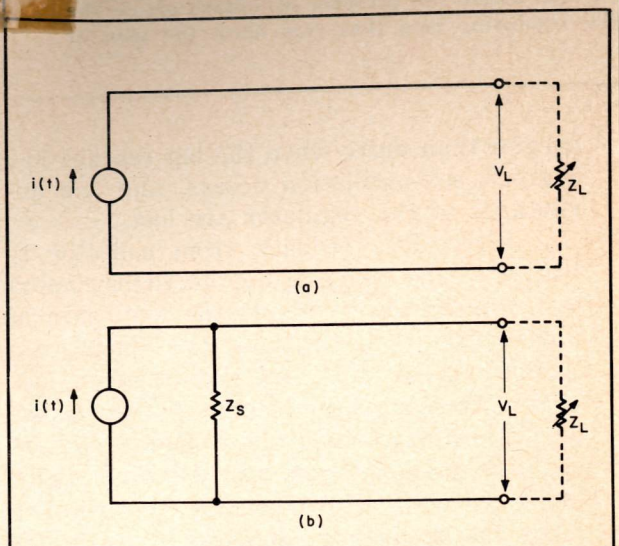


Fig. 2. Review of symbols for perfect (*a*) and practical (*b*) current sources: Though the ideal current source is by definition able to deliver a specific value of current regardless of the nature of the load, practical sources have internal shunts, Z_s, that make them sensitive to the nature of the load and the frequency of the source or any system disturbance. The ideal current source is shown as a function of time to indicate that the analysis is not necessarily limited to dc.

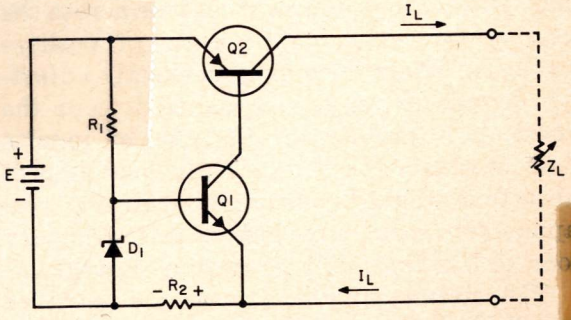


Fig. 3. Conventional constant-current sources depend entirely on sensing the load current. Changes in the load current alter the voltage across R_2 and the base-emitter voltage into Q1, which in turn alters the amount of current that Q2 passes.

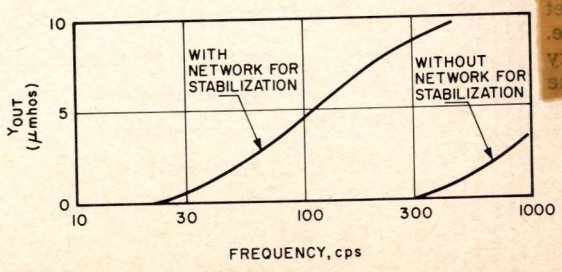
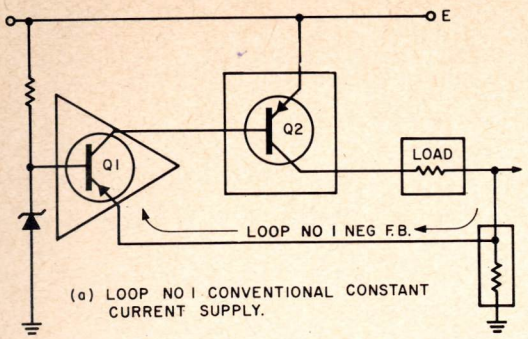
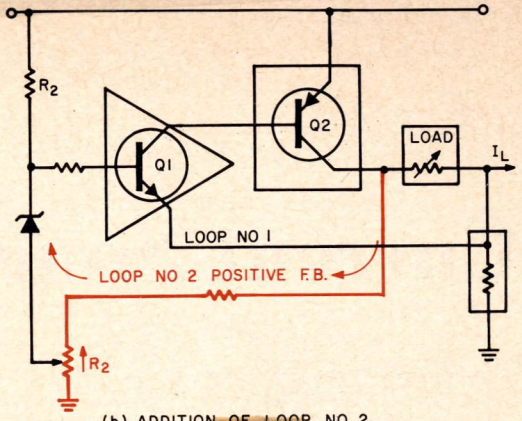


Fig. 6. Curves of admittance against frequency show the effect of the stabilization loop. It is possible to dispense with the stabilization under certain load-impedance conditions.

Purpose of Three Loops

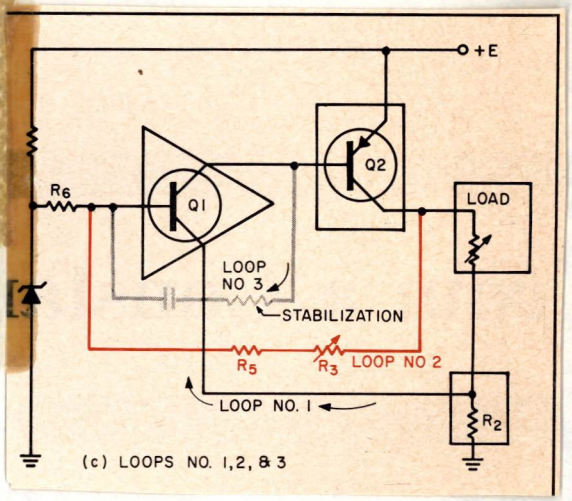


(a) LOOP NO 1 CONVENTIONAL CONSTANT CURRENT SUPPLY.

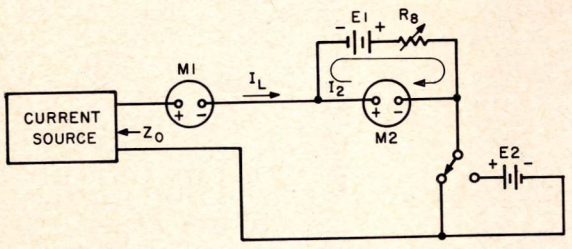


(b) ADDITION OF LOOP NO. 2

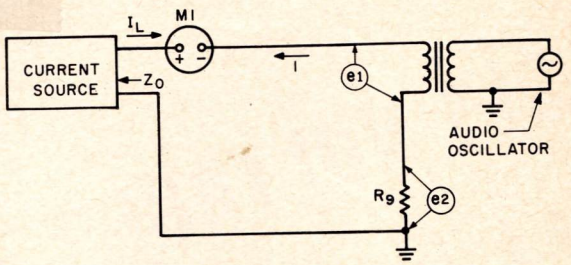
When the constant current circuit is redrawn in the classical servo format, the action of each of the three loops in the circuit is made more obvious. Loop No. 1 is the conventional current-sensing loop. Loop No. 2 is the positive feedback loop that senses the presence of voltage across the load, and loop No. 3 is the stabilizing loop that cuts down the gain at higher frequencies.



(c) LOOPS NO. 1, 2, & 3



(a) TEST FOR DC OUTPUT IMPEDANCE.



(b) TEST FOR AC OUTPUT IMPEDANCE.

Fig. 5. Test circuits to check out and adjust an infinite-impedance current source supply, showing measurement of dc output impedance (*a*) and measurement of ac output impedance (*b*).

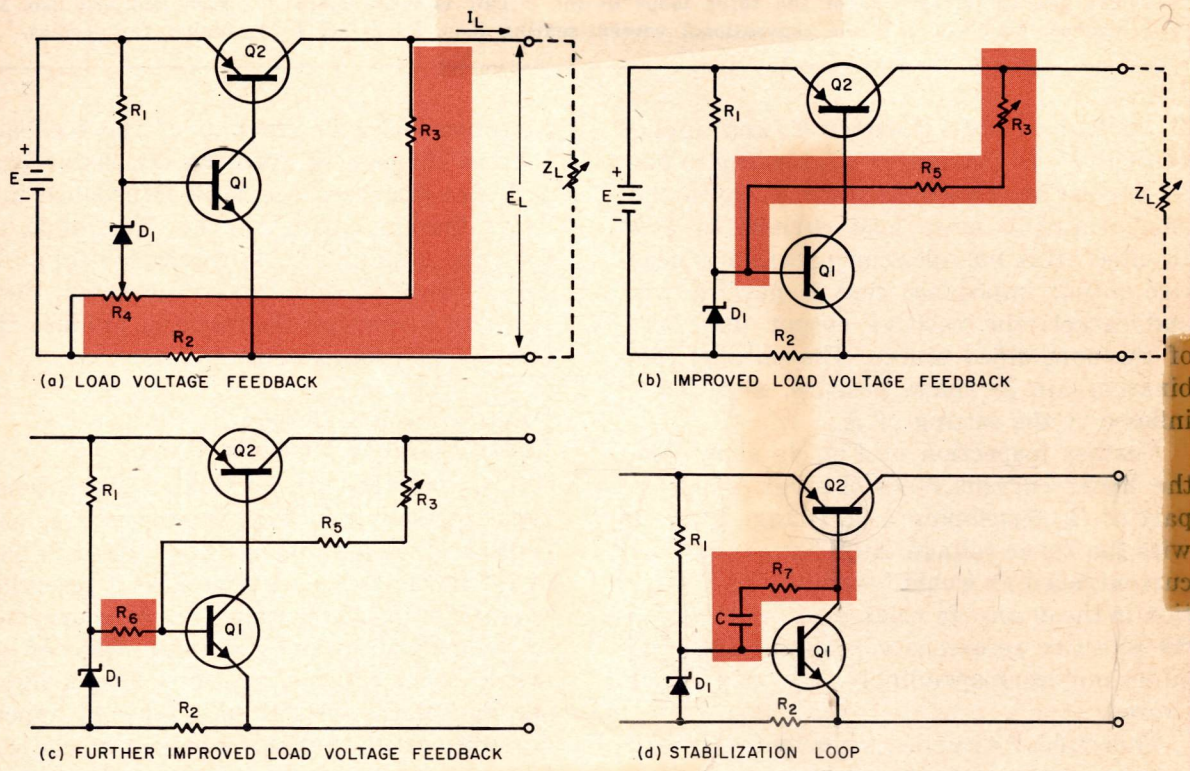


Fig. 4. The four stages of circuit development leading to the circuit in Fig. 1. (*a*) Voltage across the load is sensed and used to change the reference set by Zener *D1*. (*b*) The positive feedback can be introduced directly into the base of the transistor. (*c*) Resistor R_6 directs more of the feedback to the transistor. (*d*) Stabilization is added to offset the phase lags.

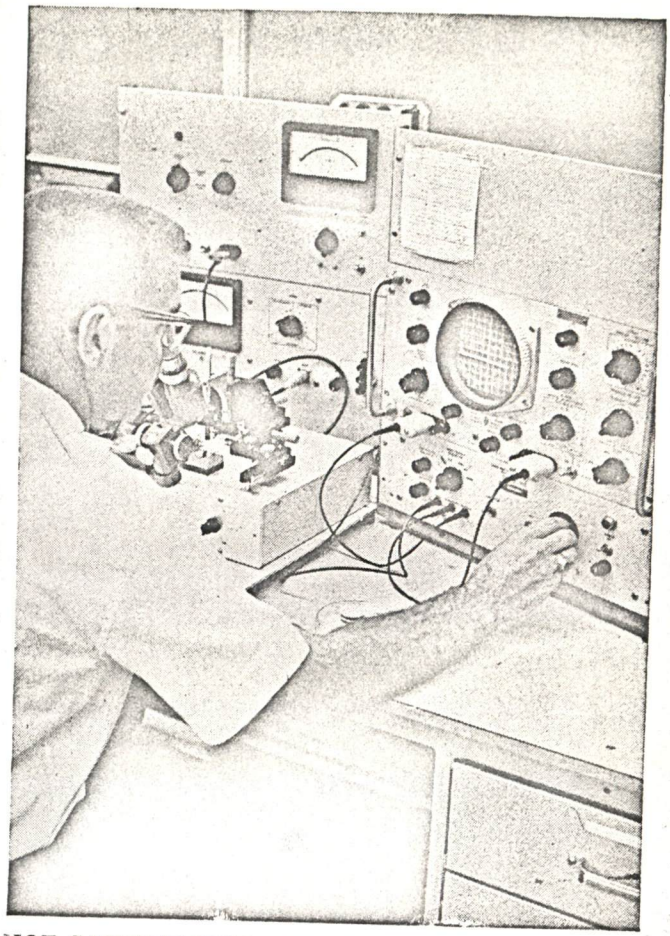
-4702-
-4703-
A
Grey

ANOTHER NEW COMPONENT

HOT CARRIER DIODES Switch in Picoseconds

Metal-semiconductor diodes increase switching speed now limited in **p-n** *junctions by minority-carrier storage. Devices need very pure materials and improved epitaxy*

By S. M. KRAKAUER, Applications Engineer
S. W. SOSHEA, Project Leader
HP Associates, Palo Alto, California

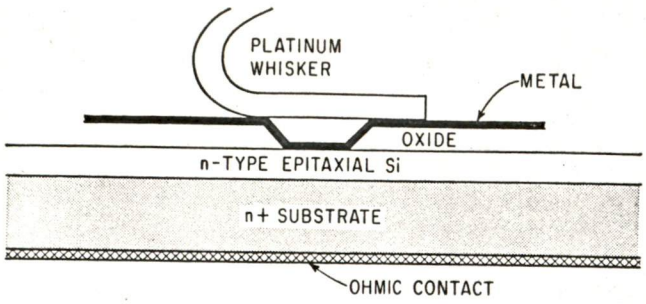


HOT CARRIER DIODE WAFER *undergoing evaluation of leakage current by project engineer Bill Baker*

EPITAXIAL SILICON hot carrier diodes were introduced by HP Associates only recently, but many of the principles of rectifying metal contacts have been known for decades. The great advances in germanium and silicon *p-n* junction devices in the last 15 years have tended to obscure the potential of metal-semiconductor contacts. The *p-n* junction diodes are, however, approaching the limit of their high-frequency performance, because of storage of minority carriers. Since minority carrier storage is virtually eliminated, metal-semiconductor diodes show renewed usefulness. The development of the modern hot carrier diode was made possible by the availability of very pure semiconductors, by improved techniques of surface cleaning and passivation, and by the epitaxial construction method.

A hot carrier diode can be made in a variety of ways. A typical epitaxial type is shown in Fig. 1. Experimental models have been made on silicon using evaporated gold, platinum, palladium, silver and many other metals. Both hot electron (on *n*-type silicon) and hot hole (on *p*-type silicon) forms are possible, but the hot electron type is generally preferable because the higher electron mobility gives better high frequency performance.

OPERATION—Hot carrier diode operation and the distinction between it and a *p-n* junction can be understood most clearly by means of the appropriate electron energy diagrams. Figure 2A shows the electron energy diagram of a hot electron diode with a Schottky-type barrier, and Fig. 2B shows the corresponding diagram for an abrupt *p-n* junction diode with the *n*-type region more heavily doped than the *p*-type region. When the hot carrier diode is forward biased, the electrons in the *n*-type semiconductor



HOT ELECTRON DIODE *construction differs from that of hot hole where substrate would be* **p-plus** *material and epitaxial layer* **p-type** *silicon—Fig. 1*

diffuse over the barrier and are injected into the metal.

The injected hot electrons interact with the lattice and the electrons of the metal and when the diode is reverse biased, these hot electrons are unable to surmount the barrier, so they do not contribute to the stored charge. When, however, the *p-n* junction is forward biased, the electrons diffuse into the *p*-type region and build up to a concentration that is limited by the rate of carrier recombination, as depicted in Fig. 2B. When the *p-n* junction is reverse biased, the stored electrons (minority carriers) flow back into the *n*-type region, thus lowering the rectification efficiency if the diode is used as a detector, or increasing the reverse recovery time if it is used as a switching diode.

The current-voltage characteristics of hot carrier diodes can be described very closely by the ideal diode equation

$$I_F = I_s\ [\exp(qV/kT) - 1]$$

in which the saturation current I_s is proportional to $\exp(-qV_i/kT)$. The type of metal can be conveniently selected to have an internal barrier V_i from 0.3 to 0.8 volt, corresponding to a saturation current, for a typical diode size (about 6×10^{-6} cm^2), from 10^{-11} to 10^{-8} amp. The junction capacitance of the hot carrier diode varies as the inverse square root of voltage and is only slightly dependent on V_i. This combination of characteristics is analogous to a family of *p-n* junctions of incrementally varying energy band gap and provides the circuit designer with an added degree of freedom that was not previously available. The reverse characteristics of hot carrier diodes appear very similar to those of *p-n* junction diodes. The reverse leakage current increases with reverse voltage gradually, owing to the internal Schottky effect, until the avalanche multiplication voltage is reached.

Hot carrier diodes are similar in concept and in operation to the ideal point-contact diode in which the contact is neither formed, alloyed nor bonded. Both the hot carrier diode and the ideal point contact diode employ a Schottky barrier, but there are many notable differences. Being of much larger area, the hot carrier diode has larger capacitance than the point contact, but it can handle greater power and is less sensitive to current transients than is the ideal point contact. The hot carrier diode, furthermore, is more stable mechanically and has more nearly ideal and reproducible electrical characteristics.

CHARACTERISTICS OF HOT-CARRIER DIODES

	HPA-2001			HPA-2101		
	Min	Typ	Max	Min	Typ	Max
Forward Current I_F in ma (at $V_F = 1$ v)	20	30		20	30	
Forward Current I_F in μa (at $V_F = 0.4$ v)	0.5	1.0		2,000	4,000	
Capacitance C_o in pf (at $V_B = 0$)		0.85	1.0		0.95	1.1
Breakdown Voltage V_B in v (at $I_R = 10$ μa)	15	25		15	25	
Leakage Current I_s in na (at $V_R = 3$ v)		1	10		30	100

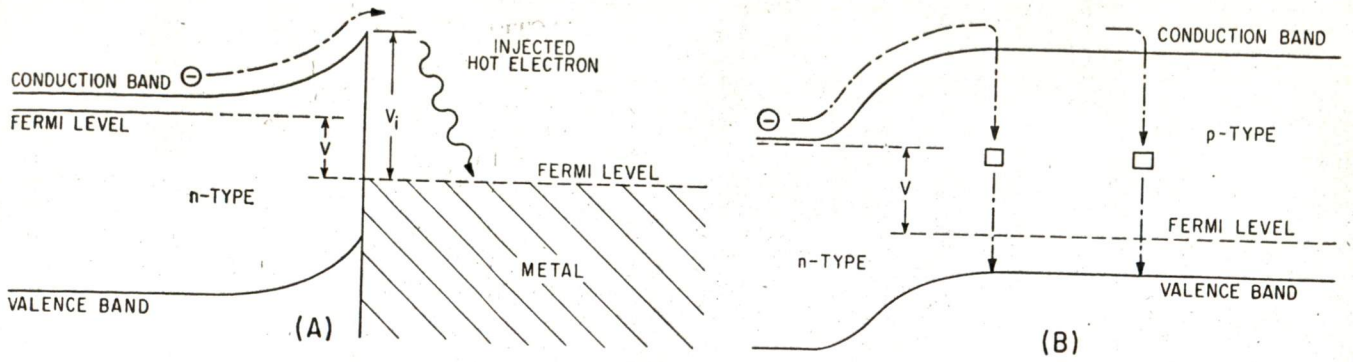
PERFORMANCE—Recovery time as a function of minority carrier lifetime for these diodes is difficult to measure. The lifetime is so low that its influence is readily obscured by diode and circuit impedances and by transient response anomalies in the associated instrumentation. It has been found best to characterize the recovery time for these diodes relative to a sinusoidal excitation[1]. The circuit is shown in Fig. 3, the resulting oscilloscope patterns in Fig. 4.

The effective minority carrier storage can be related to the amplitude of the negative spike. Diode capacitance causes the baseline to tilt, and so capacitive conduction can be separated from storage conduction by measuring the spike amplitude from this tilted reference line, as shown in upper Fig. 4.

Measurement using this technique is not completely quantitative, but it gives a convenient index of the diode recovery that corresponds to most applications. If, for example, the signal generator and amplifier are adjusted to 53 mc with sufficient output to produce a peak forward current flow of 20 ma and scope gain set to give a 5-cm deflection for the positive peak, then the amplitude of the negative spike (read as shown) will be related to lifetime as $\tau = 500$ ps/per cm for deflections less than 1.5 cm. This value is an effective rather than a true minority carrier lifetime. It is essentially the product of true minority carrier lifetime by the ratio of minority to majority carriers that is associated with forward conduction. This ratio is made smaller with reduced barrier height and reduced substrate resistivity. Currently available diodes have effective lifetimes below the resolution capability of this measurement (<50

WHAT'S A HOT CARRIER?

Nothing radioactive! In rectifying metal-semiconductor contacts, current flow is predominantly by majority carriers. When such a diode is forward biased, these majority carriers are injected into the metal and have much greater velocity than the thermal electrons—hence the name hot-carrier diodes. With minority carrier storage virtually eliminated, these diodes surpass conventional p-n junction types at high frequencies



ENERGY LEVELS *in metal-semiconductor junction (A) and* p-n *junction of normal silicon diode (B)—Fig. 2*

ps). Sufficiently high values of forward current can cause minority carrier injection, and storage.

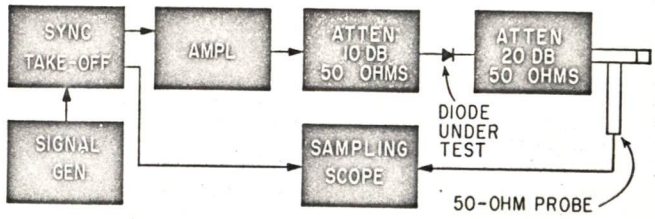
Hot carrier diodes now available are listed in the table. The static characteristics of these diodes, both in forward and reverse, are similar to conventional *p-n* junction diodes. The type 2001 resembles a conventional silicon *p-n* junction diode, and the type 2101 resembles germanium, as shown in Fig. 5.

To take full advantage of their fast response capability, care is necessary in mounting these diodes. Minimum possible lead length will reduce to a minimum performance degradation owing to shunt capacitance and series inductance. The self-inductance of the present package is approximately 3 nh. A lower inductance package is under development.

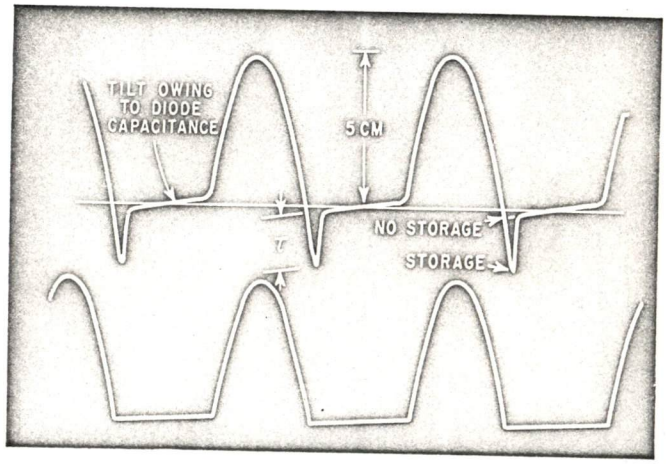
In general the same considerations that apply to the application of conventional *p-n* junction diodes will apply to the hot carrier diode. The differences between them is confined to the lower storage and wider choice of barrier height that is associated with the hot carrier diode. Accordingly, hot carrier diodes might be substituted in many existing circuits without design modifications being required, and with a substantial gain in performance.

NO DELAYS—Minority carrier storage in the hot carrier diode is so low that the turn-on and turn-off delays that are present in conventional *p-n* junction diodes will be essentially eliminated. Accordingly, hot carrier diodes can be used effectively in those pulse and high frequency applications where lag-free response is required, such as detection, mixing and limiting at microwave and high frequencies. Within fractional nanosecond limits they can be used for clamping and gating rapidly.

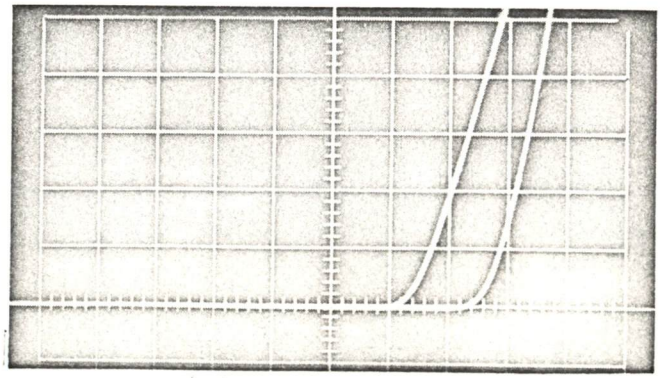
Freedom of choice in barrier height leads to applications that may or may not also require low storage. Low barrier, low storage diodes permit an approach to ideality for detection sensitivity, mixing efficiency, and harmonic generating capability because of the improved impedance matches. Also, tunnel diode logic circuits which require very low turn-on voltage for the associated diode may become possible.



EFFECTIVE LIFETIME MEASUREMENT *of diode is made with this equipment setup—Fig. 3*



HIGH-SPEED SWITCHING DIODE *(top) compared with hot carrier diode. Sweep speed is 10 nsec per cm, vertical sensitivity is 20 ma per cm. Applied signal is a 30-Mc sine wave—Fig. 4*



STATIC CHARACTERISTICS *of type 2101 (left) and type 2001 show that the former starts conduction at about 0.3 v and latter around 0.5 v. Vertical calibration is 2 ma per division and horizontal is 0.2 v per division—Fig. 5*

REFERENCE

(1) S. M. Krakauer, Harmonic Generation, Rectification, and Lifetime Evaluation with the Step Recovery Diode, *Proc. IRE*, p 1665, July 1962.

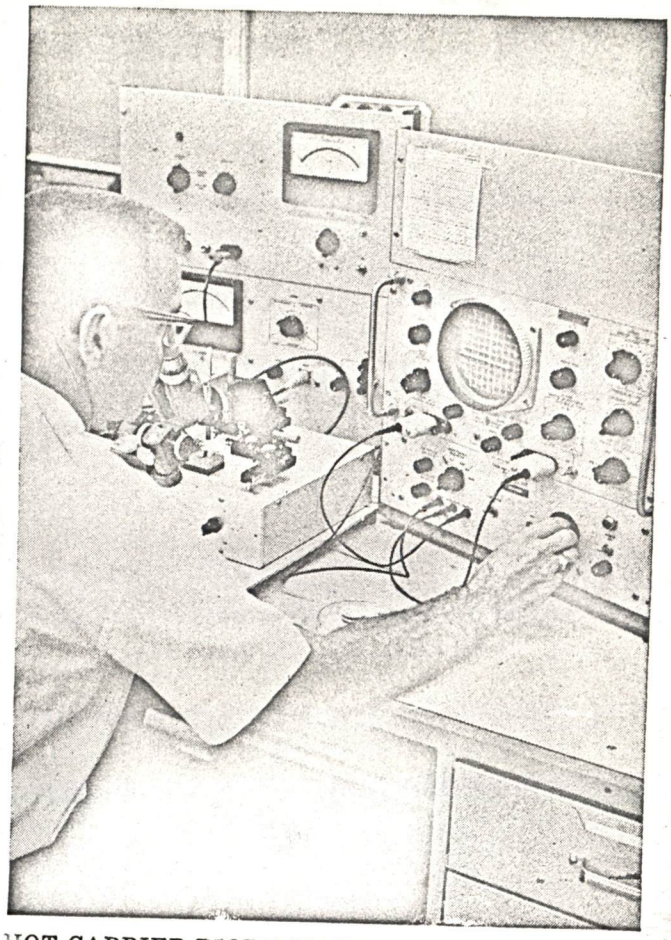
-4702-
-4703-
A
Grey

ANOTHER NEW COMPONENT

HOT CARRIER DIODES Switch in Picoseconds

Metal-semiconductor diodes increase switching speed now limited in **p-n** *junctions by minority-carrier storage. Devices need very pure materials and improved epitaxy*

By S. M. KRAKAUER, Applications Engineer
S. W. SOSHEA, Project Leader
HP Associates, Palo Alto, California

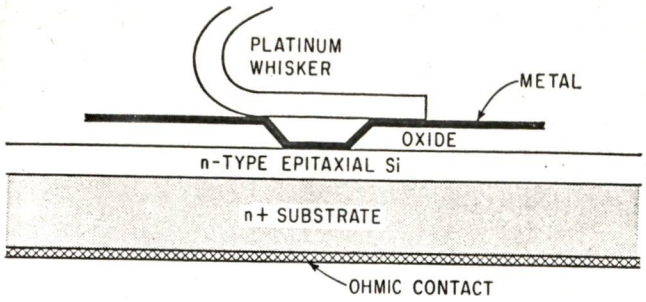


HOT CARRIER DIODE WAFER *undergoing evaluation of leakage current by project engineer Bill Baker*

EPITAXIAL SILICON hot carrier diodes were introduced by HP Associates only recently, but many of the principles of rectifying metal contacts have been known for decades. The great advances in germanium and silicon *p-n* junction devices in the last 15 years have tended to obscure the potential of metal-semiconductor contacts. The *p-n* junction diodes are, however, approaching the limit of their high-frequency performance, because of storage of minority carriers. Since minority carrier storage is virtually eliminated, metal-semiconductor diodes show renewed usefulness. The development of the modern hot carrier diode was made possible by the availability of very pure semiconductors, by improved techniques of surface cleaning and passivation, and by the epitaxial construction method.

A hot carrier diode can be made in a variety of ways. A typical epitaxial type is shown in Fig. 1. Experimental models have been made on silicon using evaporated gold, platinum, palladium, silver and many other metals. Both hot electron (on *n*-type silicon) and hot hole (on *p*-type silicon) forms are possible, but the hot electron type is generally preferable because the higher electron mobility gives better high frequency performance.

OPERATION—Hot carrier diode operation and the distinction between it and a *p-n* junction can be understood most clearly by means of the appropriate electron energy diagrams. Figure 2A shows the electron energy diagram of a hot electron diode with a Schottky-type barrier, and Fig. 2B shows the corresponding diagram for an abrupt *p-n* junction diode with the *n*-type region more heavily doped than the *p*-type region. When the hot carrier diode is forward biased, the electrons in the *n*-type semiconductor



HOT ELECTRON DIODE *construction differs from that of hot hole where substrate would be* **p-plus** *material and epitaxial layer* **p-type** *silicon—Fig. 1*

diffuse over the barrier and are injected into the metal.

The injected hot electrons interact with the lattice and the electrons of the metal and when the diode is reverse biased, these hot electrons are unable to surmount the barrier, so they do not contribute to the stored charge. When, however, the *p-n* junction is forward biased, the electrons diffuse into the *p*-type region and build up to a concentration that is limited by the rate of carrier recombination, as depicted in Fig. 2B. When the *p-n* junction is reverse biased, the stored electrons (minority carriers) flow back into the *n*-type region, thus lowering the rectification efficiency if the diode is used as a detector, or increasing the reverse recovery time if it is used as a switching diode.

The current-voltage characteristics of hot carrier diodes can be described very closely by the ideal diode equation

$$I_F = I_s\ [\exp(qV/kT) - 1]$$

in which the saturation current I_s is proportional to $\exp(-qV_i/kT)$. The type of metal can be conveniently selected to have an internal barrier V_i from 0.3 to 0.8 volt, corresponding to a saturation current, for a typical diode size (about 6×10^{-6} cm^2), from 10^{-11} to 10^{-8} amp. The junction capacitance of the hot carrier diode varies as the inverse square root of voltage and is only slightly dependent on V_i. This combination of characteristics is analogous to a family of *p-n* junctions of incrementally varying energy band gap and provides the circuit designer with an added degree of freedom that was not previously available. The reverse characteristics of hot carrier diodes appear very similar to those of *p-n* junction diodes. The reverse leakage current increases with reverse voltage gradually, owing to the internal Schottky effect, until the avalanche multiplication voltage is reached.

Hot carrier diodes are similar in concept and in operation to the ideal point-contact diode in which the contact is neither formed, alloyed nor bonded. Both the hot carrier diode and the ideal point contact diode employ a Schottky barrier, but there are many notable differences. Being of much larger area, the hot carrier diode has larger capacitance than the point contact, but it can handle greater power and is less sensitive to current transients than is the ideal point contact. The hot carrier diode, furthermore, is more stable mechanically and has more nearly ideal and reproducible electrical characteristics.

CHARACTERISTICS OF HOT-CARRIER DIODES

	HPA-2001			HPA-2101		
	Min	Typ	Max	Min	Typ	Max
Forward Current I_F in ma (at $V_F = 1$ v)	20	30		20	30	
Forward Current I_F in μa (at $V_F = 0.4$ v)	0.5	1.0		2,000	4,000	
Capacitance C_o in pf (at $V_R = 0$)		0.85	1.0		0.95	1.1
Breakdown Voltage V_B in v (at $I_R = 10$ μa)	15	25		15	25	
Leakage Current I_r in na (at $V_R = 3$ v)		1	10		30	100

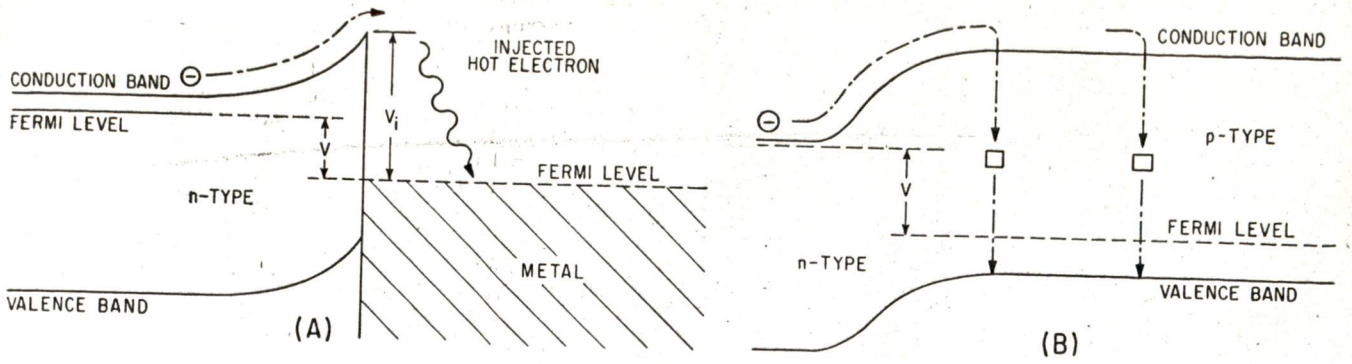
PERFORMANCE—Recovery time as a function of minority carrier lifetime for these diodes is difficult to measure. The lifetime is so low that its influence is readily obscured by diode and circuit impedances and by transient response anomalies in the associated instrumentation. It has been found best to characterize the recovery time for these diodes relative to a sinusoidal excitation[1]. The circuit is shown in Fig. 3, the resulting oscilloscope patterns in Fig. 4.

The effective minority carrier storage can be related to the amplitude of the negative spike. Diode capacitance causes the baseline to tilt, and so capacitive conduction can be separated from storage conduction by measuring the spike amplitude from this tilted reference line, as shown in upper Fig. 4.

Measurement using this technique is not completely quantitative, but it gives a convenient index of the diode recovery that corresponds to most applications. If, for example, the signal generator and amplifier are adjusted to 53 mc with sufficient output to produce a peak forward current flow of 20 ma and scope gain set to give a 5-cm deflection for the positive peak, then the amplitude of the negative spike (read as shown) will be related to lifetime as $\tau = 500$ ps/per cm for deflections less than 1.5 cm. This value is an effective rather than a true minority carrier lifetime. It is essentially the product of true minority carrier lifetime by the ratio of minority to majority carriers that is associated with forward conduction. This ratio is made smaller with reduced barrier height and reduced substrate resistivity. Currently available diodes have effective lifetimes below the resolution capability of this measurement (<50

WHAT'S A HOT CARRIER?

Nothing radioactive! In rectifying metal-semiconductor contacts, current flow is predominantly by majority carriers. When such a diode is forward biased, these majority carriers are injected into the metal and have much greater velocity than the thermal electrons—hence the name hot-carrier diodes. With minority carrier storage virtually eliminated, these diodes surpass conventional p-n junction types at high frequencies



ENERGY LEVELS *in metal-semiconductor junction (A) and* p-n *junction of normal silicon diode (B)—Fig. 2*

ps). Sufficiently high values of forward current can cause minority carrier injection, and storage.

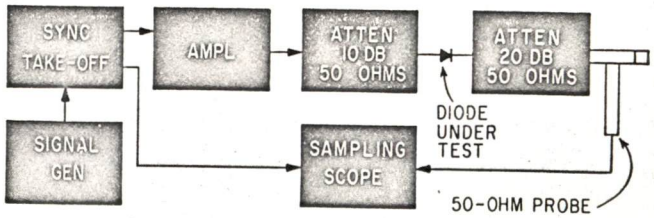
Hot carrier diodes now available are listed in the table. The static characteristics of these diodes, both in forward and reverse, are similar to conventional *p-n* junction diodes. The type 2001 resembles a conventional silicon *p-n* junction diode, and the type 2101 resembles germanium, as shown in Fig. 5.

To take full advantage of their fast response capability, care is necessary in mounting these diodes. Minimum possible lead length will reduce to a minimum performance degradation owing to shunt capacitance and series inductance. The self-inductance of the present package is approximately 3 nh. A lower inductance package is under development.

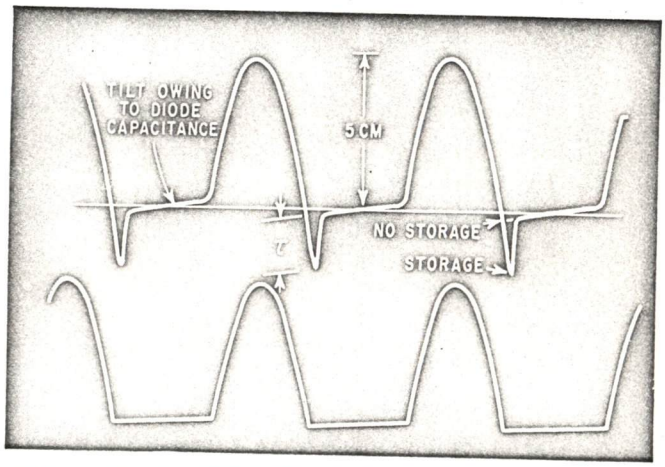
In general the same considerations that apply to the application of conventional *p-n* junction diodes will apply to the hot carrier diode. The differences between them is confined to the lower storage and wider choice of barrier height that is associated with the hot carrier diode. Accordingly, hot carrier diodes might be substituted in many existing circuits without design modifications being required, and with a substantial gain in performance.

NO DELAYS—Minority carrier storage in the hot carrier diode is so low that the turn-on and turn-off delays that are present in conventional *p-n* junction diodes will be essentially eliminated. Accordingly, hot carrier diodes can be used effectively in those pulse and high frequency applications where lag-free response is required, such as detection, mixing and limiting at microwave and high frequencies. Within fractional nanosecond limits they can be used for clamping and gating rapidly.

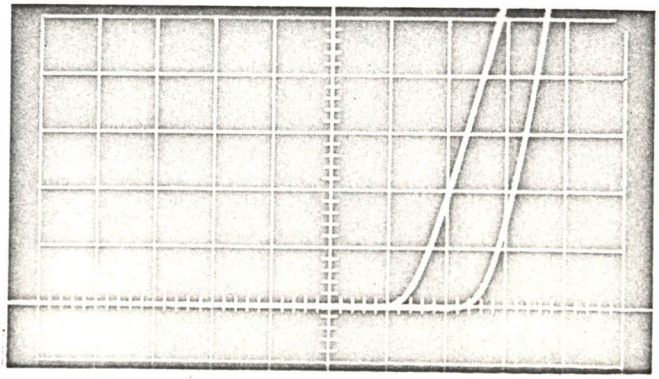
Freedom of choice in barrier height leads to applications that may or may not also require low storage. Low barrier, low storage diodes permit an approach to ideality for detection sensitivity, mixing efficiency, and harmonic generating capability because of the improved impedance matches. Also, tunnel diode logic circuits which require very low turn-on voltage for the associated diode may become possible.



EFFECTIVE LIFETIME MEASUREMENT *of diode is made with this equipment setup—Fig. 3*



HIGH-SPEED SWITCHING DIODE *(top) compared with hot carrier diode. Sweep speed is 10 nsec per cm, vertical sensitivity is 20 ma per cm. Applied signal is a 30-Mc sine wave—Fig. 4*



STATIC CHARACTERISTICS *of type 2101 (left) and type 2001 show that the former starts conduction at about 0.3 v and latter around 0.5 v. Vertical calibration is 2 ma per division and horizontal is 0.2 v per division—Fig. 5*

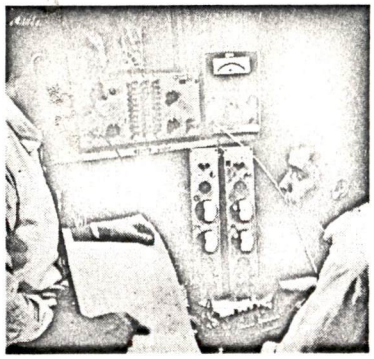
REFERENCE

(1) S. M. Krakauer, Harmonic Generation, Rectification, and Lifetime Evaluation with the Step Recovery Diode, *Proc. IRE*, p 1665, July 1962.

Designing Transistorized Differential Amplifiers

150000

Equations are developed and techniques discussed for the design of transistorized, direct-coupled differential amplifiers having good dc stability and high common mode rejection. In a second and concluding article the performance of an experimental amplifier designed with the recommended procedures will be discussed.



W. M. DeMatteis and J. W. Halligan
Philco Corp.
Lansdale Div.
Lansdale, Pa.

AN EFFICIENT method of amplifying low-level, low-frequency signals is by a transistorized, direct-coupled differential amplifier. Design procedures for such an amplifier are given here, along with techniques for designing a device capable of amplifying very low frequency signals of less than 1 mv.

Convential amplifiers often cannot be used because extraneous signals introduced by ambient temperature variations and poorly regulated supply voltages limit the minimum signal. Although modulation techniques have been used to surmount this difficulty, they sometimes introduce problems more serious than those encountered in straight dc amplification.

Two primary considerations in the design of a differential amplifier capable of amplifying very low frequency signals are:

- A high degree of dc stability.
- A high common-mode rejection.

A transistorized amplifier suitable for amplifying very low frequency signals must possess a high degree of dc stability. This is necessary if the amplifier is to amplify adequately true inputs while rejecting extraneous dc signals introduced by temperature variations and other undesirable effects. In a properly designed differential amplifier, the output signal is proportioned to the difference between input signals. Variations in environmental conditions intro-

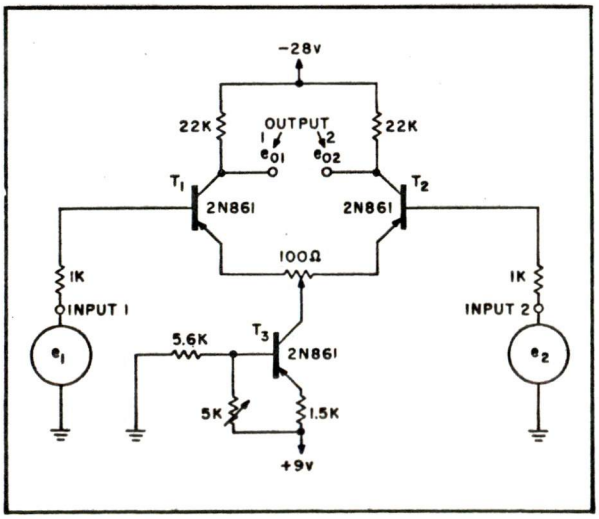


Fig. 1. Direct coupled differential amplifier. Forward bias is applied to emitters of differential transistors through *T3*, biased as common-base current source.

REPRINTED WITH PERMISSION OF ELECTRONIC DESIGN AUGUST 2 AND 16, 1962

duces equivalent signals at both input terminals that are cancelled in the output.

Well-matched transistors are required for high dc stability. The simple differential amplifier circuit shown in Fig. 1, with one of its input terminals maintained at ground potential, inherently possesses a high degree of dc stability.

High Common-Mode Rejection Is a Major Requirement

The differential amplifier circuit has further utility in that two isolated input terminals are available. It thus is useful in providing an output proportional to the small difference between two very large signals. The degree to which the amplifier performs this function is indicated by the quantity "common mode rejection." This is a measure of the ability of the amplifier to reject a signal common to both its input terminals.

In addition to a high degree of dc stability and a high common mode rejection, it is also desirable that both the input and output terminals be at zero dc potential. Several differential amplifiers then can be conveniently cascaded to provide very high gain dc amplifiers. Also, the fact that both input and output terminals are at the same potential will allow the introduction of heavy dc feedback between output and input. This configuration is particularly suited to such applications as operational amplifiers.

Referring to the circuit of Fig. 1, several comments can be made regarding the general procedure for designing the dc bias networks. To obtain a high gain per differential stage, the load resistors should be relatively large. Also, the input impedance appearing at input *1* with input *2* shorted to ground is approximately twice the input impedance of a single common-emitter amplifier at the same operating point conditions. Therefore, to obtain a relatively high input impedance and to allow the use of high load resistance with reasonable values of supply voltage, the differential amplifier transistors should be biased at relatively low values of current.

Silicon Transistors Provide Stable Operation

The use of silicon transistors allows extremely stable operation—even at elevated temperatures—at emitter currents of the order of 100 μa or less per transistor. So that the input terminals will be at ground potential, the required forward bias must be applied to the emitter terminals of the respective units. For a high common mode rejection the impedance between the emitters and ground should be extremely high.

Thus, the emitter bias current should be supplied from a current source. This is done readily through the use of a third transistor in the emitter circuits. It is biased as a common-base current source so that the effective impedance seen from the emitters of the differential transistors is the very high output impedance of a common-base transistor.

The emitter bias current is applied to the emitters through a small potentiometer connected between the emitters of the two differential transistors. This permits adjustments of minor imbalance between the transistors or other components.

For true differential operation the base of each transistor in a differential circuit should see approximately the same resistance. Generally, no matching of source resistance is required for values of less than several thousand ohms if silicon transistors are used in the amplifier.

For higher values of source resistance, however, the source resistance seen at each input terminal should be well matched. Since the base-to-collector reverse saturation and leakage current (I_{CBO}) of each transistor flows in its respective source resistance, it introduces a small voltage that appears as an input signal to each transistor. However, for silicon transistors with extremely low I_{CBO}'s, the effect is negligible for source resistances of less than several thousand ohms.

In the circuit of Fig. 1, an output signal proportional to the difference between the signals appearing at each input is available at each collector. However, these output signals are 180 deg out of phase with each other. Consequently, an additional gain of two can be realized in the amplifier by defining the output as the difference signal appearing between the collector of transistor *1* and the collector of transistor *2*.

The use of a double-ended output has the added advantage that any difference in the small signal gain between transistor *1* and transistor *2* is cancelled in the output.

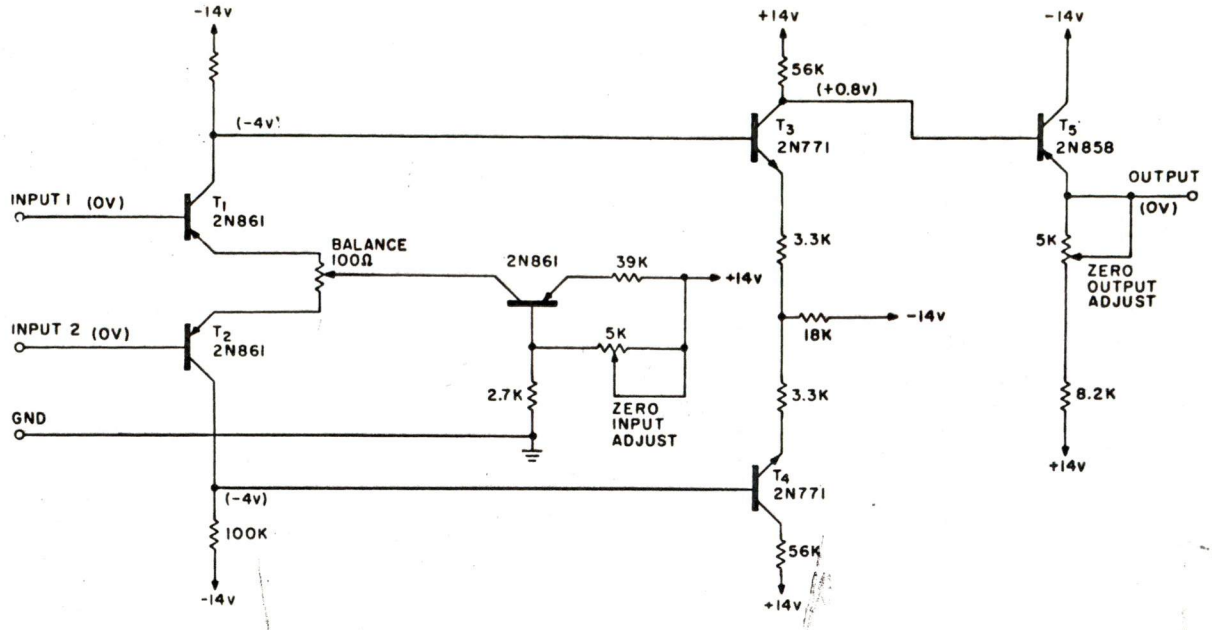


Fig. 2. Multi-stage differential amplifier using complementary transistors to permit both input and output terminals to be biased at ground potential. Stages thus can be cascaded for high gain.

Input, Output at Zero DC Potential: Why and How

Since the base of each transistor in the differential circuit has been established at ground potential, the collector of each transistor is, of necessity, not at ground potential for proper biasing. Thus, if the differential amplifier circuit contains only a single stage of amplification, a voltage translation network must be introduced in order to return the output terminals to ground potential under zero input-signal conditions.

For single-ended output this may be accomplished either by a passive resistor network translating the voltage to ground potential, or by an emitter follower circuit. However, if more than one differential stage of amplification is required, the necessary voltage translation is much more easily accomplished by constructing the second stage from a complementary type of transistor.

This second stage also is a differential amplifier and the double-ended output of the first stage can be used to drive the second stage directly. The technique of voltage translation via complementary transistors is illustrated in a three-stage differential amplifier circuit, Fig. 2. Here the input stage uses pnp transistors and the second stage uses npn transistors.

The stages are direct-coupled. By using differential amplifiers in both stages, imbalance between the units in the first stage is attenuated by the common-mode rejection of the input circuits in the second stage. The third stage is a pnp transistor used as an emitter-follower to a single-ended output.

Use of complementary-type transistors in alternating stages allows both the input and output terminals to be readily biased at ground potential. The technique is applicable to any number of cascaded stages, alternating the type of transistor in each successive stage.

After several stages of differential amplification the signal is at a sufficiently high level so that single-ended outputs may be used from either side of the last differential amplifier. For most direct-coupled amplifier applications a very low output impedance is desirable. Therefore, the final stage in any such multistage differential amplifier is usually an emitter-follower circuit.

In a derivation section, included with the second part of this article in the next issue, the differential output voltage for the circuit of Fig. 3 is shown to be:

$$e_{o_1} - e_{o_2} = \frac{2R_L}{R_s\left[\frac{h_{FE_1} + h_{FE_2}}{h_{FE_1}\,h_{FE_2}}\right] + R_E} \times \left\{(e_2 - e_1) + (\Delta V_{BE} + R_s \Delta I_{CO}) - (R_s I_{CO_2} + V_{EE} - V_{BE_2} - e_2)\ \frac{(R_{E_1} - R_{E_2})}{2R_{EE}}\right\} + R_L \Delta I_{CO} \quad (1)$$

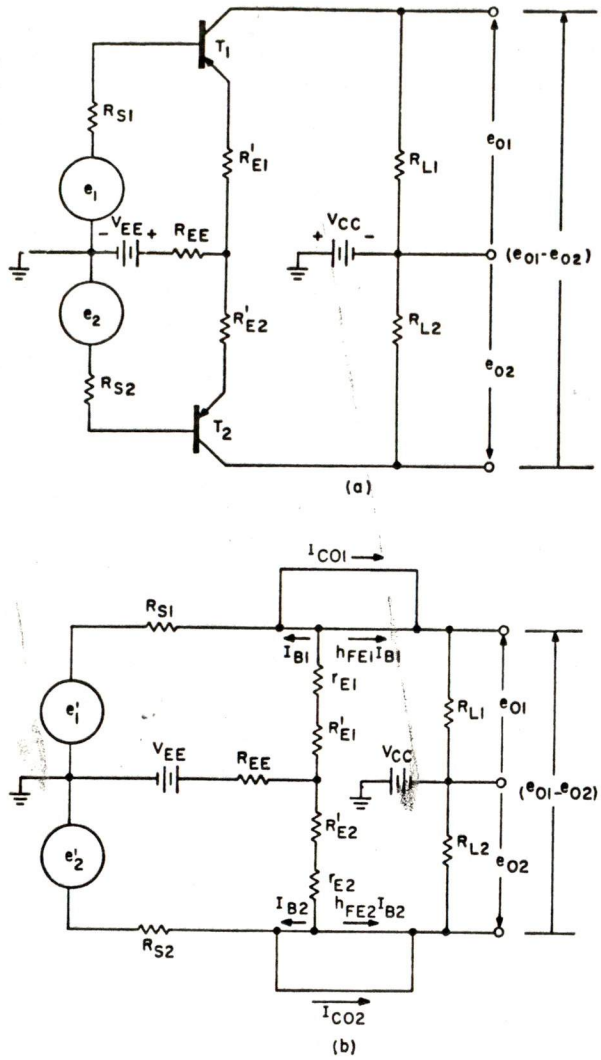


Fig. 3a. Basic circuit for differential amplifier with **(b)** its approximate equivalent circuit.

Note that the expression contains four terms. The first is the desired output and is proportional to the difference between the input voltages e_1 and e_2.

The second is the output due to differences in V_{BE} and I_{CO} between T_1 and T_2. In order for this to be zero, V_{BE_1} and V_{BE_2} must be equal, I_{CO_1} and I_{CO_2} must be equal, and they must remain equal with variations in ambient temperature.

The third term is the error voltage and always appears in the output of a differential amplifier. If R_{E_1} and R_{E_2} are parts of a potentiometer connected between the emitters of the two transistors, the potentiometer may be adjusted to produce zero differential output for zero input at any reference temperature. That is, the dc balance is obtained at a particular temperature if

$$\Delta V_{BE} + R_S \Delta I_{CO} = \frac{V_{EE}}{2R_{EE}} (R_{E_1} - R_{E_2}), \qquad (2)$$

assuming that $V_{EE} >> R_S I_{CO_2} - V_{BE_2}$.

Any drift of the output after balance will be due to an unequal variation of V_{BE} and I_{CO}, and unequal junction temperatures between the transistors. The junction temperatures of the transistors may be made approximately equal by mounting the transistors in a simple heat sink.

For reasonable values of R_L, the fourth term of Eq. 1, i.e., $R_L \Delta I_{CO}$, usually may be neglected for silicon transistors.

The above holds true if R_{S_1} and R_{S_2} are equal. If they are not, then (1) becomes

$$e_{o_1} - e_{o_2} = \frac{2R_L}{\frac{R_{S_2}}{h_{FE_2}} + \frac{R_{S_1}}{h_{FE_1}} + R_E} \times \left\{ (e_2 - e_1) + (\Delta V_{BE} + \Delta R_S I_{CO}) - [R_{S_2} I_{CO_2} + V_{EE} - V_{BE_2} - e_2] \left[\frac{R_{E_1} - R_{E_2}}{2R_{EE}} + \frac{\frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}}}{2R_{EE}} \right] \right\}, \qquad (3)$$

assuming that $\frac{R_{S_2}}{h_{FE_2}} + R_{E_2} << 2R_{EE}$.

In order to balance the output with zero input,

$$\Delta V_{BE} + \Delta R_S I_{CO} = \frac{V_{EE}}{2R_{EE}} \left(R_{E_1} - R_{E_2} + \frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}} \right) \qquad (4)$$

The V_{BE} of silicon transistors decreases as temperature is increased. Fig. 4 is a plot of V_{BE} vs temperature of four 2N861 Silicon Precision Alloy Transistors (SPAT*). ΔV_{BE} is the magnitude of the maximum observed difference in V_{BE} between any two of the transistors. If ΔV_{BE} remains constant throughout the temperature range, then the differential output due to difference in V_{BE} will have a constant value different from zero. As shown in Fig. 4, ΔV_{BE} increases by approximately 1 mv between −25 C and +125 C. The effect of ΔV_{BE} may be made small by adjusting the emitter balance potentiometer so that zero differential output is obtained at any particular temperature. The output then will remain essentially constant through-

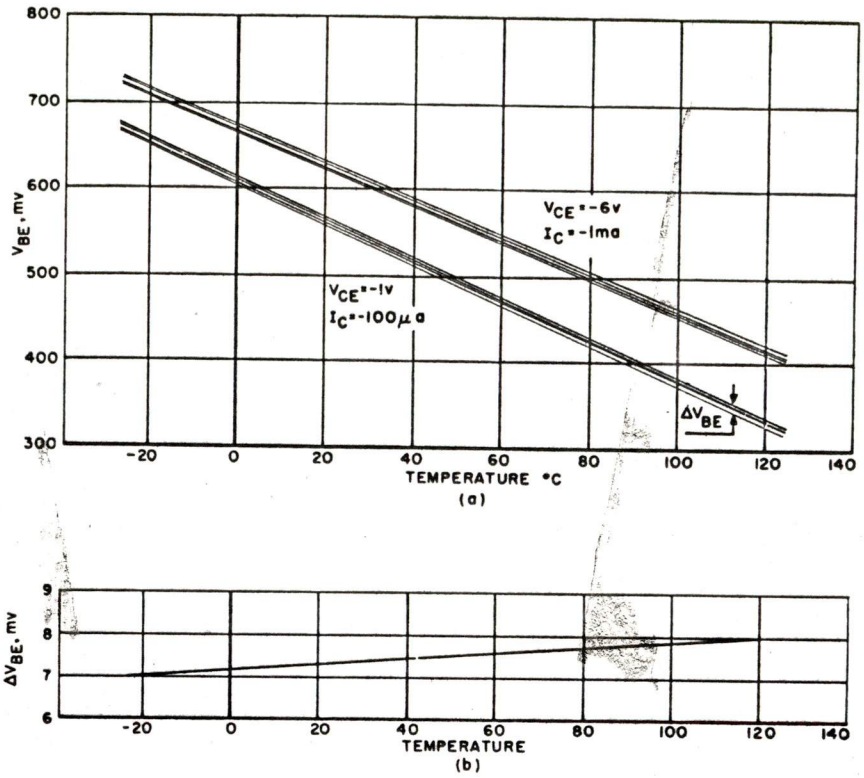


Fig. 4a. Variation in V_{BE} with temperature for four 2N861 silicon transistors. (**b**) ΔV_{BE} is the magnitude of the maximum observed difference in V_{BE} for any two of the transistors.

out the temperature range, differing from zero only by the relatively small change in ΔV_{BE} of approximately 6 μv/C.

It should be noted that a difference of 0.01 C between the junction temperatures of the transistors results in an equivalent input drift of approximately 20 μv, even for a perfectly matched pair of transistors. If the junction temperatures of the two transistors are held approximately equal by using an adequate heat sink, the drift of the amplifier can be reduced appreciably.

Common Mode Rejection Depends on V_{BE} Match

Common mode rejection (CMR) may be defined as the ratio of the magnitude of the smaller input signal to the magnitude of the output signal. CMR exhibits the following relationship (derivation is included in Part 2 of this article):

$$CMR \geqslant \frac{V_{EE}}{\Delta V_{BE} + \Delta R_S I_{CO}}. \qquad (5)$$

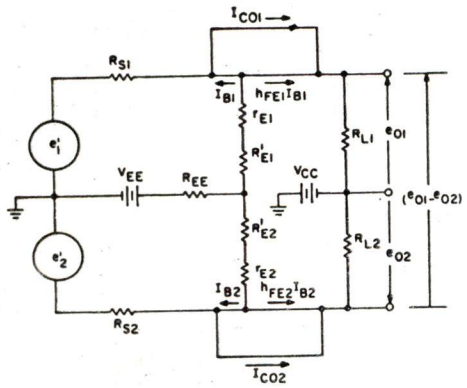
Eq. (5) will hold true under balanced conditions, i.e., if Eq. (4) is satisfied. Eq. (5) also signifies that for any particular mismatch in transistor parameters a desired common-mode rejection may be obtained by selecting V_{EE} for the desired rejection.

The differential gain of the amplifier stage is:

$$G_d = \frac{2R_1}{\frac{R_{S_2}}{h_{FE_2}} + \frac{R_{S_1}}{h_{FE_1}} + R_E}. \qquad (6)$$

Derivations of Differential Output Voltage And Common-Mode Rejection

Derivations of differential output voltage and commonmode rejection are given here.



The approximate equivalent circuit of the differential amplifier is shown. The differential output voltage is $e_{O_1} - e_{O_2}$. The output voltage (e_{O_1}) of T_1 is:

$$e_{O_1} = \frac{RL}{\frac{R_S\,(h_{FE_1}+h_{FE_2})}{h_{FE_1}\;h_{FE_2}} + R_E} \left\{(e_2 - e_1) + \Delta V_{BE} + R_S \Delta I_{CO} \frac{R_{E_2}}{R_{EE}}\left[e_1 + V_{BE_1} - R_S I_{CO_1} - V_{EE}\right]\right\} + I_{CO_1} R_L - V_{CC}, \quad (1)$$

where $R_{E_1} = R'_{E_1} + r_{e_1}$;

$R_{E_2} = R_{E'_2} + r_{e_2}$;

$R_E = R_{E_1} + R_{E_2}$;

$\Delta V_{BE} = V_{BE_2} - V_{BE_1}$; $\Delta I_{CO} = I_{CO_1} - I_{CO_2}$;

and assuming that:

$(h_{FE} + 1) \approx h_{FE}$;

$R_S = R_{S_1} = R_{S_2}$;

$R_L = R_{L_1} = R_{L_2}$;

$R_{EE} >> R_E + R_S$;

$\frac{R_{E_2}}{2R_{EE}} << 1$.

A similar expression is obtained for the output voltage (e_{O_2}) of T_2. The differential output voltage $(e_{O_1} - e_{O_2})$ is:

$$e_{O_1} - e_{O_2} = \frac{2R_L}{R_S\left(\frac{h_{FE_1}+h_{FE_2}}{h_{FE_1}h_{FE_2}}\right) + R_E} \left\{(e_2 - e_1) + (\Delta V_{BE} + R_S \Delta I_{CO}) - [\,R_S I_{CO_2} + V_{EE} - V_{BE_2} - e_2\,]\left[\frac{R_{E_1} - R_{E_2}}{2R_{EE}}\right]\right\} + R_L \Delta I_{CO}.$$

Common Mode Rejection Depends on V_{BE} Match

Common mode rejection (CMR) may be defined as the ratio of the magnitude of the smaller input signal to the magnitude of the output signal. It may be derived as follows:

If $(\Delta V_{BE} + \Delta R_S I_{CO}) - \frac{V_{EE}}{2R_{EE}}\left(R_{E_1} - R_{E_2} + \frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}}\right)$

is adjusted to zero, thus satisfying Eq. 4 the differential output is

$$(e_{O_1} - e_{O_2}) = \frac{R_L}{R_{EE}}\left[2(e_2 - e_1)\left(\frac{1+G_2}{G_1 + G_2}\right) + e_2\left(\frac{G_1 - G_2}{G_1 + G_2}\right)\right], \quad (3)$$

where

$$G_1 = \frac{R_{EE}}{R_{E_1} + \frac{R_{S_1}}{h_{FE_1}}} \text{ and } G_2 = \frac{R_{EE}}{R_{E_2} + \frac{R_{S_2}}{h_{FE_2}}}$$

The first term of Eq. 6 is proportional to the desirable differential gain, and the second term is proportional to the undesirable common-mode gain. The common-mode rejection is the ratio of the differential gain to the common-mode gain, so that

$$CMR = \frac{2\,(1+G_2)}{G_1 - G_2} = \frac{2\left(R_{EE} + \frac{R_{S_2}}{h_{FE_2}} + R_{E_2}\right)}{\left(\frac{R_{S_2}}{h_{FE_2}} + R_{E_2}\right) - \left(\frac{R_{S_1}}{h_{FE_1}} + R_{E_1}\right)}. \quad (4)$$

It may safely be assumed that

$$R_{E_2} + \frac{R_{S_2}}{h_{FE_2}} << R_{EE}.$$

If Eq. 4 now is substituted into Eq. 4 (above), the expression for commonmode rejection becomes:

$$CMR \geqslant \frac{V_{EE}}{\Delta V_{BE} + \Delta R_S I_{CO}}$$

TRANSISTOR direct-coupled differentia. amplifiers can be an efficient means of amplifying low-level, low-frequency signals.

Fig. 1 shows the single stage, three-transistor, direct-coupled differential amplifier. In order to insure that $R_{EE} >> R_S$ and R_L, the effective R_{EE} is simulated by the output resistance of T_3 and is connected to the emitters of T_1 and T_2 through the 100-ohm potentiometer. The effective value of R_{EE} is obtained by calculating the output resistance of the stage containing T_3. For the particular transistor used as T_3, the effective value of R_{EE} was found to be approximately 30 K.

T_1 and T_2 were mounted as close together as possible in a transistor clip* that is designed to keep the temperatures of the two TO-18 transistor cases equal. This clip maintains the junctions of the two transistors at approximately the same temperature. The dc drift of the experimental amplifier was measured at the output terminals for a temperature variation from +25 C to +125 C. At the elevated temperature the output drift was 75 mv, which corresponds to an equivalent input drift of 4.5 μv per deg C. Referring to Fig. 4 in the first part of this article (*ED*, Aug. 2, p 72) it is noted that the maximum change in ΔV_{BE} for a random group of 2N861's over this temperature range is approximately 0.6 mv, or 6 μv/deg C. The fact that the measured dc drift with ambient temperature variations was 4.5 μv/deg C indicates the efficiency of the clip in maintaining the transistor junctions at the same temperature.

*Atlee Corp. Part No. 100-300-1-9

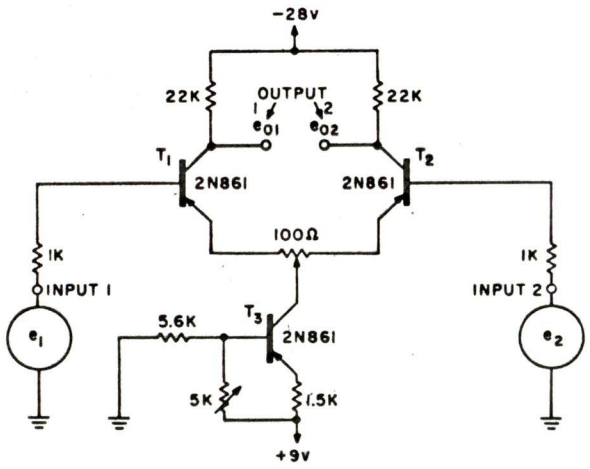


Fig. 1. A single-stage, three transistor, direct-coupled differential amplifier.

Fig. 2 shows the frequency response of the experimental amplifier at +25 C and at +125 C. The low frequency voltage gain of the stage is 45 db at both ambient temperatures. The response at +25 C is 3 db down at 40 kc, decreasing at a rate of 8 db per octave to unity gain at 4 mc. The response at +125 C is 3 db down at 35 kc and is within 2 db of the room temperature response at the higher frequencies.

For the transistors used in the experimental amplifier,

$$h_{FE_1} = 44,$$
$$R_{S_1} = 1\,\text{K} + 700 = 1700,$$
$$h_{FE_2} = 42,$$
$$R_{S_2} = 1\,\text{K} + 460 = 1460.$$

Remembering that R_E includes the dynamic emitter resistances (kT/qI_E),

$$R_E = 100 + 26 + 26 = 152.$$

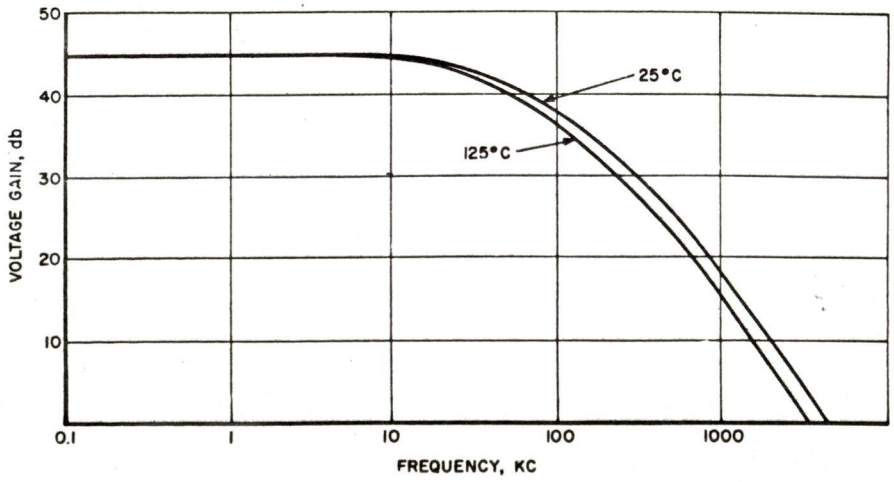


Fig. 2. Response of experimental amplifier at 25 C and 125 C. Low frequency voltage gain is 45 db at both ambients. Response is 3 db down at 40 kc and 35 kc respectively, for the 25-C and 125-C curves.

The differential gain, therefore, is:

$$G_d = \frac{2R_L}{\frac{R_{S_1}}{h_{FE_1}} + \frac{R_{S_2}}{h_{FE_2}} + R_E}$$

$$= \frac{44\text{ K}}{226} = 194 = 45.8\text{ db}$$

This is in agreement with the measured differential gain of 45 db.

Measured CMR Agrees With Calculated Value

The common-mode rejection of the experimental amplifier at 1000 cps was measured as 8000, or 78 db, and linear amplification was obtained up to approximately 20 v peak-to-peak output signal. For the particular transistors used in the amplifier, ΔV_{BE} was measured as 7.5 mv at the operating conditions of

$$V_{CE} = -6\text{ v},$$
$$I_E = +1.0\text{ ma}.$$

Since the current source T_3 supplies the total emitter current of 2.0 ma through the effective value of

$$R_{EE} = 30\text{ K},$$

the effective value of $+V_{EE}$ in the experimental amplifier is (2.0 ma) (30 K) = 60 v.

The calculated *CMR*, therefore, is:

$$CMR = \frac{V_{EE}\ (effective)}{V_{BE}} = \frac{60}{7.5 \times 10^{-3}} = 8000.$$

Again, this is in excellent agreement with the experimental results obtained.

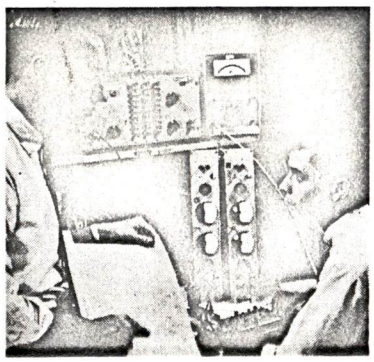
The experimental direct-coupled differential amplifier described here is indicative of the performance that may be obtained by the use of proper design techniques. The performance of the amplifier is in excellent agreement with the design equations developed.

The design equations also are useful in determining the desirable transistor parameters for differential amplifier applications. Eq. (1) shows that the common mode rejection is proportional to the degree to which the base-to-emitter voltages are matched at the desired operating point. Eq. (2) shows that, while a high dc current gain (h_{FE}) is desirable, the differential gain is insensitive to differences between h_{FE_1} and h_{FE_2}. For high differential gain, the source resistance should be low. In addition, Eq. (2) implies an inverse dependence of the differential gain on the V_{BE} match between units. For high gain, R_E should be small. However, the minimum value of the potentiometer connected between the emitters of T_1 and T_2 is equal to $\Delta V_{BE}(max)$ divided by the emitter current of one transistor. ■■

Designing Transistorized Differential Amplifiers

150,000

Equations are developed and techniques discussed for the design of transistorized, direct-coupled differential amplifiers having good dc stability and high common mode rejection. In a second and concluding article the performance of an experimental amplifier designed with the recommended procedures will be discussed.



W. M. DeMatteis and J. W. Halligan
Philco Corp.
Lansdale Div.
Lansdale, Pa.

AN EFFICIENT method of amplifying low-level, low-frequency signals is by a transistorized, direct-coupled differential amplifier. Design procedures for such an amplifier are given here, along with techniques for designing a device capable of amplifying very low frequency signals of less than 1 mv.

Convential amplifiers often cannot be used because extraneous signals introduced by ambient temperature variations and poorly regulated supply voltages limit the minimum signal. Although modulation techniques have been used to surmount this difficulty, they sometimes introduce problems more serious than those encountered in straight dc amplification.

Two primary considerations in the design of a differential amplifier capable of amplifying very low frequency signals are:

- A high degree of dc stability.
- A high common-mode rejection.

A transistorized amplifier suitable for amplifying very low frequency signals must possess a high degree of dc stability. This is necessary if the amplifier is to amplify adequately true inputs while rejecting extraneous dc signals introduced by temperature variations and other undesirable effects. In a properly designed differential amplifier, the output signal is proportioned to the difference between input signals. Variations in environmental conditions intro-

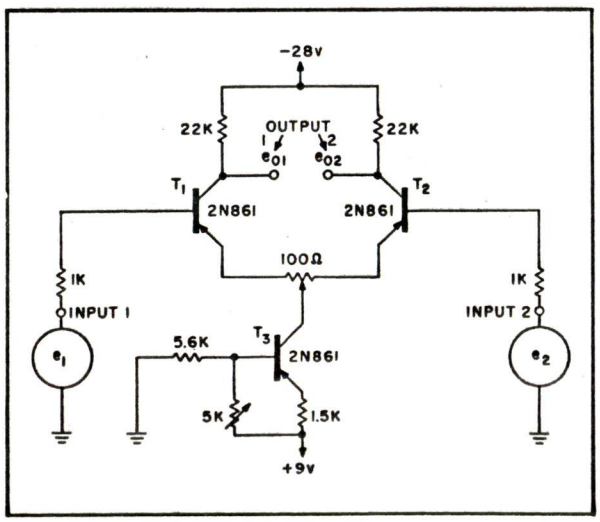


Fig. 1. Direct coupled differential amplifier. Forward bias is applied to emitters of differential transistors through *T3*, biased as common-base current source.

REPRINTED WITH PERMISSION OF ELECTRONIC DESIGN AUGUST 2 AND 16, 1962

duces equivalent signals at both input terminals that are cancelled in the output.

Well-matched transistors are required for high dc stability. The simple differential amplifier circuit shown in Fig. 1, with one of its input terminals maintained at ground potential, inherently possesses a high degree of dc stability.

High Common-Mode Rejection Is a Major Requirement

The differential amplifier circuit has further utility in that two isolated input terminals are available. It thus is useful in providing an output proportional to the small difference between two very large signals. The degree to which the amplifier performs this function is indicated by the quantity "common mode rejection." This is a measure of the ability of the amplifier to reject a signal common to both its input terminals.

In addition to a high degree of dc stability and a high common mode rejection, it is also desirable that both the input and output terminals be at zero dc potential. Several differential amplifiers then can be conveniently cascaded to provide very high gain dc amplifiers. Also, the fact that both input and output terminals are at the same potential will allow the introduction of heavy dc feedback between output and input. This configuration is particularly suited to such applications as operational amplifiers.

Referring to the circuit of Fig. 1, several comments can be made regarding the general procedure for designing the dc bias networks. To obtain a high gain per differential stage, the load resistors should be relatively large. Also, the input impedance appearing at input *1* with input *2* shorted to ground is approximately twice the input impedance of a single common-emitter amplifier at the same operating point conditions. Therefore, to obtain a relatively high input impedance and to allow the use of high load resistance with reasonable values of supply voltage, the differential amplifier transistors should be biased at relatively low values of current.

Silicon Transistors Provide Stable Operation

The use of silicon transistors allows extremely stable operation—even at elevated temperatures—at emitter currents of the order of 100 μa or less per transistor. So that the input terminals will be at ground potential, the required forward bias must be applied to the emitter terminals of the respective units. For a high common mode rejection the impedance between the emitters and ground should be extremely high.

Thus, the emitter bias current should be supplied from a current source. This is done readily through the use of a third transistor in the emitter circuits. It is biased as a common-base current source so that the effective impedance seen from the emitters of the differential transistors is the very high output impedance of a common-base transistor.

The emitter bias current is applied to the emitters through a small potentiometer connected between the emitters of the two differential transistors. This permits adjustments of minor imbalance between the transistors or other components.

For true differential operation the base of each transistor in a differential circuit should see approximately the same resistance. Generally, no matching of source resistance is required for values of less than several thousand ohms if silicon transistors are used in the amplifier.

For higher values of source resistance, however, the source resistance seen at each input terminal should be well matched. Since the base-to-collector reverse saturation and leakage current (I_{CBO}) of each transistor flows in its respective source resistance, it introduces a small voltage that appears as an input signal to each transistor. However, for silicon transistors with extremely low I_{CBO}'s, the effect is negligible for source resistances of less than several thousand ohms.

In the circuit of Fig. 1, an output signal proportional to the difference between the signals appearing at each input is available at each collector. However, these output signals are 180 deg out of phase with each other. Consequently, an additional gain of two can be realized in the amplifier by defining the output as the difference signal appearing between the collector of transistor *1* and the collector of transistor *2*.

The use of a double-ended output has the added advantage that any difference in the small signal gain between transistor *1* and transistor *2* is cancelled in the output.

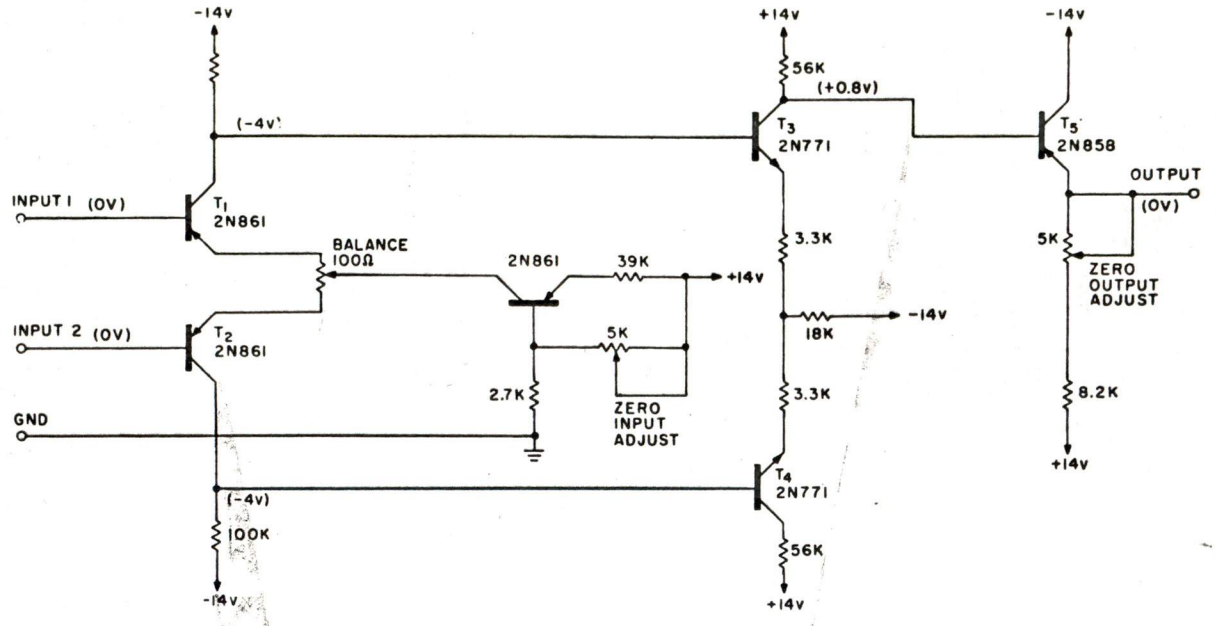


Fig. 2. Multi-stage differential amplifier using complementary transistors to permit both input and output terminals to be biased at ground potential. Stages thus can be cascaded for high gain.

Input, Output at Zero DC Potential: Why and How

Since the base of each transistor in the differential circuit has been established at ground potential, the collector of each transistor is, of necessity, not at ground potential for proper biasing. Thus, if the differential amplifier circuit contains only a single stage of amplification, a voltage translation network must be introduced in order to return the output terminals to ground potential under zero input-signal conditions.

For single-ended output this may be accomplished either by a passive resistor network translating the voltage to ground potential, or by an emitter follower circuit. However, if more than one differential stage of amplification is required, the necessary voltage translation is much more easily accomplished by constructing the second stage from a complementary type of transistor.

This second stage also is a differential amplifier and the double-ended output of the first stage can be used to drive the second stage directly. The technique of voltage translation via complementary transistors is illustrated in a three-stage differential amplifier circuit, Fig. 2. Here the input stage uses pnp transistors and the second stage uses npn transistors.

The stages are direct-coupled. By using differential amplifiers in both stages, imbalance between the units in the first stage is attenuated by the common-mode rejection of the input circuits in the second stage. The third stage is a pnp transistor used as an emitter-follower to a single-ended output.

Use of complementary-type transistors in alternating stages allows both the input and output terminals to be readily biased at ground potential. The technique is applicable to any number of cascaded stages, alternating the type of transistor in each successive stage.

After several stages of differential amplification the signal is at a sufficiently high level so that single-ended outputs may be used from either side of the last differential amplifier. For most direct-coupled amplifier applications a very low output impedance is desirable. Therefore, the final stage in any such multistage differential amplifier is usually an emitter-follower circuit.

In a derivation section, included with the second part of this article in the next issue, the differential output voltage for the circuit of Fig. 3 is shown to be:

$$e_{o_1} - e_{o_2} = \frac{2R_L}{R_s\left[\dfrac{h_{FE_1} + h_{FE_2}}{h_{FE_1}\,h_{FE_2}}\right] + R_E} \times \left\{(e_2 - e_1) + (\Delta V_{BE} + R_s \Delta I_{CO}) - (R_s I_{CO_2} + V_{EE} - V_{BE_2} - e_2)\ \frac{(R_{E_1} - R_{E_2})}{2R_{EE}}\right\} + R_L \Delta I_{CO} \quad (1)$$

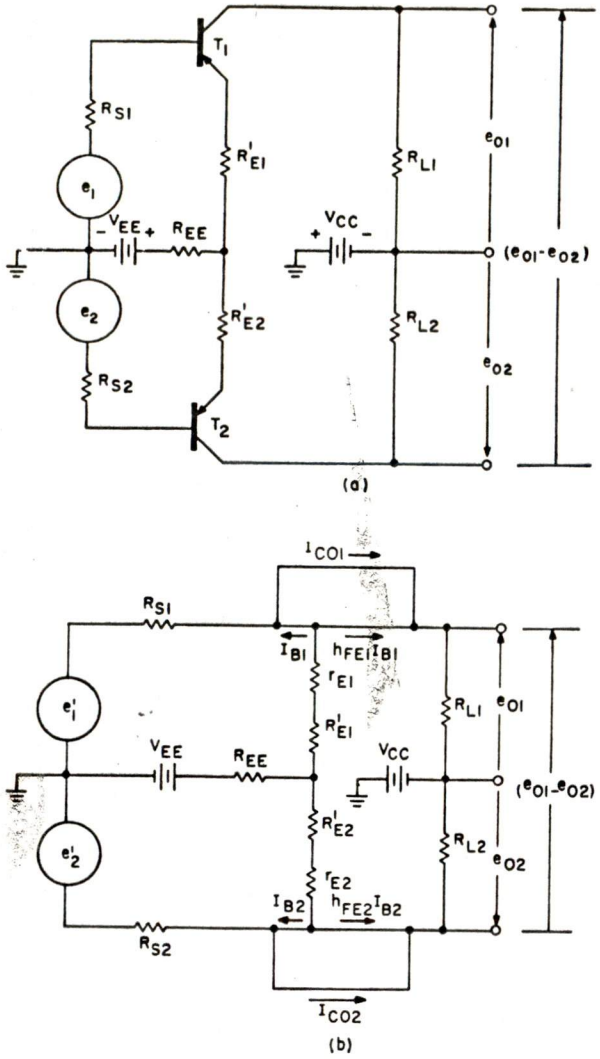


Fig. 3a. Basic circuit for differential amplifier with (b) its approximate equivalent circuit.

Note that the expression contains four terms. The first is the desired output and is proportional to the difference between the input voltages e_1 and e_2.

The second is the output due to differences in V_{BE} and I_{CO} between T_1 and T_2. In order for this to be zero, V_{BE_1} and V_{BE_2} must be equal, I_{CO_1} and I_{CO_2} must be equal, and they must remain equal with variations in ambient temperature.

The third term is the error voltage and always appears in the output of a differential amplifier. If R_{E_1} and R_{E_2} are parts of a potentiometer connected between the emitters of the two transistors, the potentiometer may be adjusted to produce zero differential output for zero input at any reference temperature. That is, the dc balance is obtained at a particular temperature if

$$\Delta V_{BE} + R_S \Delta I_{CO} = \frac{V_{EE}}{2R_{EE}} (R_{E_1} - R_{E_2}), \qquad (2)$$

assuming that $V_{EE} >> R_S I_{CO_2} - V_{BE_2}$.

Any drift of the output after balance will be due to an unequal variation of V_{BE} and I_{CO}, and unequal junction temperatures between the transistors. The junction temperatures of the transistors may be made approximately equal by mounting the transistors in a simple heat sink.

For reasonable values of R_L, the fourth term of Eq. 1, i.e., $R_L \Delta I_{CO}$, usually may be neglected for silicon transistors.

The above holds true if R_{S_1} and R_{S_2} are equal. If they are not, then (1) becomes

$$e_{o_1} - e_{o_2} = \frac{2R_L}{\frac{R_{S_2}}{h_{FE_2}} + \frac{R_{S_1}}{h_{FE_1}} + R_E} \times \left\{ (e_2 - e_1) + (\Delta V_{BE} + \Delta R_S I_{CO}) - [R_{S_2} I_{CO_2} + V_{EE} - V_{BE_2} - e_2] \left[\frac{R_{E_1} - R_{E_2}}{2R_{EE}} + \frac{\frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}}}{2R_{EE}} \right] \right\}, \qquad (3)$$

assuming that $\frac{R_{S_2}}{h_{FE_2}} + R_{E_2} << 2R_{EE}$.

In order to balance the output with zero input,

$$\Delta V_{BE} + \Delta R_S I_{CO} = \frac{V_{EE}}{2R_{EE}} \left(R_{E_1} - R_{E_2} + \frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}} \right) \qquad (4)$$

The V_{BE} of silicon transistors decreases as temperature is increased. Fig. 4 is a plot of V_{BE} vs temperature of four 2N861 Silicon Precision Alloy Transistors (SPAT*). ΔV_{BE} is the magnitude of the maximum observed difference in V_{BE} between any two of the transistors. If ΔV_{BE} remains constant throughout the temperature range, then the differential output due to difference in V_{BE} will have a constant value different from zero. As shown in Fig. 4, ΔV_{BE} increases by approximately 1 mv between −25 C and +125 C. The effect of ΔV_{BE} may be made small by adjusting the emitter balance potentiometer so that zero differential output is obtained at any particular temperature. The output then will remain essentially constant through-

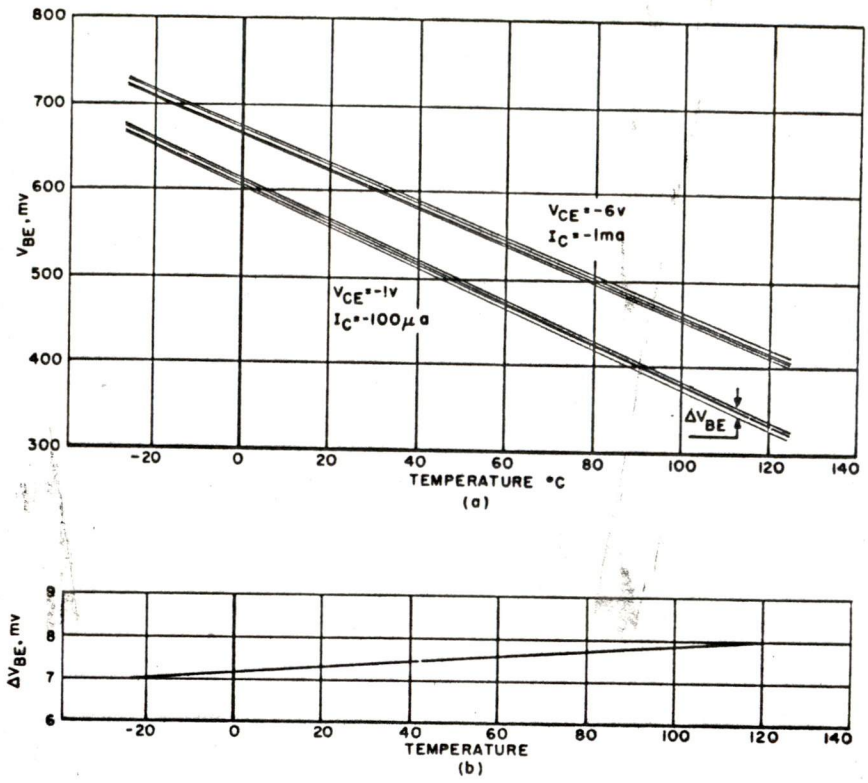


Fig. 4a. Variation in V_{BE} with temperature for four 2N861 silicon transistors. **(b)** ΔV_{BE} is the magnitude of the maximum observed difference in V_{BE} for any two of the transistors.

out the temperature range, differing from zero only by the relatively small change in ΔV_{BE} of approximately 6 μv/C.

It should be noted that a difference of 0.01 C between the junction temperatures of the transistors results in an equivalent input drift of approximately 20 μv, even for a perfectly matched pair of transistors. If the junction temperatures of the two transistors are held approximately equal by using an adequate heat sink, the drift of the amplifier can be reduced appreciably.

Common Mode Rejection Depends on V_{BE} Match

Common mode rejection (CMR) may be defined as the ratio of the magnitude of the smaller input signal to the magnitude of the output signal. CMR exhibits the following relationship (derivation is included in Part 2 of this article):

$$CMR \geq \frac{V_{EE}}{\Delta V_{BE} + \Delta R_S I_{CO}}. \qquad (5)$$

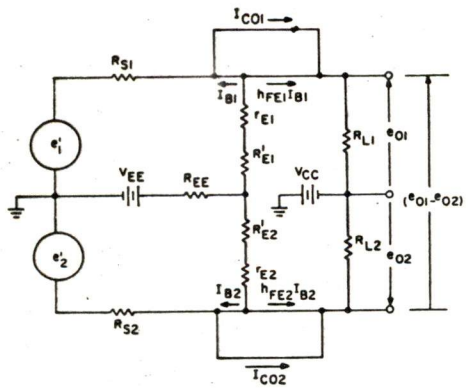
Eq. (5) will hold true under balanced conditions, i.e., if Eq. (4) is satisfied. Eq. (5) also signifies that for any particular mismatch in transistor parameters a desired common-mode rejection may be obtained by selecting V_{EE} for the desired rejection.

The differential gain of the amplifier stage is:

$$G_d = \frac{2R_1}{\frac{R_{S_2}}{h_{FE_2}} + \frac{R_{S_1}}{h_{FE_1}} + R_E}. \qquad (6)$$

Derivations of Differential Output Voltage And Common-Mode Rejection

Derivations of differential output voltage and commonmode rejection are given here.



The approximate equivalent circuit of the differential amplifier is shown. The differential output voltage is $e_{O_1} - e_{O_2}$. The output voltage (e_{O_1}) of T_1 is:

$$e_{O_1} = \frac{RL}{\frac{R_S\,(h_{FE_1}+h_{FE_2})}{h_{FE_1}\;h_{FE_2}} + R_E}$$

$$\left\{(e_2 - e_1) + \Delta V_{BE} + R_S\Delta I_{CO} \right.$$

$$\left. \frac{R_{E_2}}{R_{EE}}\left[e_1 + V_{BE_1} - R_S I_{CO_1} - V_{EE}\right]\right\} \qquad (1)$$

$$+ I_{CO_1} R_L - V_{CC},$$

where $R_{E_1} = R'_{E_1} + r_{e_1}$;

$R_{E_2} = R_E{}'_2 + r_{e_2}$;

$R_E = R_{E_1} + R_{E_2}$;

$\Delta V_{BE} = V_{BE_2} - V_{BE_1}$; $\Delta I_{CO} = I_{CO_1} - I_{CO_2}$;

and assuming that:

$(h_{FE} + 1) \approx h_{FE}$;

$R_S = R_{S_1} = R_{S_2}$;

$R_L = R_{L_1} = R_{L_2}$;

$R_{EE} >> R_E + R_S$;

$$\frac{R_{E_2}}{2R_{EE}} << 1.$$

A similar expression is obtained for the output voltage (e_{O_2}) of T_2. The differential output voltage $(e_{O_1} - e_{O_2})$ is:

$$e_{O_1} - e_{O_2} = \frac{2R_L}{R_S\left(\frac{h_{FE_1}+h_{FE_2}}{h_{FE_1}\,h_{FE_2}}\right) + R_E}$$

$$\left\{(e_2 - e_1) + (\Delta V_{BE} + R_S\Delta I_{CO}) - [\,R_S I_{CO_2}\right.$$

$$\left. + V_{EE} - V_{BE_2} - e_2\,]\left[\frac{R_{E_1} - R_{E_2}}{2R_{EE}}\right]\right\} + R_L\Delta I_{CO}.$$

Common Mode Rejection Depends on V_{BE} Match

Common mode rejection (CMR) may be defined as the ratio of the magnitude of the smaller input signal to the magnitude of the output signal. It may be derived as follows:

If $(\Delta V_{BE} + \Delta R_S I_{CO}) - \frac{V_{EE}}{2R_{EE}}\left(R_{E_1} - R_{E_2} + \frac{R_{S_1}}{h_{FE_1}} - \frac{R_{S_2}}{h_{FE_2}}\right)$

is adjusted to zero, thus satisfying Eq. 4 the differential output is

$$(e_{O_1} - e_{O_2}) = \frac{R_L}{R_{EE}}\left[2(e_2 - e_1)\left(\frac{1 + G_2}{G_1 + G_2}\right)\right.$$

$$\left. + e_2\left(\frac{G_1 - G_2}{G_1 + G_2}\right)\right], \qquad (3)$$

where

$$G_1 = \frac{R_{EE}}{R_{E_1} + \frac{R_{S_1}}{h_{FE_1}}} \text{ and } G_2 = \frac{R_{EE}}{R_{E_2} + \frac{R_{S_2}}{h_{FE_2}}}$$

The first term of Eq. 6 is proportional to the desirable differential gain, and the second term is proportional to the undesirable common-mode gain. The common-mode rejection is the ratio of the differential gain to the common-mode gain, so that

$$CMR = \frac{2\,(1 + G_2)}{G_1 - G_2} = \frac{2\left(R_{EE} + \frac{R_{S_2}}{h_{FE_2}} + R_{E_2}\right)}{\left(\frac{R_{S_2}}{h_{FE_2}} + R_{E_2}\right) - \left(\frac{R_{S_1}}{h_{FE_1}} + R_{E_1}\right)}. \qquad (4)$$

It may safely be assumed that

$$R_{E_2} + \frac{R_{S_2}}{h_{FE_2}} << R_{EE}.$$

If Eq. 4 now is substituted into Eq. 4 (above), the expression for commonmode rejection becomes:

$$CMR \geqslant \frac{V_{EE}}{\Delta V_{BE} + \Delta R_S I_{CO}}$$

TRANSISTOR direct-coupled differentia. amplifiers can be an efficient means of amplifying low-level, low-frequency signals.

Fig. 1 shows the single stage, three-transistor, direct-coupled differential amplifier. In order to insure that $R_{EE} >> R_S$ and R_L, the effective R_{EE} is simulated by the output resistance of T_3 and is connected to the emitters of T_1 and T_2 through the 100-ohm potentiometer. The effective value of R_{EE} is obtained by calculating the output resistance of the stage containing T_3. For the particular transistor used as T_3, the effective value of R_{EE} was found to be approximately 30 K.

T_1 and T_2 were mounted as close together as possible in a transistor clip* that is designed to keep the temperatures of the two TO-18 transistor cases equal. This clip maintains the junctions of the two transistors at approximately the same temperature. The dc drift of the experimental amplifier was measured at the output terminals for a temperature variation from +25 C to +125 C. At the elevated temperature the output drift was 75 mv, which corresponds to an equivalent input drift of 4.5 μv per deg C. Referring to Fig. 4 in the first part of this article (*ED*, Aug. 2, p 72) it is noted that the maximum change in ΔV_{BE} for a random group of 2N861's over this temperature range is approximately 0.6 mv, or 6 μv/deg C. The fact that the measured dc drift with ambient temperature variations was 4.5 μv/deg C indicates the efficiency of the clip in maintaining the transistor junctions at the same temperature.

*Atlee Corp. Part No. 100-300-1-9

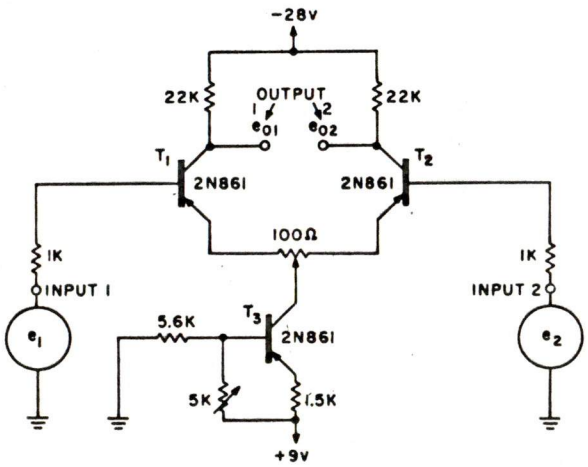


Fig. 1. A single-stage, three transistor, direct-coupled differential amplifier.

Fig. 2 shows the frequency response of the experimental amplifier at +25 C and at +125 C. The low frequency voltage gain of the stage is 45 db at both ambient temperatures. The response at +25 C is 3 db down at 40 kc, decreasing at a rate of 8 db per octave to unity gain at 4 mc. The response at +125 C is 3 db down at 35 kc and is within 2 db of the room temperature response at the higher frequencies.

For the transistors used in the experimental amplifier,

$$h_{FE_1} = 44,$$
$$R_{S_1} = 1\,\text{K} + 700 = 1700,$$
$$h_{FE_2} = 42,$$
$$R_{S_2} = 1\,\text{K} + 460 = 1460.$$

Remembering that R_E includes the dynamic emitter resistances (kT/qI_E),

$$R_E = 100 + 26 + 26 = 152.$$

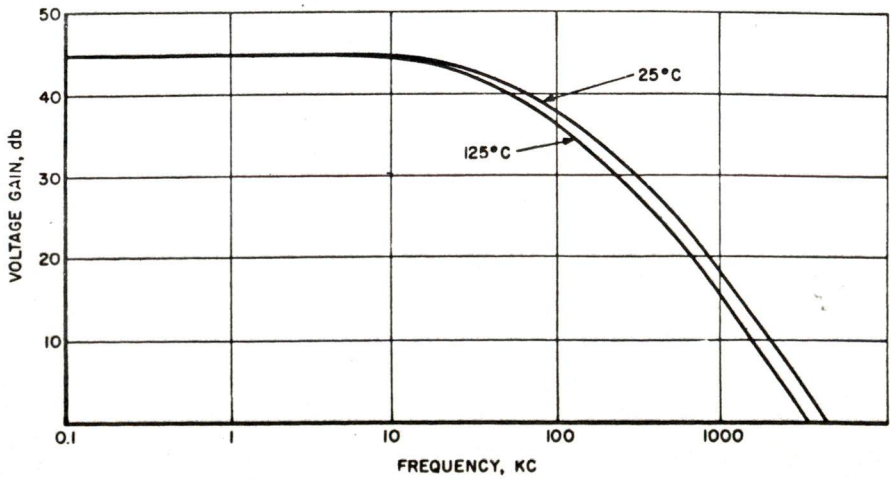


Fig. 2. Response of experimental amplifier at 25 C and 125 C. Low frequency voltage gain is 45 db at both ambients. Response is 3 db down at 40 kc and 35 kc respectively, for the 25-C and 125-C curves.

The differential gain, therefore, is:

$$G_d = \frac{2R_L}{\frac{R_{S_1}}{h_{FE_1}} + \frac{R_{S_2}}{h_{FE_2}} + R_E}$$

$$= \frac{44\text{ K}}{226} = 194 = 45.8\text{ db}$$

This is in agreement with the measured differential gain of 45 db.

Measured CMR Agrees With Calculated Value

The common-mode rejection of the experimental amplifier at 1000 cps was measured as 8000, or 78 db, and linear amplification was obtained up to approximately 20 v peak-to-peak output signal. For the particular transistors used in the amplifier, ΔV_{BE} was measured as 7.5 mv at the operating conditions of

$$V_{CE} = -6\text{ v},$$
$$I_E = +1.0\text{ ma}.$$

Since the current source T_3 supplies the total emitter current of 2.0 ma through the effective value of

$$R_{EE} = 30\text{ K},$$

the effective value of $+V_{EE}$ in the experimental amplifier is (2.0 ma) (30 K) = 60 v.

The calculated *CMR*, therefore, is:

$$CMR = \frac{V_{EE}\ (effective)}{V_{BE}} = \frac{60}{7.5 \times 10^{-3}} = 8000.$$

Again, this is in excellent agreement with the experimental results obtained.

The experimental direct-coupled differential amplifier described here is indicative of the performance that may be obtained by the use of proper design techniques. The performance of the amplifier is in excellent agreement with the design equations developed.

The design equations also are useful in determining the desirable transistor parameters for differential amplifier applications. Eq. (1) shows that the common mode rejection is proportional to the degree to which the base-to-emitter voltages are matched at the desired operating point. Eq. (2) shows that, while a high dc current gain (h_{FE}) is desirable, the differential gain is insensitive to differences between h_{FE_1} and h_{FE_2}. For high differential gain, the source resistance should be low. In addition, Eq. (2) implies an inverse dependence of the differential gain on the V_{BE} match between units. For high gain, R_E should be small. However, the minimum value of the potentiometer connected between the emitters of T_1 and T_2 is equal to $\Delta V_{BE}(max)$ divided by the emitter current of one transistor. ■■

CORRELATION BETWEEN THE BASE-EMITTER VOLTAGE AND ITS TEMPERATURE COEFFICIENT

ALFONS TUSZYNSKI
AMERICAN OPTICAL COMPANY
INSTRUMENT DIVISION
Buffalo • New York

A. INTRODUCTION

Transistors used in the first stage of DC Differential Amplifiers are frequently matched for current transfer ratio in an attempt to reduce the thermal drift of the amplifiers. In low impedance circuits, however, null stability is greatly affected by the temperature dependence of the base-emitter voltage (V_{be}) and is limited by the degree of mismatch in the thermal coefficients of V_{be} of the two transistors used in the differential pairs.

This paper presents experimental evidence of correlation between the thermal coefficients of V_{be} and the actual value of V_{be}, under specified conditions, and attempts to show that significant improvement in the null stability of DC amplifiers can be achieved by pairing transistors for V_{be} at a fixed temperature.

B. THERMAL DRIFT IN TRANSISTORS

Transistor characteristics are notoriously sensitive to temperature variations; the greatest difficulties being caused by changes in:

1. Collector cut-off current I_{co}
2. Current transfer ratio h_{fe} or "β"
3. Base-emitter voltage V_{be}

The effect of the various drift components, which may originate in different parts of an amplifier, is best expressed in terms of "equivalent input drift," defined as the input signal which will counterbalance the disturbing force. The thermal component of the equivalent input drift, representing the sum of errors due to variation in the ambient temperature of an amplifier, is called here the "thermal drift" ($e_{G\theta}$) and is expressed in Volts per °C. It is shown in the appendix that in the case of a single transistor

$$e_{G\theta} \approx (\triangle I_{co} - I_b \triangle\beta/\beta)\, R_G - \triangle V_{be} \qquad (1)$$

where $\triangle X$ denotes a change in parameter X per °C, and R_G represents the source impedance, as seen by the base of the transistor.

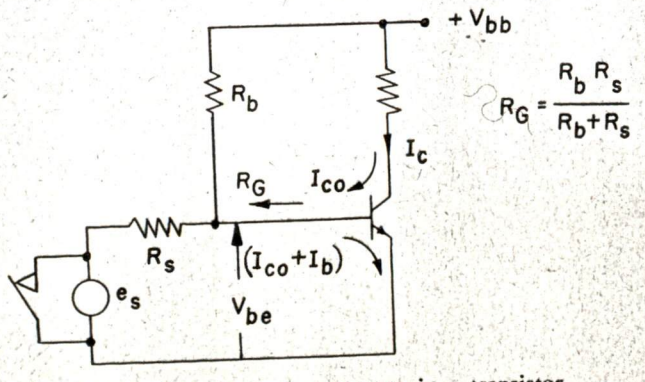


Figure 1 — Operating currents in a transistor.

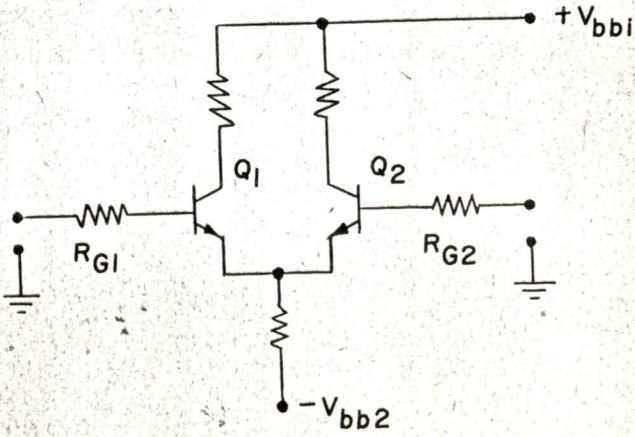


Figure 2 — Common emitter differential amplifier.

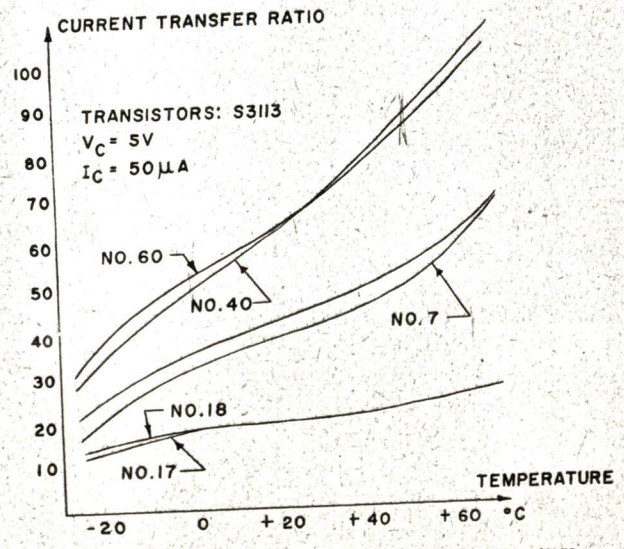
The common emitter differential amplifier, shown in skeleton form in Figure 2, is probably the most popular DC amplifier connection. Errors common to transistors Q_1 and Q_2 cancel out, so that the equivalent thermal drift of this circuit is, by repeated use of Equation 1:

$$e_{G\theta} \approx (\triangle I_{co1} - I_{b1}\,\triangle\beta_1/\beta_1)R_{G1} - \triangle V_{be1} - (\triangle I_{co2} - I_{b2}\triangle\beta_2/\beta_2)R_{G2} - \triangle V_{be2} \qquad (2)$$

The collector cut-off current I_{co} of good planar silicon transistors is, at 5 volts and 25°C, of the order of a Nano Ampere, with a temperature coefficient of about 10 per cent per °C; the actual value of this low current varies considerably from transistor to transistor, making $\triangle I_{co1} - \triangle I_{co2}$ roughly equal to the larger of the two increments.

To reduce the effects of variations in the current transfer ratio (β), one selects transistors of good "betas" and operates them at relatively low collector currents (I_b in Equations 1 and 2 is approximately equal to the collector current divided by the transfer ratio). There are many types of silicon transistors which have current transfer ratios of 20 or more at collector currents of 50 micro Amperes. The thermal coefficient of beta depends on the temperature range, the collector current, beta itself, and the type of transistor, but seldom exceeds 1 per cent per °C. The data shown in Graph 1, suggests that transistors of the same type and comparable betas have similar thermal coefficients of beta. The method of estimation of the probable value of the drift component due to changes in current transfer ratio, depends on the absolute and relative values of R_{G1} and R_{G2}, as well as the mismatch of the base currents. In the extreme case of source impedance mismatch ($R_{G2}=0$), it works out to be $10R_{G1}$ Nano Volts per °C, at a collector current of 50 micro Amperes and a transfer ratio of 25, with a temperature coefficient of 1/2% per °C.

The temperature coefficient of the base-emitter voltage is about 2 mill Volts per °C. The tracking of the temperature coefficients of V_{be} of two transistors of the same type is extremely good (better than 1 part in 50), reducing the probable value of $\triangle V_{be1} - \triangle V_{be2}$ to some 20 micro Volts per °C.



Graph 1 — Current transfer ratio vs. temperature.

In circuits in which the source impedances are low and differ by less than 2 kilo Ohms, the thermal coefficient of the base-emitter voltage predominates, especially in the temperature range 0—60°C. The zero stability of the network could be improved by the introduction of components adjustable on test at two or more temperatures, but this leads to complications in production and servicing.

C. TEMPERATURE COEFFICIENT OF V_{be}

The average temperature coefficients of a batch of one-hundred transistors, type number 2N1613, are plotted in Graph 2, against the room temperature values of their base-emitter potentials. The same information for a different type of transistor (2N2318) is presented in Graph 3. There is a distinct correlation between the value of the base emitter voltage of a transistor, and the magnitude of the thermal coefficient of that voltage. The thermal coefficients shown in the two graphs represent average values over a temperature range of 36°C (24 to 60°C). The tolerance on the ambient temperature at which the base-emitter was measured, was ± 1°C for the batch shown in Graph 2, ± 1/2°C for Graph 3.

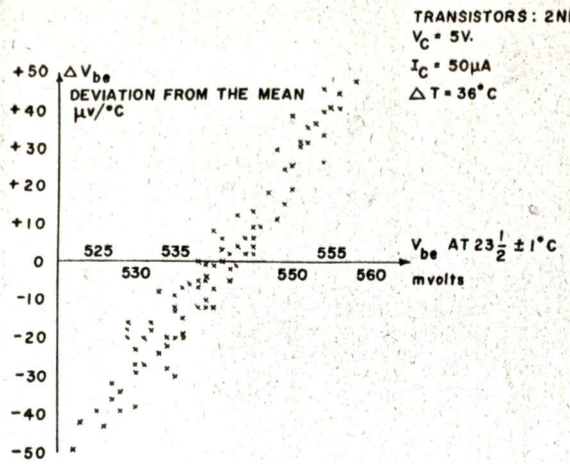
The significant figures for Graph 2 are brought out in Table 1. The standard deviation of $\triangle V_{be}$ is 23μV/°C for the whole batch, with 96μV/°C as the maximum difference in the thermal coefficients of any two transistors in the batch. Grouping of the transistors into classes according to V_{be}, each class covering a spread of 5mV in V_{be}, decreases the standard deviation of the (classified) batch to 6.1 and the maximum deviation to 31μV/°C.

An additional test was performed to check the validity of the argument for selection, and a standard deviation of 2.6μV/°C was obtained for a set of 10 transistors differing in V_{be} by less than 2mV at an ambient temperature of 24 ± 1/2°C.

TABLE I

Classification of Data in Graph 2.

Group #	# of Transistors in group	V_{be} Spread	Deviation of $\triangle V_{be}$ from batch mean	Standard deviation μV/°C	$\triangle V_{be}$ max. minus $\triangle V_{be}$ min.
1	3	522-525	−43	3.8	10
2	12	526-530	−30	8.1	27
3	14	531-535	−19	6.1	22
4	27	536-540	−7	5.8	28
5	20	541-545	+4	4.7	18
6	10	546-550	+21	8	31
7	13	551-555	+34	5	19
8	3	556-558	+44	5	7
TOTAL	102	522-558	0	23	96
Classified Batch	102			6.1	31



Graph 2 — Dispersion of the thermal coefficient of the base-emitter voltage.

D. CONCLUSION

By means of two simple tests carried out entirely at room temperature, transistors can be paired for zero stability of the order of a few micro Volts/°C (source impedance below 1 k Ohm). Matching for 5 to 10% in current transfer ratio and 2 to 5 milli Volts in base-emitter voltage, presents no difficulties and no transistors need be rejected if batches of at least 200 transistors of specified minimum current transfer ratio are processed at a time. Care must be taken in packaging to prevent any temperature differentials between the two transistors. A dual unit (two transistors in one case) of appropriate characteristics would be ideal, though expensive at the present time.

The above discussion applies mainly to the temperature range 0—60°C. At temperatures below 0°C, drift due to current transfer ratio increases, whereas at high temperatures, the collector cut-off current becomes troublesome.

References

1. Middlebrook and Taylor, "Differential Amplifier With Regulator Achieves High Stability, Low Drift", *Electronics*, July, 1961, Page 65.

2. D. W. Slaughter, "The Emitter-Coupled Differential Amplifier", *IRE Transactions — Circuit Theory — 3*, March, 1956, Page 51.

APPENDIX: THERMAL DRIFT IN TRANSISTORS

The collector, emitter and base currents of a transistor are related as follows:

$$I_e = I_c + I_b \qquad (1)$$

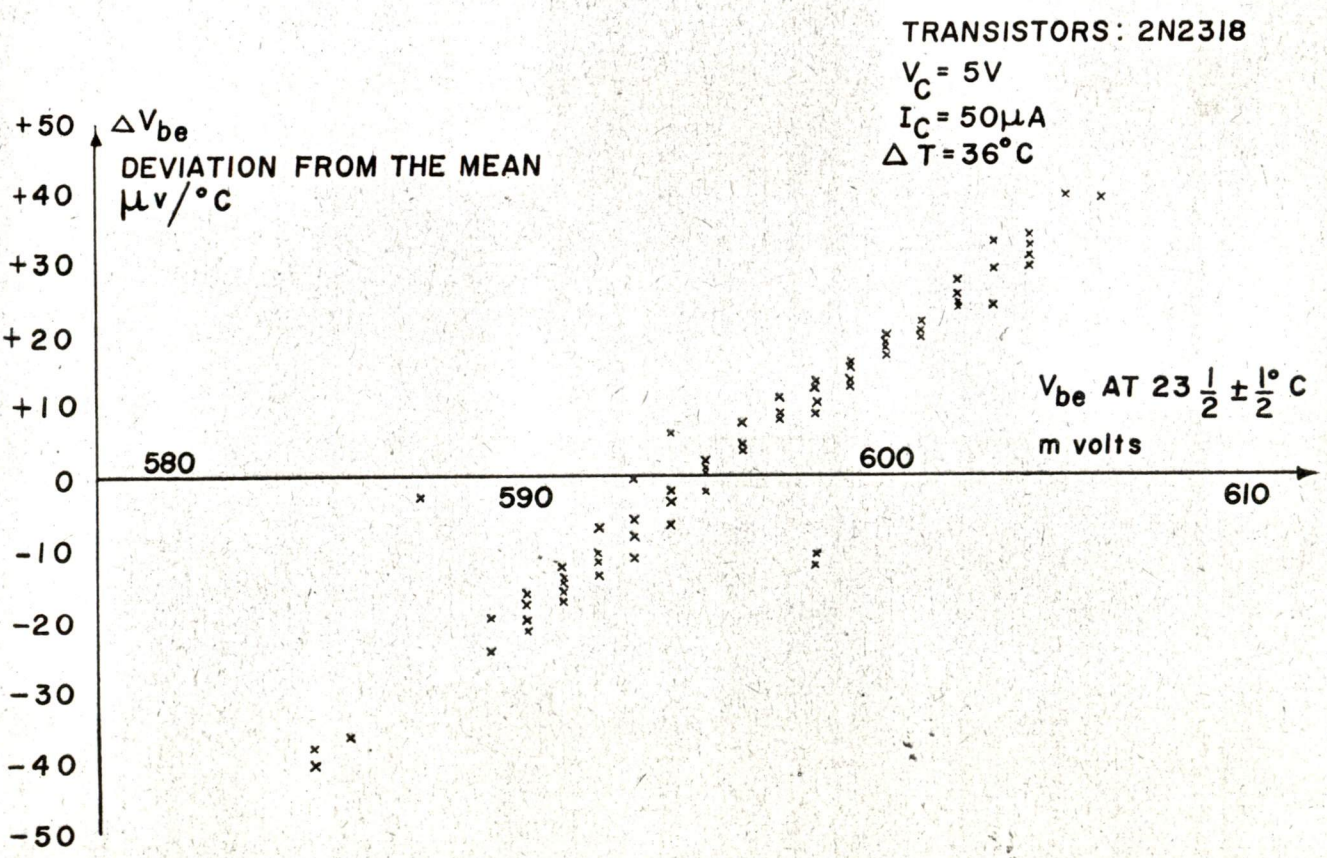
$$I_c = \alpha I_e + I_{co} \qquad (2)$$

whence,

$$I_c = \beta I_b + (\beta + 1) I_{co} \qquad (3)$$

where,

$$\beta = (1 - \alpha)/\alpha \qquad (4)$$



Graph 3 — Dispersion of the thermal coefficient of the base-emitter voltage.

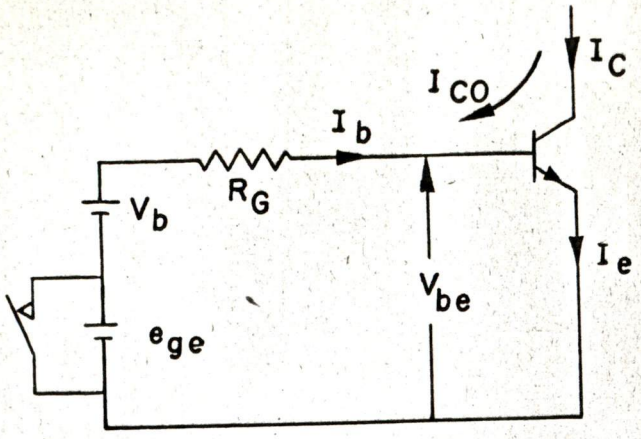


Figure 3 — Equivalent bias circuit.

The base current can be evaluated from the mesh equation,

$$I_b R_g - V_b + V_{be} = 0 \qquad (5)$$

and Equation 3 can be written as:

$$I_c = \beta(V_b - V_{be})/R_g + (\beta + 1)I_{co} \qquad (6)$$

Differentiation of Equation 6 with respect to temperature, followed by resubstitution of Equation 5, yields:

$$\triangle I_c = (I_b + I_{co})\triangle\beta - \beta\triangle V_{be}/R_g + (\beta + 1)\triangle I_{co} \qquad (7)$$

where:

$\triangle I_c$ = increment of collector current caused by 1°C change in temperature

$\triangle V_{be}$ = change in V_{be}/°C, etc.

To offset $\triangle I_c$, the base current must be changed by approximately $\triangle I_c/\beta$; this requires a voltage $e_{g\theta}$ of magnitude $\triangle I_c/\beta R_g$. Rewriting $\triangle I_c$ in terms of $e_{g\theta}$ we get:

$$\triangle I_c \approx \beta e_{g\theta}\ /R_g \qquad (8)$$

which, inserted in Equation 7, gives the equivalent thermal drift per °C as:

$$e_{g\theta} = R_g(I_b + I_{co})\triangle\beta/\beta - \triangle V_{be} + R_g\triangle I_{co}(\beta + 1)/\beta \qquad (9)$$

$$\approx R_g(\triangle I_{co} + I_b\triangle\beta/\beta) - \triangle V_{be} \qquad (10)$$

The increment $\triangle V_{be}$, in the above equation, is negative, while $\triangle\beta$ and $\triangle I_{co}$ are positive, so that the numerical value of $e_{g\theta}$ is:

$$e_{g\theta} \approx R_g(\triangle I_{co} + I_b\triangle\beta/\beta) + |\triangle V_{be}| \qquad (11)$$

CORRECTION

The following minor typographical errors appeared in the paper, "Thermoelectric Generators as Power Sources for Thermoelectric Refrigerators", by Barry Stern, *solid/state/design*, May 1962:

1. The expression following equation (10),

$$\frac{\alpha g\ (Th - Tah)}{Rext + Rc},$$

and appearing in the left hand side of equation (11), should be

$$\frac{\alpha g\ (Th - Tah)}{Rext + Rg}.$$

2. In Figure 3, the designation P and N should be reversed on the cooling couple, so that Figure 3 corresponds to Figure 1.
3. In the Table of Contents and in The Picture in Brief, "Joffe" should be "Ioffe".

crossover of the curves in Fig. 9(C). A close examination of the data under the heading "1/4-Inch Gap" in Table III suggests that the data for the 1/4-inch gap during radiation may be suspect.

Radiation accounted for at least a 3-decade lowering in gap d-c insulation resistance. Changes in d-c insulation resistance, under radiation conditions, as a result of varying temperatures, were less than half a decade.

Knowledge in the field of radiation effects is still sparse. The amount of general information on components under radiation is particularly limited. It is hoped that the experimental data obtained in these tests will serve to shed some light upon the theories advanced, and will be helpful in evaluating and conducting subsequent tests on all types of electronic components. These tests will serve as the basis for a designer to establish safe voltage breakdown distances between terminals, and between terminals and ground, for transformer applications under extreme altitude and radiation conditions.

Reference

1. Terminal Spacings for High-Altitude and Ultrahigh-Temperature Electronic Transformer Applications, G. I. Duncan, W. A. Rectanus. *AIEE Transactions*, pt. II (*Applications and Industry*), vol. 78, Mar. 1959, pp. 16–19.

Stable Transistor Wide-Band D-C Amplifiers

ROBERT H. OKADA
NONMEMBER AIEE

STABLE D-C amplification is required in many engineering applications. In the past, vacuum-tube circuits have been developed which meet most low drift requirements by utilizing differential amplification, chopper, and chopper stabilization techniques. Transistors offer many inherent advantages such as small size, long life, no filament power, low power dissipation and good reliability, but they in turn present new problems in design and techniques. Since the major source of drift in tube d-c amplifiers is filament power variations, the transistor might seem to have inherent advantages concerning drift. However, the variations of transistor parameters with temperature presents as formidable a problem as vacuum-tube filaments. Stable d-c amplifiers using transistors can be designed and constructed by modifying the basic techniques used in vacuum-tube circuits.

All amplifiers have specifications concerning the following: gain, gain stability, gain and phase response versus frequency, drift, dynamic range, input and output impedance, and stability against self-oscillation if feedback is used. Transistors differ from tubes notably in the impedance levels and d-c drift. Although most transistor circuit configurations present a low input impedance, it is not very difficult to build feedback amplifiers which have input impedances of 100 kilohms to 1 megohm. D-c drift in transistor amplifiers is due mainly to temperature variations of I_{co} and V_{be}. Silicon transistors can be used to almost eliminate I_{co} variations, since room temperature value of I_{co} is so small that that the temperature effects are several orders of magnitude below V_{be} effects. Thus, the change in V_{be} with temperature (approximately 2.5 millivolts per degree centigrade) is the major problem for stable d-c transistor amplification.

Methods of Reducing Drift

The main methods of reducing drift are: nonlinear elements, differential amplifiers, chopper amplifiers, chopper stabilization, and combinations of the foregoing. Nonlinear elements such as thermistors and diodes can be obtained which will cancel out drift due to amplifier components, but this method usually involves careful matching and selection, which does not lend itself to production. Also, variations with age are difficult to predict, which adds to the delicateness of this technique. If another transistor is used as the nonlinear element, the result is usually a differential amplifier. If carefully matched transistors are used, and age variations are similar, the differential amplifier can have very low drift. Slaughter presented a differential amplifier which had excellent drift characteristics (due to well-matched transistors).[1] Fig. 1 is a schematic of this amplifier.

Chopper amplifiers, which modulate the input signal at a fixed frequency, amplify the resulting carrier and side bands, and then demodulate, are essentially drift-free, provided the chopper itself is a stable element. Mechanical choppers can be used which introduce negligible drift but are limited in frequency to about 1 kc. The over-all bandwidth of such an amplifier is a few hundred cycles. Transistors can be used as the chopping element which utilize chopping frequencies as high as 100 kc with resulting over-all bandwidth in the range of 10 kc. Frequency limitations of the transistor chopping circuits prevent greater bandwidths, and the drift of the transistor choppers must be considered. The latter has been analyzed by Chaplin,[2] and the results look promising.

To obtain wide-band stable d-c amplification the chopper stabilization technique of Goldberg[3] can be modified for transistor circuits. Blecher describes such an amplifier which has a drift referred to the input of ±5 millivolts over a 50-degree-centigrade temperature range using a single-ended germanium input stage.[4] Goldberg's technique consists of providing a separate path for the d-c component, which contains a high-gain

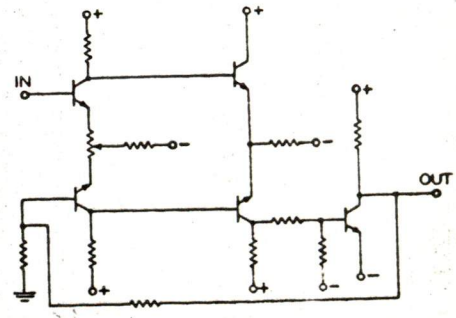


Fig. 1. Schematic of Slaughter amplifier

Paper 60-6, recommended by the AIEE Electronics Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Winter General Meeting, New York, N. Y., January 31-February 5, 1960. Manuscript submitted April 24, 1959; made available for printing November 23, 1959.

Robert H. Okada is with the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, Pa., and the Burroughs Corporation, Paoli, Pa.

The author wishes to acknowledge the contribution of Harold Noffz, who did all the experimental work, and of Harvey Rosenberg, who was in charge of the project under which the work was done. The co-operation of the Burroughs Corporation is also appreciated.

low-drift chopper amplifier preceding the main amplifier.

By a combination of silicon transistor differential amplifier and chopper stabilization, a transistor amplifier can be built with the use of transistors which have normal production spreads of V_{be} temperature variations, and resulting in a drift referred to input of less than one half of a millivolt. By selecting matched V_{be} temperature characteristics the drift can be reduced much farther.

Transistor Operational Amplifiers

The standard voltage operational amplifier is shown in Fig. 2(A) with A_v a voltage gain (negative). The gain of the amplifier is Z_f/Z_i if $|A_v|>>1$ and $Z_i>>(Z_f/A_v)$. For a transistor amplifier the last requirement is usually not true. However, consider the current amplifier in Fig. 2(B). From the Appendix is obtained

$$\frac{I_L}{I_i} \doteq -\frac{Z_f}{Z_L} \text{ if } Z_f \gg \frac{Z_i}{A_i} \text{ and } Z_L \gg \frac{Z_f+Z_i}{1+A_i}$$

and:

$$\frac{E_L}{E_i} = \frac{-Z_f}{Z_i} \text{ if } A_i\beta_i \gg 1,\ A_i\beta_i \gg \frac{2Z_i}{Z_i}, \text{ and}$$

$$\beta_i \equiv \frac{I_f}{I_L} \doteq \frac{Z_L}{Z_f}$$

Thus, for the voltage gain to be the ratio of two impedances, the input impedance at the A_i amplifier terminals should ideally be zero, and a low input impedance transistor amplifier is ideal for an operational amplifier. These equations are a special case of Blecher's results which utilize the short-circuit current gain and source impedance in place of a current amplifier.[4] However, if it is remembered that the value of A_i is a function Z_L (and Z_f in general), these equations are useful because of their simplicity, especially since the above conditions are usually valid in practice.

Possible Chopper Stabilization Configurations

Goldberg's circuit is shown in block diagram form in Fig. 3(A). This circuit does not correct for drifts due to grid current leakage at A since the circuit interprets potential at this point as signal. Since the transistor drift components (I_{co} and V_{be}) appear at this point, the circuit must be modified as shown in Fig. 3(B). If the ratio of the primed impedances is the same as the unprimed ones, point B will be zero potential if the output is equal to the gain (E_0/E_i) of the amplifier times the input. Any discrepancy is due to drift components and the chopper amplifier will attempt to correct for the drift. Incidentally, aside from drift signals appearing at point A, point B is at the same potential as point A, and therefore the primed impedances can be removed and the two points A and B tied together. This reduces Fig. 3(B) to Fig. 3(A) and is an excellent means of explaining to the uninitiated the (at first) puzzling configuration of Fig. 3(A).

The chopper-stabilizing effects are limited (as will be shown later), which makes a differential input stage desirable in order to reduce the amount of stabilization required. In a typical example the V_{be} drift over a range of 70 F (degrees Fahrenheit) to 150 F will be 100 millivolts, which is to be compared with a ΔV_{be} in a differential input stage of 17 millivolts. The 17-millivolt figure is the maximum ΔV_{be} variation of 32 silicon transistors, type *2N479*, over the same temperature range. Thus if these figures are typical of production results, the differential input amplifier has an inherent drift reduction of about 5 to 1 without selection. The differential stage can also supply the high common mode rejection necessary if a single-ended output is required. If silicon transistors are used, the I_{co} effects are negligible compared with V_{be} effects, but the same order of inherent drift reduction applies to I_{co} drift in germanium transistor differential stages.

Analysis of Chopper Stabilization

Fig. 3(C) represents the same circuit as 3(B), in which A_i and A_i' represent actual current gains of amplifiers $G1$ and $G2$, and I_k is the current required to reduce the output to zero because of drift.

Writing Kirchoff's equations yields

$$I_0+I_0' = I_{in} \tag{1}$$

$$-(I_1+I_0+I_f+I_k)A_i = I_f+I_f'+I_L \tag{2}$$

$$-(I_0'+I_f')A_i' = I_1 \tag{3}$$

and if (see the Appendix)

$$Z_L \gg \frac{Z_{in}}{A_i},\ Z_f \gg \frac{Z_{in}}{A_i},\ Z_L \gg \frac{Z_{in}'}{A_i'}$$

and

$$Z_f' \gg \frac{Z_{in}'}{A_i'}$$

$$Z_L I_L = Z_f I_f = Z_f' I_f'$$

and if (see the Appendix)

$$A_i \gg 1,\ A_i' \gg 1,\ Z_i \gg \frac{Z_{in}}{A_i\beta_i} \tag{4}$$

and

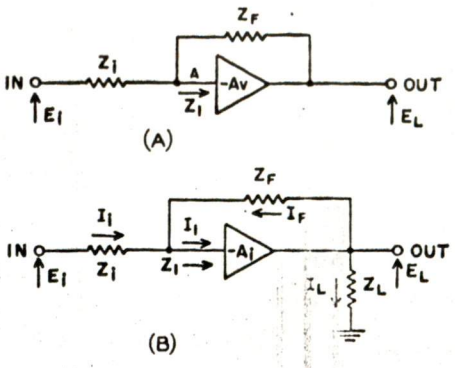


Fig. 2. Operational amplifiers. A—Voltage mode. B—Current mode

$$Z_f' \gg \frac{Z_{in}'}{A_i'\beta_i'}$$

then

$$Z_i I_0 = Z_i' I_0' \tag{5}$$

Solving equations 1 and 5 simultaneously yields the load current in terms of the input current

$$I_L = -\frac{Z_f'}{Z_L}\frac{Z_i}{Z_i+Z_i'}\left[I_{in}+\frac{Z_i+Z_i'}{Z_i}\frac{I_k}{A_i'}\right] \tag{6}$$

if, as is usually the case,

$$A_i \gg 1,\ A_i' \gg \frac{1}{A_i},\ \frac{A_i'}{Z_f'} \gg \frac{1}{Z_f}+\frac{1}{AZ_L},\ \frac{A_i'Z_i}{Z_i'} \gg 1$$

If equation 6 is compared with the analogous result using no chopper stabilization (see the Appendix), i.e.,

$$I_L = -\frac{Z_f}{Z_L}(I_{in}+I_k)$$

if

$$Z_f \gg \frac{Z_{in}}{A_i}$$

and

$$Z_L \gg \frac{Z_f+Z_1}{1+A_i} \tag{7}$$

it is seen that the drift current is reduced by a factor of

$$\frac{Z_i+Z_i'}{Z_i}\frac{1}{A_i'}$$

In terms of voltage, since in Fig. 3(C)

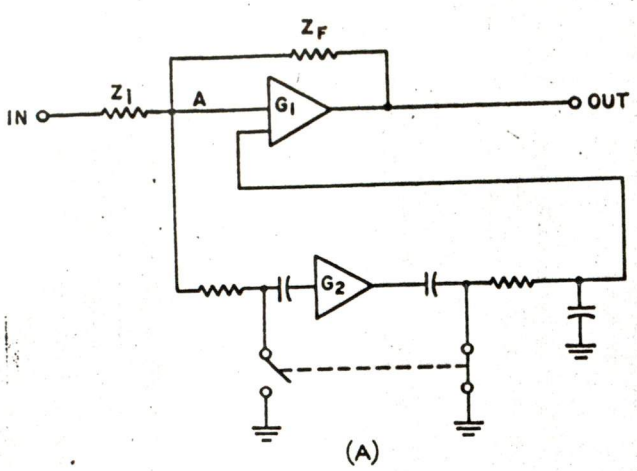
$$E_L = I_L R_L,\ E_{in} \doteq I_{in}\frac{Z_iZ_i'}{Z_i+Z_i'}$$

the output voltage is

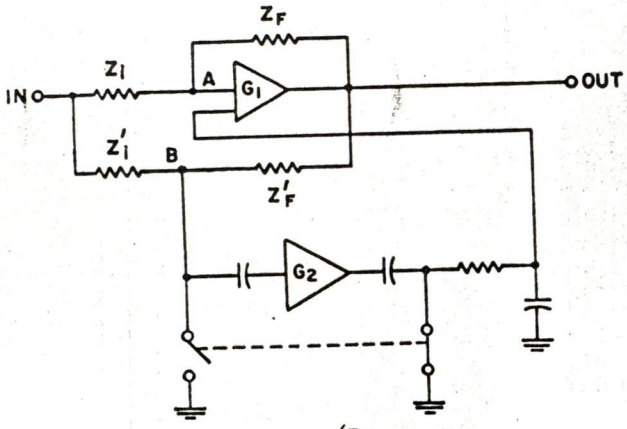
$$E_L = -\frac{Z_f'}{Z_i'}\left[E_{in}-\frac{Z_i'I_k}{A_i'}\right] \tag{8}$$

The analogous result for no chopper stabilization is

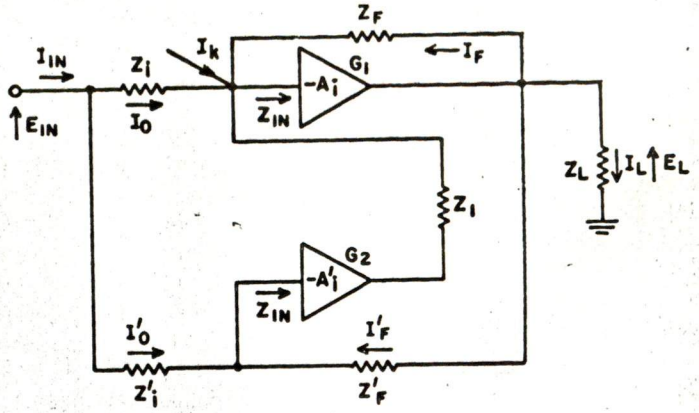
$$E_L = \frac{-Z_f}{Z_i}[E_{in}-Z_iI_k] \tag{9}$$



(A)



(B)



(C)

Fig. 3. A—Schematic of Goldberg's chopper stabilization scheme. B—Modification for transistor circuits. C—Schematic of B used for determining drift

and the ratio of the drift terms is

$$\frac{\frac{Z_1' I_k}{A'}}{\frac{Z_1 I_k}{Z_1 A_1}} = \frac{Z_1'}{Z_1 A_1}$$

or voltage drift reduction is

$$\frac{Z_1'}{Z_1}\,\frac{1}{A_1'} \qquad (10)$$

The choice of the ratio Z_1'/Z_1 is determined by the required input impedance which is equal to the parallel combination of Z_1 and Z_1'. The chopper current gain A_1' can be very high and in mechanical choppers is limited only by the electrostatic and electromagnetic pickup at the vibrating contracts, and the minimum value of Z_1 in Fig. 3(C).

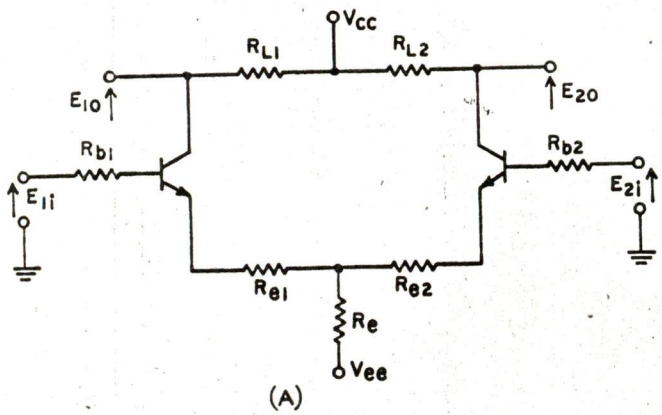
Analysis of Differential Amplifiers

A differential amplifier is shown in Fig. 4(A). To study the effects of I_{co}, V_{be}, and dissimilar transistors, the equivalent circuit of Fig. 4(B) will be used. (This circuit treats the base-to-emitter voltage as an externally available voltage which can be measured experimentally. Both d-c and varying signals are included. The basic assumptions for this equivalent circuit are: 1. The constant α_n. 2. The collector impedance is very high compared with load impedance. 3. Any internal base resistance is included in R_{b1} and R_{b2}.) The branch equations are

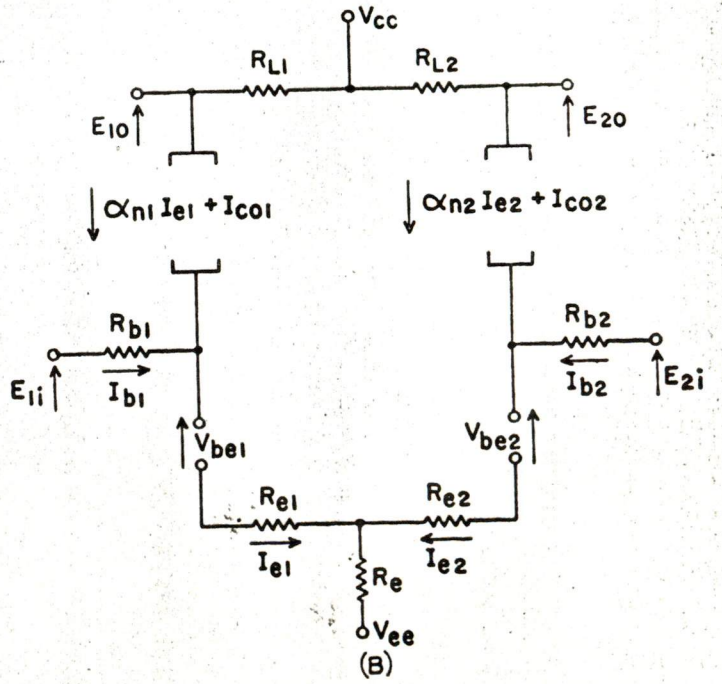
$$E_1 - V_{be1} - V_{ee} = R_{b1} I_{b1} + (R_{e1} + R_e) I_{e1} + R_e I_{e2} \qquad (11)$$

$$E_2 - V_{be2} - V_{ee} = R_{b2} I_{b2} + R_e I_{e1} + (R_{e2} + R_e) I_{e2} \qquad (12)$$

$$I_{e1} = I_{b1} + \alpha_{n1} I_{e1} + I_{co1} \qquad (13)$$



(A)



(B)

Fig. 4. A—Transistor differential amplifier stage. B—An equivalent circuit for A

$$I_{e2}=I_{b2}+\alpha_{n2}I_{e2}+I_{co2} \tag{14}$$

Using the relation, $\beta_n=\alpha_n/1-\alpha_n$, and solving equation 13 for I_{b1} yields

$$I_{b1}=I_{e1}\left(\frac{1}{1+\beta_{n1}}\right)-I_{co1} \tag{15}$$

Similarly,

$$1_{b2}=I_{e2}\left(\frac{1}{1+\beta_{n2}}\right)-I_{co2} \tag{16}$$

Equations 11, 12, 15, and 16 can be solved for I_{e2}, with the use of the knowledge that

$$E_{20}=V_{cc}-(\alpha_{n2}I_{e2}+I_{co2})R_{L2} \tag{17}$$

The result is

$$E_{20}=V_{cc}-\alpha_{n2}R_{L2}\times\left[\frac{(E_{2i}-E_{1i})+(V_{be1}-V_{be2})+R_{b2}I_{co2}-R_{b1}I_{co1}-\left(\frac{R_{b1}}{\beta_{n1}}+R_{e1}\right)\frac{V_{ee}}{R_e}}{R_{e1}+R_{e2}+\frac{R_{b2}}{\beta_{n2}}+\frac{R_{b1}}{\beta_{n1}}+\frac{R_{b1}R_{b2}}{\beta_{n1}\beta_{n2}R_e}}\right]-I_{co2}R_{L2} \tag{18}$$

provided $R_{b1}/\beta_{n1}R_e \ll 1$. A similar result is obtained for E_{1o} and, taking the difference,

$$E_{20}-E_{1o}=\frac{(E_{1i}-E_{2i})(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\frac{(V_{be2}-V_{be1})(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\frac{(R_{b1}I_{co1}-R_{b2}I_{co2})}{\Delta}\times\frac{(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{1}+\left(\frac{R_{b1}}{\beta_{n1}}+R_{e1}-\frac{R_{b2}}{\beta_{n2}}-R_{e2}\right)\times\frac{\frac{V_{ee}}{R_e}(\alpha_{n1}R_{L1}+\alpha_{n2}R_{L2})}{\Delta}+\frac{(I_{co1}R_{L1}-I_{co2}R_{L2})}{\Delta} \tag{19}$$

where

$$\Delta=R_{e1}+R_{e2}+\frac{R_{b2}}{\beta_{n2}}+\frac{R_{b1}}{\beta_{n1}}+\frac{R_{b1}R_{b2}}{\beta_{n1}\beta_{n2}R_e} \tag{20}$$

The first term in equation 19 is the desired output for a given input and it represents the voltage gain of the differential amplifier. If the transistors and external resistances are equal, the gain simplifies to

$$\frac{E_{2o}-E_{1o}}{E_{1i}-E_{2i}}=\frac{2\alpha_{n12}R_{L12}}{R_{e12}+\frac{R_{b12}}{\beta_{n12}}+\frac{R_{b}{}^{2}{}_{12}}{2\beta_{n}{}^{2}{}_{12}R_e}} \tag{21}$$

The second term of equation 19 is the output due to differences in V_{be}. For this to be zero at all temperatures, V_{be1} and V_{be2} must be equal and must drift equally. The third and fifth terms are due to differences in I_{co1} and I_{co2}, and the fourth term is an output due to differences in β_{n1} and β_{n2}.

If R_{e1} and R_{e2} are made into a potentiometer with the movable tap connected to the common emitter resistor, the potentiometer can be adjusted to give zero output for zero input at a particular temperature. That is, the potentiometer can be of such value that it can adjust the fourth term to cancel out the sum of the second, third, and fifth terms. However, if the variation of V_{be} and I_{co} of both transistors does not vary exactly the same with temperature, the difference will be revealed as drift in the output. Since V_{be} and I_{co} variations with temperature are intrinsically a function of the semiconductor, the variations in drift between transistors is considerably less than the drift itself.

To determine the common mode rejection of the differential amplifier, the common mode gain is calculated as follows. Taking only the terms of equation 18 containing input signals, but without making the assumption of $R_{b1}/\beta_{n1}R_e \ll 1$ (since later algebraic addition makes the term important for common mode signal results in

$$E_{2o}=\frac{1}{\Delta}\left[-\alpha_{n2}R_{L2}\left(\frac{R_{b1}}{\beta_{n1}R_e}E_{2i}+E_{2i}-E_{1i}\right)\right.$$

Similarly for E_{1o}, and subtracting, yields

$$E_{2o}-E_{1o}=\frac{1}{\Delta}\left[-\alpha_{n2}R_{L2}\left(\frac{R_{b1}}{\beta_{n1}R_e}E_{2i}+E_{2i}-E_{1i}\right)+\alpha_{n1}R_{L1}\left(\frac{R_{b2}}{\beta_{n2}R_e}E_{1i}+E_{1i}-E_{2i}\right)\right] \tag{23}$$

For a common mode signal $E_{2i}=E_{1i}=E_{cm}$, and the gain is

$$G_{cm}=\frac{E_{2o}-E_{1o}}{E_{cm}}=\frac{1}{\Delta}\times\left(-\alpha_{n2}R_{L2}\frac{R_{b1}}{\beta_{n1}R_e}+\alpha_{n1}R_{L1}\frac{R_{b2}}{\beta_{n2}R_e}\right) \tag{24}$$

The major effect is due to difference in β; therefore it is assumed that

$$R_{L1}=R_{L2}=R_{L12},\ R_{b1}=R_{b2}=R_{b12},\ \alpha_{n2}=\alpha_{n1}=\alpha_{n12}$$

which simplifies equation 24 to

$$G_{cm}=\frac{1}{\Delta}\left[\frac{\alpha_{n12}R_{L12}R_{b12}}{R_e}\left(\frac{1}{\beta_{n2}}-\frac{1}{\beta_{n1}}\right)\right] \tag{25}$$

Under the same assumption the differential mode gain in given by equation 21, and the ratio is

$$\text{CMR}=\frac{G_{dm}}{G_{cm}}=\frac{2R_e\beta_{n1}\beta_{n2}}{R_{b12}(\beta_{n1}-\beta_{n2})} \tag{26}$$

where CMR = common mode rejection.

As might be expected from similar tube amplifiers, the common mode rejection varies directly with the common emitter resistor R_e, and inversely with the differences in β. If identical β transistors were used, the previous assumptions would have to be re-evaluated.

It should be emphasized that the equivalent circuit of Fig. 4(B) assumes an internal collector resistance which is high compared with the collector load resistances; when this is not true, the resulting analysis becomes even lengthier than that shown here.

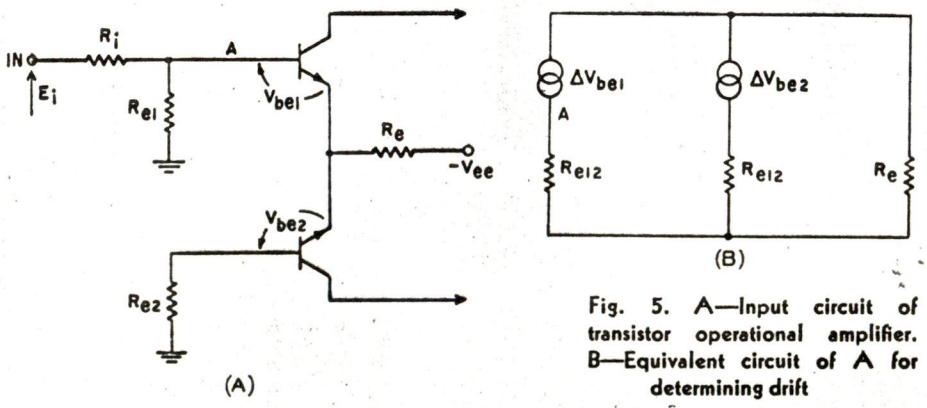


Fig. 5. A—Input circuit of transistor operational amplifier. B—Equivalent circuit of A for determining drift

Analysis of Drift in Differential Amplifiers

Fig. 5(A) illustrates a typical operational amplifier with a differential input. R_{e2} is small in order to keep the impedance at point A small, and R_{e1} is made equal to R_{e2} in order to keep the circuit balanced for drift effects. R_{e1} and R_{e2} are small compared with R_f as is required for operational amplifiers. This results in some of the amplifier input current being shunted into R_{e1} where it produces no

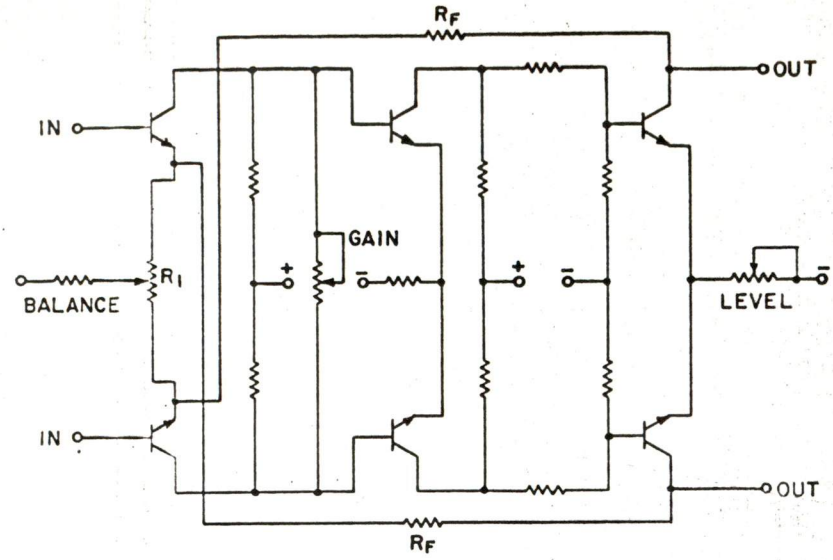


Fig. 6. Differential transistor amplifier with feedback to first-stage emitters

open-loop current gain; however, the gain can be made up in later stages.

The equivalent circuit for drift due to ΔV_{be} is shown in Fig. 5(B), where $R_{e1}=R_{e2}=R_{e12}$, and the internal emitter base impedance is considered small in comparison with R_{e12}. R_f can be neglected or the parallel combination of R_f, and R_{e1} made equal to R_{e12}.

An equal change in V_{be1} and V_{be2} causes current to flow in R_e and the result is attenuated by the common mode rejection of the differential amplifier. By Ohm's law the equivalent drift current at point A due to a difference in the change of V_{be1} and V_{be2} is

$$I_d=\frac{\Delta V_{be1}-\Delta V_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}} \qquad (27)$$

To calculate the output voltage due to drift current I_d, the equation for output voltage per input current of an operational amplifier is from equation 48:

$$\frac{E_L}{I_f}=-R_f$$

if

$$Z_f\gg\frac{Z_1}{A_f}$$

and

$$Z_L\gg\frac{Z_f+Z_1}{1+A_f} \qquad (28)$$

and

$$E_{L1}=-I_dR_f=-\frac{\Delta V_{be1}-\Delta V_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}}R_f \qquad (29)$$

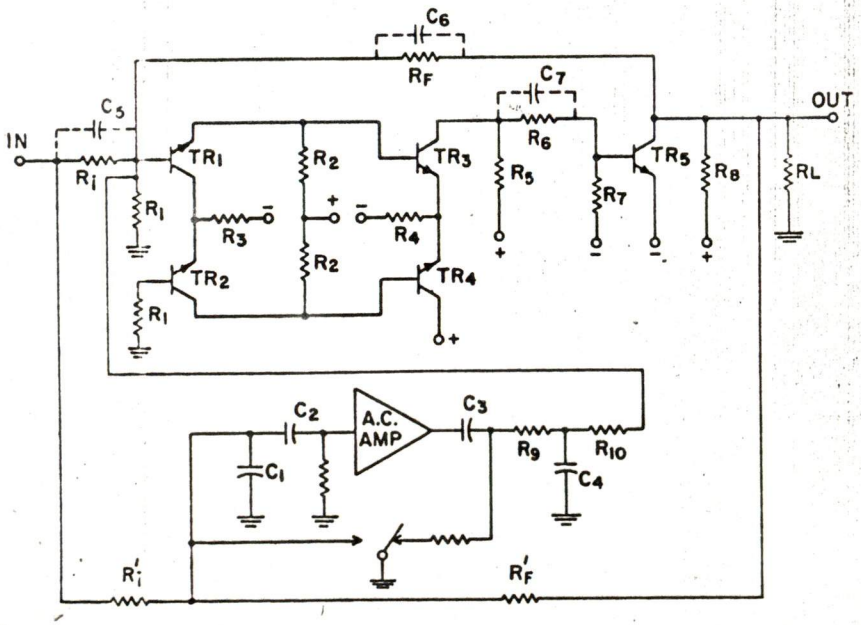


Fig. 7. Chopper stabilized operational transistor amplifier with differential input stages

For unequal I_{co} changes with temperature, the equation is

$$E_{L2}=-\left(\frac{\Delta I_{co1}}{\alpha_1}-\frac{\Delta I_{co2}}{\alpha_2}\right)R_f \qquad (30)$$

where E_{L1} and E_{L2} are the output potentials due to V_{be} and I_{co} drifts respectively. These output drifts can be referred to the input by dividing by the closed loop voltage gain given by $-R_f/-R_f$, and they become

$$E_{d1}=\frac{\Delta V_{be1}-\Delta_{be2}}{R_{e12}+\dfrac{R_{e12}R_e}{R_{e12}+R_e}}R_f \qquad (31)$$

and

$$E_{d2}=\left(\frac{\Delta I_{co1}}{\alpha_1}-\frac{\Delta I_{co2}}{\alpha_2}\right)R_f \qquad (32)$$

where E_{d1} and E_{d2} are the voltage drifts due to V_{be} and I_{co} effects respectively, both being referred to the input.

In a typical example, $\Delta V_{be1}-\Delta V_{be2}=17$ maximum millivolts (for a sample of 29 type *2N479*) over a temperature range of 70 to 150 F, $R_{e12}=10$ kilohms, $R_e=29$ kilohms, $R_f=100$ kilohms, $E_{d1}=100$ millivolts. This can be reduced by chopper stabilization to 0.5 millivolt if drift reduction is 200; see equation 10.

For silicon transistors type *2N479*, $(\Delta I_{co1}-\Delta I_{co2})=0.01$ microampere maximum (for a sample of ten units), and this corresponds to an $E_{d2}=1$ millivolt, which is negligible when reduced by a factor of 200. However, the same theory will result in low drifts if germanium differential stages are used.

Practical Differential Amplifiers

The Slaughter amplifier is a differential amplifier with feedback, and it incorporates all the advantages of this type of operation. However, it requires that the output be single-ended and that it be in phase with the input. Fig. 6 illustrates a differential amplifier with feedback that allows for a single-ended output in either phase, or a differential output. The input can be single- or double-ended; in the former case the other input is grounded through an impedance equal to the source impedance. The input impedance is made high by feedback operation. The gain of this amplifier is the reciprocal of the feedback β ($A_f=(2R_f+R_1)/R_1$, double-ended output) if the open-loop gain times the feedback β is much greater than unity. These results are standard for feedback amplifiers except that the feedback β is the resistor ratio given above; it does not depend on the impedance seen looking back into the first emitter. This rather surprising

result has been confirmed by analysis. Incidentally, the analysis applies to a single-ended amplifier in which the feedback is introduced at the first-stage emitter in grounded emitter operation.

An amplifier was constructed according to Fig. 6 and performed as expected. The balance, gain, and level controls behave as they would in an equivalent tube-type amplifier. The gain control is optional but it is useful practically to stop oscillations if the open-loop gain is too high for the phase shifts introduced by the transistors. A gain of 10 was achieved with 30-db (decibel) feedback without frequency compensation with *2N479* silicon transistors. The divider network between the second and third stage is required to bring the output to zero d-c volts and is an ideal place to introduce frequency compensation if the transistor phase shifts cause oscillations. (The details of frequency stabilization are discussed in a later section.) The drift was, of course, a function of the difference in drift characteristics of the first-stage transistors.

To take advantage of chopper stabilization and differential stage drift reduction, an operational amplifier can be used with a differential input and a single-ended output. The chopper stabilization is effective, however, only for a single-ended input signal; Fig. 7 illustrates this type of amplifier. The gain is given by the ratio of R_f to R_i, the input impedance is the parallel combination of R_i and R_i', and the drift reduction is equal to R_i'/R_i times the net current gain of the a-c amplifier.

The a-c amplifier has a net phase reversal so that the single-pole double-throw chopping yields no net phase reversal for the direct current. It is possible to feed the output of the chopper amplifier into the base of TR_2 if the a-c amplifier and chopper have a net d-c phase reversal.

The important features in the design of this type of amplifier are summarized as follows:

D-C Design

The first and second stages are the differential type, and to determine their operating points the collector current is first specified. Since the emitter potentials are very close to the base, the common emitter resistor is calculated to have twice the individual collector currents flowing when it is operating between the base potential and the negative supply.

R_3 and R_4 are chosen high for good common mode rejection (see equation 26 for the common mode rejection formula)

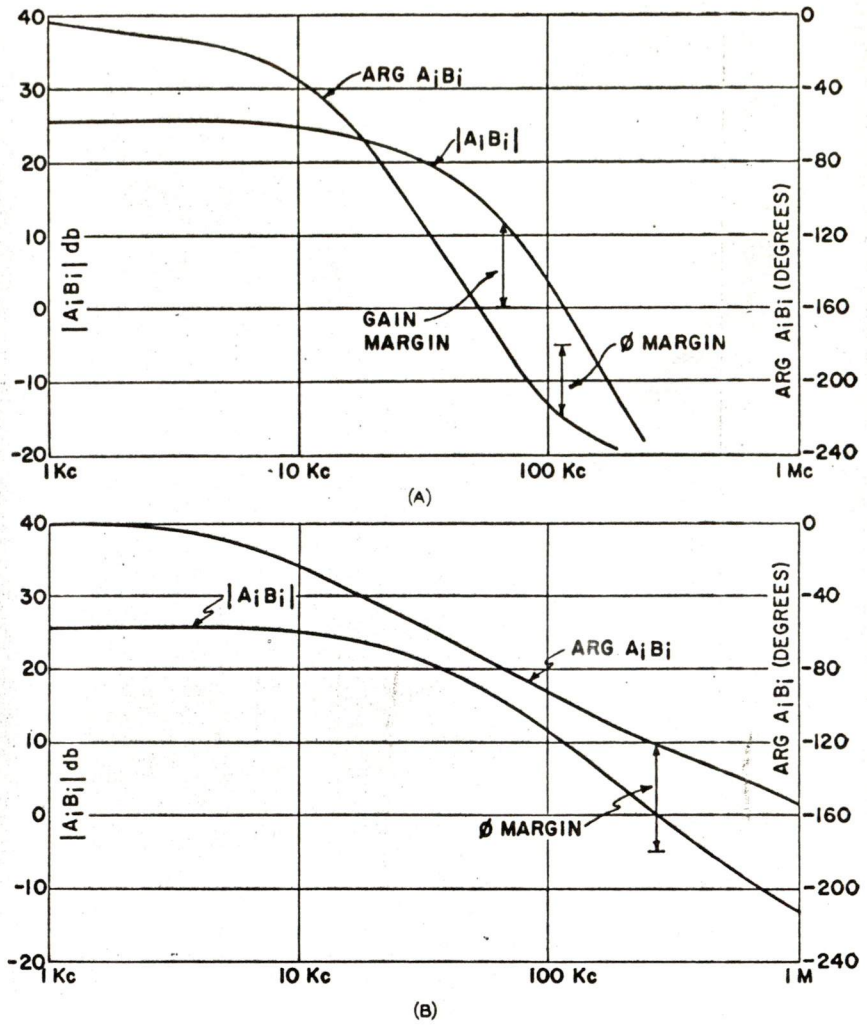


Fig. 8. A—Open-loop gain and phase response of transistor differential input operational amplifier with no frequency compensation. B—After stabilization to self-oscillation

and this determines their negative return potential.

The collector load resistance can then be selected so that it will be high (perhaps ten times as high) compared to the input impedance of the following stage. This is required since, ideally, all of the signal current should flow into the next stage. The collector supply can then be chosen large enough to allow for the signal swing, with the low value of a-c load of the next stage taken into account. For quiescent d-c conditions the small value of base bias current of the next stage can usually be neglected. The collector load of TR_4 is omitted so that it is actually an emitter follower. Since this stage is a double-ended to single-ended converter, it is not necessary, and the design is still the same since the collectors can usually be considered as current sources of high impedance.

R_6 and R_7 constitute a divider network to allow the output to be at zero d-c potential. Knowing the collector potential of TR_3 and realizing that the base of TR_5 is essentially at emitter supply potential, lead to the choice of R_5, R_6, and R_7, so that the base current of TR_5 is at the desired quiescent value. R_6 and R_5 divide the signal current so as to reduce the gain, and it is therefore desirable to keep R_6 at a minimum. However, this demands a maximum of bleeder current through R_5, R_6, and R_7 and a compromise is necessary. R_7 is made much higher than the input impedance of TR_5 in order not to divert a lot of signal current, and this usually requires its negative return supply to be rather high negatively.

It is possible to use an opposite semiconductor material for the middle stage and eliminate the divider loss; however, R_6 is an ideal location to shunt a capacitor to form a lead network for frequency stabilization. This will be discussed later.

Drift Design

Equations 31 and 32 yield the drift referred to the input where $R_{e12}=R_1$, R_i is the same, $R_f>>R_1$, $\Delta V_{be1}-\Delta V_{be2}$ is

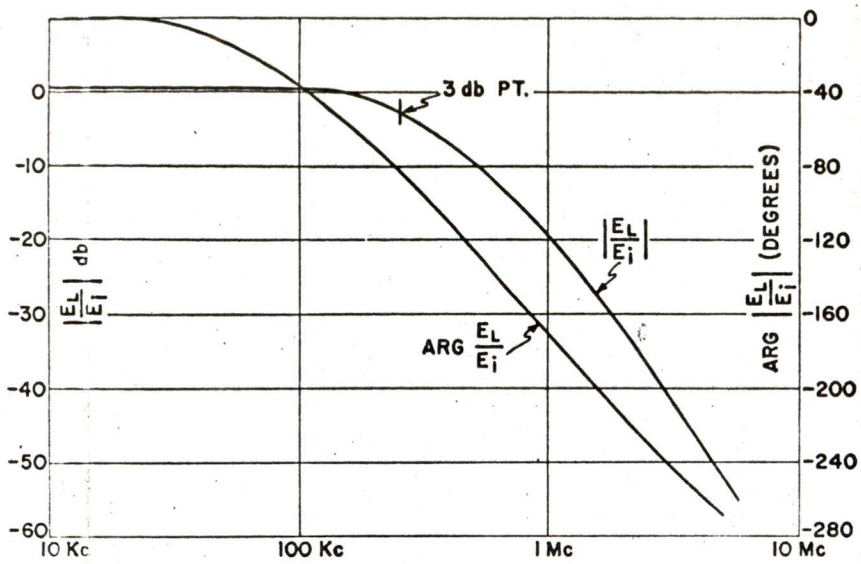


Fig. 9. **Closed-loop response of transistor differential input operational amplifier**

the maximum production spread of V_{be} versus temperature, and $\Delta I_{co1} - \Delta I_{co2}$ is the maximum production spread of I_{co} versus temperature.

Noise

The noise due to the transistors is mainly a function of the collector current in the first stage. This study involved very nominal noise specifications and the first-stage collector current was made relatively large (one milliampere). If this value is reduced to decrease noise effects, care must be taken to ensure that the transistor does not drift out of its active region with temperature. The small common mode gain of the differential amplifier is a distinct advantage in this respect compared with a single-ended type of amplifier.

A-C Design

If a large dynamic range is required, the d-c design must allow for the specified swing without saturating or cutting off. Again, the differential amplifier, due to its low common mode gain, helps maintain the room temperature operating conditions when the temperature varies.

To a first approximation, each transistor stage introduces a 6 db per octave slope with a corner frequency equal to the β cutoff frequency. Fig. 8(A) is a typical open-loop gain and phase response for a 3-stage amplifier shown in Fig. 7. Since there are three time constants involved, self-oscillations are possible, and indeed Fig. 7 indicates a negative gain and phase margin. Note that the gain drops off at a rate of 18 db per octave, indicating that the three β cutoff corner frequencies are close to each other. Note also that although the high-frequency equivalent circuit of a transistor would indicate more than one time constant, the additional time constants have corner frequencies far above the β cutoff frequency and therefore do not enter in the stabilization problem.

Fig. 8(B) represents the gain and phase response of the same amplifier after addition of capacitors C_2 and C_6 in Fig. 7. Note that a positive phase margin of 60 degrees has been obtained. The values of C_7 and C_8 are determined by standard techniques and will not be discussed here. There are many stabilization networks available; the two chosen are typical. Note that the use of the bleeder network R_5, R_6, and R_7 allow C_2 to form a simple cascade lead network.

Fig. 9 represents the closed loop (output divided by input) of the amplifier in Fig. 8(A) when the open-loop gain and phase were adjusted as in Fig. 8(B). The half-power point is at 250 kc. This could be raised by adjusting C_2 and C_6 in order to have a peaking in the closed loop response, but Fig. 9 represents the maximum response without peaking. Of course, peaking would reduce the phase margin of 60 degrees, but this is more than adequate to ensure against self-oscillation.

The capacitor C_5 of Fig. 7, can be used to shape the closed loop response without affecting the open-loop response of the amplifier. The reason for this is that the low input impedance at the base of TR_1 effectively short-circuits R_1 and C_5. The combination of R_1, C_5, and the input impedance of TR_1 form a lead network which which will introduce a rising 6 db per octave slope in the closed loop response, permitting a greater bandwidth than shown in Fig. 9.

Chopper Amplifier and Filter

The a-c amplifier shown in Fig. 7 represents a fairly simple design since it must amplify only the chopper frequency. A 3-stage common emitter germanium transistor amplifier was used with large emitter resistances to ensure a low stability factor. These resistors are heavily by-passed, so they do not reduce the gain at the chopper frequency. The combination of net phase reversal in the chopper amplifier and the single-pole double-throw chopping contacts yield a zero net phase shift for d-c signals. The d-c output is fed back to the base of the first transistor. If other combinations of a-c amplifier phase reversal and chopping contacts yield a net d-c phase reversal, the output can be fed into the base of TR_2 in Fig. 7.

C_1 together with R_1' form a filter to prevent chopping frequency signals to be fed back to the input, and C_2 and C_3 are blocking capacitors. R_9, C_4, and R_{10} are the output filter. To reduce the chopping frequency ripple to a negligible amount, the band pass of the entire chopper amplifier is reduced to a few cps (cycles per second). This is sufficient for correcting d-c drifts but it should be noted that a more complex filter will provide additional advantages. If the filter were designed to have a flat band pass close to the chopping frequency and the fall rapidly to give the desired attenuation at the chopping frequency, stabilization could be obtained over a reasonably wide bandwidth, one example might be to use a 400-cps chopping frequency, and a filter which was flat out to > 120 cps. This combination would allow simple power supplies to be used since the 60- to 120- cps ripple introduced by such supplies would be attenuated sharply by the chopper stabilization technique. The analysis of such stabilization would be similar to that already shown.

Appendix

The current operational amplifier of Fig. 2(B) can be analyzed assuming $-A_i$ is the open-loop current gain which in general is a function of Z_L and Z_f.

From Fig. 2(B):

$$I_1 = I_i + I_f = \frac{I_L + I_f}{-A_i} \tag{33}$$

Define

$$\beta_i = \frac{I_f}{I_R} \tag{34}$$

Equation 33 becomes

$$I_i + \beta_i I_L = \frac{I_L + \beta_i I_L}{-A_i}$$

$$\frac{I_L}{I_i}=\frac{-A_i}{(1+\beta_i+A_i\beta_i)}=\frac{-A_i}{1+\beta_i(1+A_i)} \quad (35)$$

From Fig. 2(B):

$$E_L=I_LZ_L=I_fZ_f+(I_i+I_f)Z_1 \quad (36)$$

where Z_1 is the input impedance of the current amplifier; from equation 33

$$I_i=-\frac{I_f(A_i+1)+I_L}{A_i}$$

Putting this into equation 36 results in

$$\frac{I_f}{I_L}\equiv\beta_i=\frac{Z_LA_i+Z_1}{A_iZ_f-(A_i+1)Z_1+Z_1A_i}$$

Dividing by A_i, and simplifying yields

$$\frac{I_f}{I_L}\equiv\beta_i=\frac{Z_L+\frac{Z_1}{A_i}}{Z_f-\frac{Z_1}{A_i}} \quad (37)$$

If, as is usually the case,

$$Z_L\gg\frac{Z_1}{A_i};\ Z_f\gg\frac{Z_1}{A_i}$$

$$\beta_i\doteq\frac{Z_L}{Z_f} \quad (38)$$

From equations 35 and 37

$$\frac{I_L}{I_f}=-\frac{(A_iZ_f-Z_1)}{Z_f+Z_1+Z_L(1+A_i)} \quad (39)$$

$$\doteq-\frac{Z_f}{Z_L} \text{ if } Z_f\gg\frac{Z_1}{A_i},\ Z_L\gg\frac{Z_f+Z_1}{1+A_i} \quad (40)$$

From Fig. 2(B)

$$E_L=I_LZ_L,\ E_i=Z_iI_i+(I_i+I_f)Z_1$$
$$=(Z_i+Z_1)I_i+Z_1I_f \quad (41)$$

and since $I_f=\beta_iI_L$, the result is

$$\frac{E_L}{E_i}=\frac{Z_LI_L}{(Z_i+Z_1)I_i+\beta_iZ_1I_L} \quad (42)$$

from equation 33

$$I_i=-\frac{I_f(A_i+1)+I_L}{A_i}$$

With the use of equation 34

$$I_i=-\frac{\beta_iI_L(A_i+1)+I_L}{A_i} \quad (43)$$

and equation 43 becomes

$$\frac{E_L}{E_i}=\frac{-A_iZ_L}{Z_i[\beta_i(A_i+1)+1]+2Z_i} \quad (44)$$

$$\doteq-\frac{Z_L}{Z_i\beta_i} \text{ if } A_i\beta_i\gg1,\ A_i\beta_i\gg\frac{2Z_i}{Z_i} \quad (45)$$

If, further, $\beta_i\doteq Z_L/Z_f$, then

$$\frac{E_L}{E_i}\doteq\frac{Z_f}{Z_i} \quad (46)$$

Summing up the approximate results yields

$$\frac{I_L}{I_i}\doteq-\frac{Z_f}{Z_L} \text{ if } Z_f\gg\frac{Z_1}{A_i} \text{ and } Z_L\gg\frac{Z_f+Z_1}{1+A_i}$$

$$\beta_i\equiv\frac{I_f}{I_L}\doteq\frac{Z_L}{Z_f} \text{ if } Z_L\gg\frac{Z_1}{A_i} \text{ and } Z_f\gg\frac{Z_1}{A_i}$$

$$\frac{E_L}{E_i}\doteq\frac{Z_f}{Z_i} \text{ if } A_i\beta_i\gg1,\ A_i\beta_i\gg\frac{2Z_1}{Z_i}, \text{ and}$$

$$\beta_i\doteq\frac{Z_L}{Z_f} \quad (47)$$

The approximate results are extremely useful since the conditions upon which they depend are usually very closely met.

Another useful result is the ratio of output voltage to input current. Define the gain impedance as $Z_g\equiv E_L/I_i$. Since $E_L=I_LZ_L$, $Z_g=I_L/I_i\ Z_L$. But $I_L/I_i\doteq-Z_f/Z_L$ and

$$Z_g\doteq-Z_f \quad (48)$$

That is, the gain impedance, or ratio of output voltage to input current is equal to the negative of the feedback impedance, which states that if A_i is high enough, and Z_1 is low enough, all of the input current flows through the feedback impedance and produces the output voltage.

References

1. Feedback Stabilized Transistor Amplifiers, D. W. Slaughter. *Electronics*, New York, N. Y., vol. 28, 1955, p. 174.

2. Some Transistor Input Stages for High Gain DC Amplifiers, G. B. B. Chaplin, A. R. Owens. Institution of Electrical Engineers, London, England, vol. 105, 1958, p. 249.

3. Stabilization of Wide-Band Direct-Current Amplifiers for Zero and Gain, A. W. Goldberg. *RCA Review*, Princeton, N. J., vol. 11, 1950, p. 296.

4. Transistor Circuits for Analog and Digital Systems, F. H. Blecher. *Bell System Technical Journal*, New York, N. Y., vol. 35, 1956, p. 295.

Instabilities of Push-Pull Magnetic Amplifiers Feeding the Field of an Electric Machine

H. F. STORM
MEMBER AIEE

THE CAUSES of instability of inductively loaded, center-tap magnetic amplifiers, and the remedies for this instability are well known.[1-4] If such a stabilized magnetic amplifier is used to energize the field of an electric machine, such as the shunt field of a d-c generator, the armature voltage of this machine will increase or decrease smoothly with the control voltage of the magnetic amplifier. Since the load current of the magnetic amplifier is unidirectional, the armature voltage of the d-c generator will also be unidirectional. However, in many closed-loop control systems it is necessary to obtain duodirectional output of the d-c generator. In this case the d-c generator will be equipped with two shunt fields; each shunt field is connected to a magnetic amplifier with such polarity that the magnetomotive forces produced by the two shunt fields are subtractive. Then, if the first magnetic amplifier is controlled "full on" and the second magnetic amplifier is controlled "full off," the armature of the d-c generator develops a voltage of one polarity, and if the the control to the magnetic amplifiers is reversed, the armature voltage of the d-c generator is also reversed. One may expect, furthermore, that if each magnetic amplifier is turned "one half on," the net magnetomotive force will be zero, and, hence, that the armature voltage of the d-c generator will also be zero; one may also expect that for a gradual increase of control current of one amplifier, and a similar gradual decrease of control current of the other amplifier, the armature voltage of the d-c generator will show a gradual change. The last two expectations, however, are not likely to be fulfilled. Instead, a minute change of the magnetic amplifier control currents will cause a sudden jump of armature voltage from perhaps +50% to −50% (or vice versa) of rated d-c generator armature voltage, and intermediate values of armature voltage will not be obtainable. This instability renders almost useless this otherwise inviting control scheme. It is suspected

Paper 56-730, recommended by the AIEE Magnetic Amplifiers Committee and approved by the AIEE Technical Operations Department for presentation at the AIEE Summer and Pacific General Meeting, San Francisco, Calif., June 25–29, 1956. Manuscript submitted April 12, 1956; made available for printing October 14, 1959.

H. F. Storm is with the General Electric Company, Schenectady, N. Y.

The author wishes to thank Manuel Ares for the experimental information contained in this paper.

PART 2

part 2. Differential amplifiers

Direct-coupled amplifiers are commonly employed in applications where a need exists to amplify small signals or small differences between two large signals. Often, these signals are no greater than a few millivolts and require a large amount of amplification. Many cases involve very slowly varying or d-c signals making the use of R-C coupling impractical or impossible and the designer must necessarily resort to a direct-coupling technique.

In some instances, chopper stabilized a-c amplifiers have been employed wherein the signal is chopped or converted to a-c at the chopping frequency, amplified, and then demodulated. Limitations involved in this scheme are the ability of choppers to work only at relatively low frequencies and the requirement that the signal frequency be much less than the chopping frequency. However, the latter system does avoid the problem of changes in the d-c bias conditions with time, or more importantly, with temperature.

The use of direct-coupled amplifiers without chopping is attractive from the standpoint of reduction in design complication. This approach affords savings in size, weight, cost and power, and, if needed, offers higher frequency capabilities. The biggest drawback to the use of d-c amplifiers in the past has been the problem of equivalent input drift voltage. The word "drift" refers to changes in the d-c bias conditions. It is significant in low level d-c amplifiers because bias variations are indistinguishable from the input signal, thereby introducing an amplified error at the output.

Since the signal level is lowest at the input, the drift problem in the first stage is of primary importance. One method of compensating for this unwanted signal is to use a balanced input stage more commonly referred to as a differential amplifier. If both input transistors are truly identical and both vary identically with temperature, any error signal introduced by one is balanced out by the other. Sperry has recently developed a low level, silicon planar differential amplifier unit which, for reasons explained below, will permit low level signal amplification, without chopping, over the temperature extremes of $-40^{\circ}C$ to $+80^{\circ}C$ with less than $3\mu v/^{\circ}C$ of drift error.

Assuming that the two input transistors are matched at room temperature, the problem of drift is due primarily to unequal changes with temperature of three temperature - dependent parameters: leakage current, I_{CO}; current gain, h_{FE} and the base-emitter voltage, V_{BE}. Since silicon devices are used, the contribution due to leakage may be minimized by designing for low d-c stability factor and by employing extremely low leakage transistors. The variation of h_{FE} with temperature is approximately 1%/°C. If the collector current is small and is held constant by employing a constant current source in the emitter circuit, the change in the base current due to changes in h_{FE} will be small. This small change in base drive then will cause only slight changes in the d-c base voltage.

The remaining and most important contributor to drift is the variation of V_{BE} with temperature. This parameter accounts for almost all of the error signal. The temperature coefficient, $\Delta V_{BE}/\Delta T$, is approximately -2500 μvolts/°C. Thus if the temperature of one transistor is 1/1000 of one degree different from the other, an equivalent input error voltage of 2.5 μvolts will result. It is therefore advantageous to mount the two transistors of the differential input stage on a common thermal base preferably in one package. In this way, each will experience very nearly the same junction to ambient temperature gradient. Differences in junction temperature due to power dissipation differences between transistors may be minimized by reducing the absolute power dissipated per junction. The most direct method of power reduction involves decreasing collector current, preferably down to the 1 to 10 μamp region. While operating at these low currents, the device requirement for useful current gain must be met. Sperry differential amplifier transistors have matched beta of 50 to 150 at the 1 to 10μA level.

As an illustration of the advantage of reduced power operation, consider the following example: the power dissipation of a device with a beta of 50 at I_C = 5μA and V_{CE} = 1.5v is approximately 7.5μW. A thermal resistance of 300°C/W, produces an associated rise in junction temperature of 0.0023°C. With conventional bias conditions of I_C = 100μA and V_{CE} = 5v, the power dissipation is approximately 500μW, thereby raising the junction temperature 0.151°C. A ten percent mismatch in either collector current or collector to emitter voltage would produce approximately 0.6μv drift voltage in the first case compared to 37.5μv error in the latter.

Another important consideration in low level amplifiers is noise. The noise decreases at low values of current as long as the leakage is much less than the collector current and good current gain is still available. Therefore, running at low currents as suggested above tends to reduce the equivalent input noise.

All of the characteristics outlined above are incorporated in Sperry low level transistors such as the 2N929A, 2N930A, 2N2523 and 2N2524. The Sperry 2N2722 differential amplifier transistor consists of two units similar in gain characteristics to the 2N930A mounted in a single can to minimize temperature differences as explained above. The individual current gains are matched at room temperature to within ten percent at a collector current of 1μA. Also, the base emitter voltages (V_{BE}) are matched to within 5mv under the same current and temperature conditions.

As a demonstration of low current differential amplifier operation, the circuit in Figure 10 was designed and built. The first stage is a Sperry 2N2722 unit (bias conditions: I_C = 5μA, V_{CE} = 1.5v) while the second stage consists of two low level complementary PNP planar transistors. Since the signal level is much higher in the second stage, it is not necessary to mount this pair in a single can. The fact that low level complementary PNP units are available, offers another significant advantage in that the signal level is automatically returned towards ground. The usual resistive divider networks are thereby eliminated, and only one power supply voltage is needed.

In Figure 11, the base-emitter voltage difference versus temperature of four 2N2722 units is plotted. Selecting the worst case, that of Unit No. 2, it can be seen that the ΔV_{BE} is less than 1mv from -50°C to +100°C.

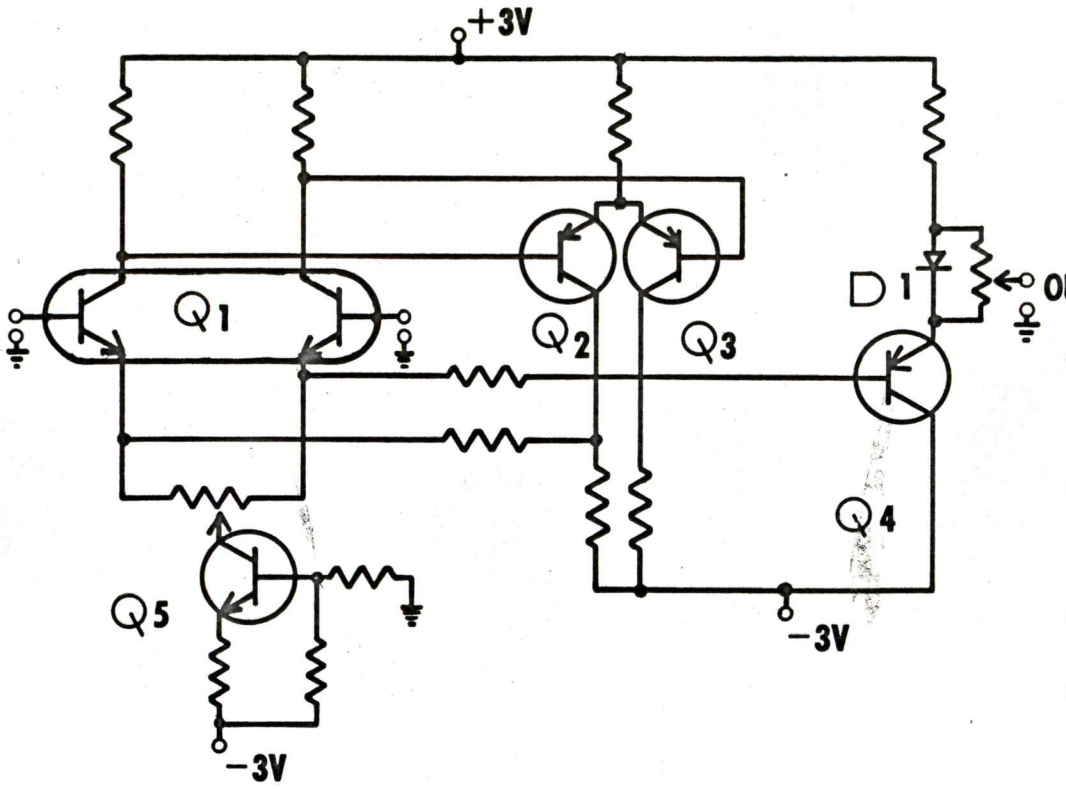
The equivalent input drift voltage versus temperature of the complete amplifier is shown in Figure 12, with four different 2N2722 units in the input stage. Unit No. 3 produced the largest observed error.
The equivalent input drift for this worst case is only 2.25 μvolts/°C from -40°C to +80°C.

The equivalent input noise voltage of the amplifier versus source resistance is shown in Figure 13, for three different bandwidths. Naturally, as the bandwidth is widened, the noise level increases. It is interesting that even a bandwidth of 1,570 cps, produces a noise level less than 0.5 μvolts for normal values of source resistance.

When the value of source resistance becomes very large, the noise level begins to rise. This increase in noise voltage is due to two effects. As R_S increases, the stability factor ($S=\Delta Ic/\Delta Ico$) increases. Thus the leakage current contribution to the noise voltage is enhanced. The other contribution is due to the pure thermal noise voltage of R_S, given by $\sqrt{4KT\Delta F\, R_S}$. This latter term obviously increases as R_S becomes large.*

*For further information on differential amplifiers, write for Applications Lab Report No. 3008 - "Improving Differential Amplifier Performance by Low Current Transistor Operation".

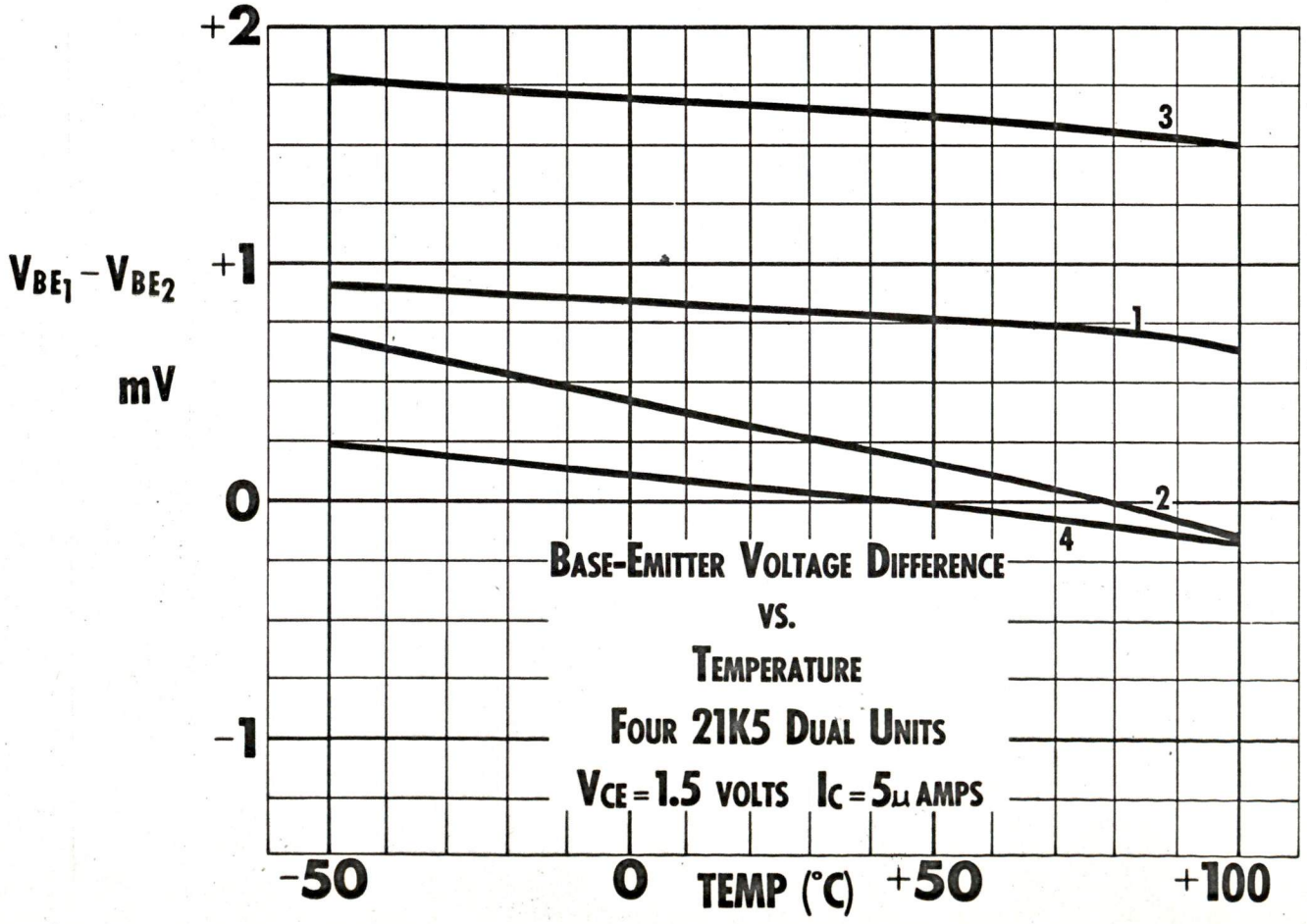
LOW LEVEL DIFFERENTIAL AMPLIFIER



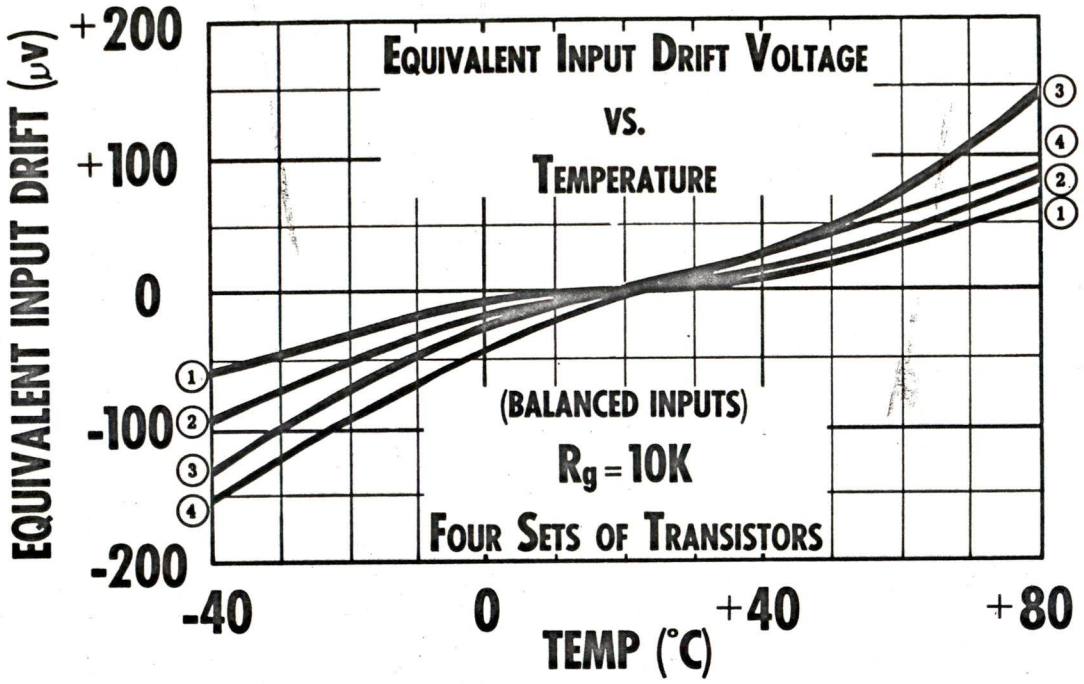
D_1	1N482
Q_1	2IK5
Q_2	2N2605
Q_3	2N2605
Q_4	2N2605
Q_5	2N2524

IO

II

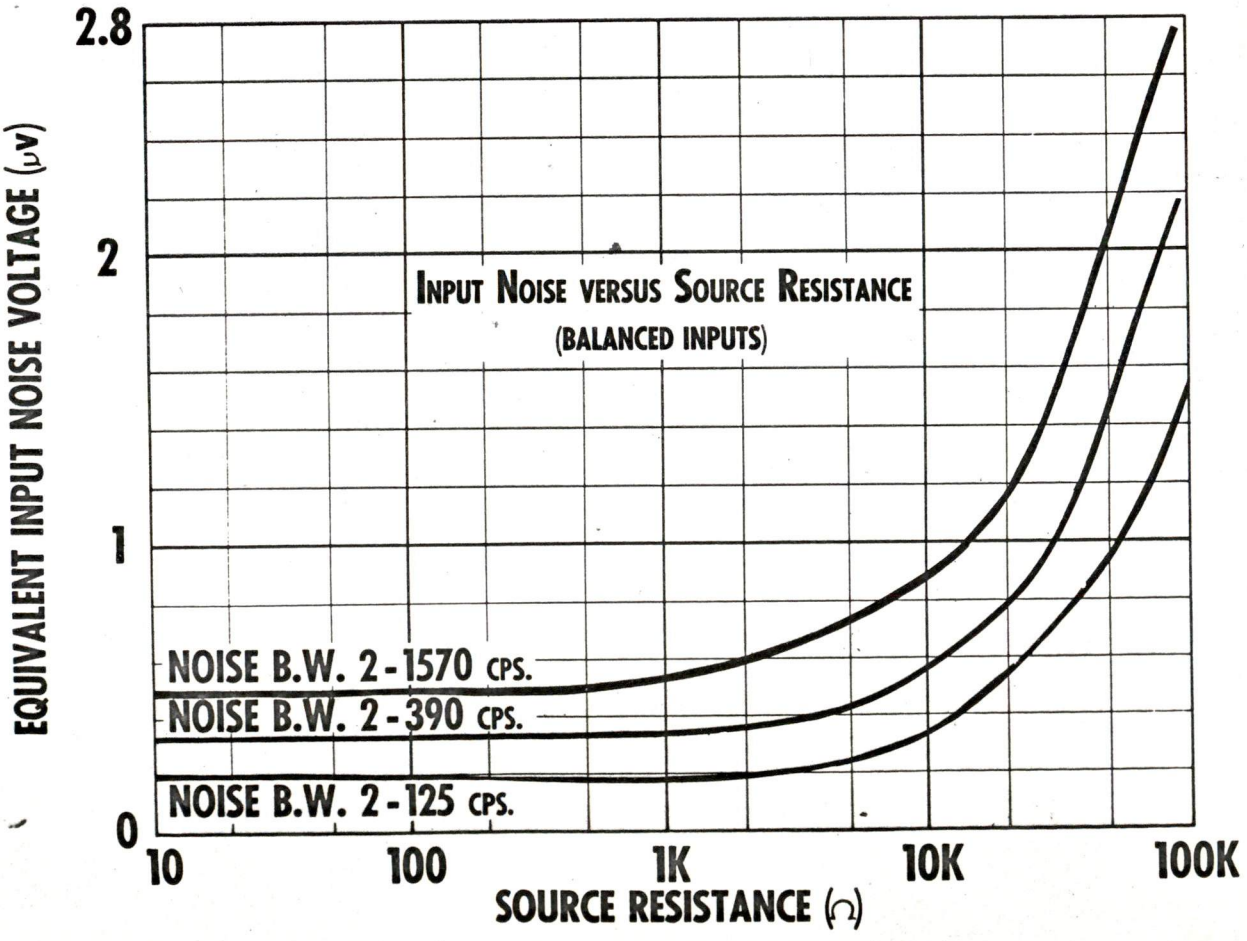


TYPICAL LOW LEVEL DIFFERENTIAL AMPLIFIER PERFORMANCE



12

13



DESIGN DECISIONS

Constant-Current Modules Use One Basic Board

Ronald M. Mann
Transistor Products Div.
Texas Instruments Inc.
Dallas, Tex.

DESIGN OF transistor testing equipment requires modules capable of delivering wide ranges of constant current to devices under test. To keep labor and overhead costs to a minimum, a printed-circuit module was developed that has wide versatility.

The basic circuit module in Fig. 1 can be used for currents from the microampere to the ampere range. For low currents, the main transistor is mounted on the board. For higher currents, the transistor is mounted externally on a heat sink and a cable of three wires is run to the socket holes on the module, as in Fig. 2. The printed-circuit board is shown in Fig. 3.

Values for components and the type and number of diodes and transistors are determined from; required current, or current range if the regulator is to be made variable; worst possible condition of impedance or resistance at the output terminals of the regulator; and power available.

Input should be a relatively constant voltage source. If the requirements are not too rigid, a good Zener-diode regulated supply may be used. The heart of the system is the reference voltage across CR_1 and CR_2 and resistor, R_3 in Fig. 4. These form a stable voltage reference for the series transistors.

As the output load varies, current I_4 will instantaneously try to change. This in turn will change the voltage across R_1 and R_2, causing a change in the voltage from the emitter of Q_1 to common, with respect to the reference point. This change in emitter voltage of Q_1 compared to the constant base voltage, E_3, will cause the series transistors to change their V_{CE}, thereby causing a change in the output to maintain constant current.

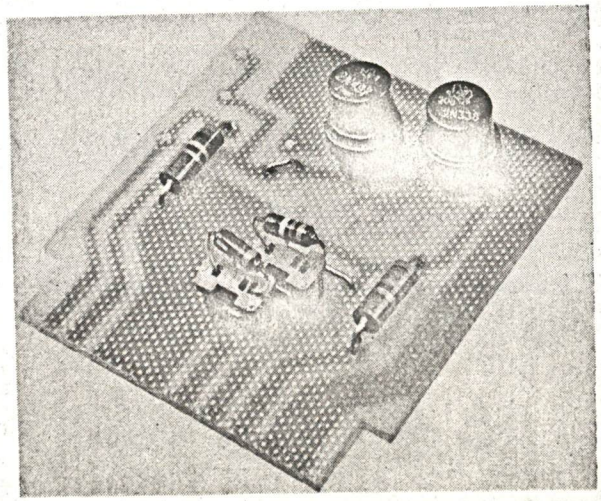


Fig. 1. Constant-current module for low-current application uses two board-mounted transistors.

General Design Procedure For Constant-Current Module

In the design of a specific unit, current I_4, and the highest impedance that may be obtained by the output of the regulator must be predetermined. These will determine the maximum output voltage at a constant current.

$$E_{max} = I_4 (Z_{max}) \quad (1)$$

The input voltage to the regulator is

$$E_{in} = 3E_{max} \quad (2)$$

With I_4 given, the transistors may be selected. For extremely low values of I_4, silicon transistors should be used so that the I_{CBO} is much less than I_4. In the moderate milliampere range, germanium transistors, such as the 2N1302, 2N1306 and 2N1308, can be used. Average dissipation required of Q_1 will be:

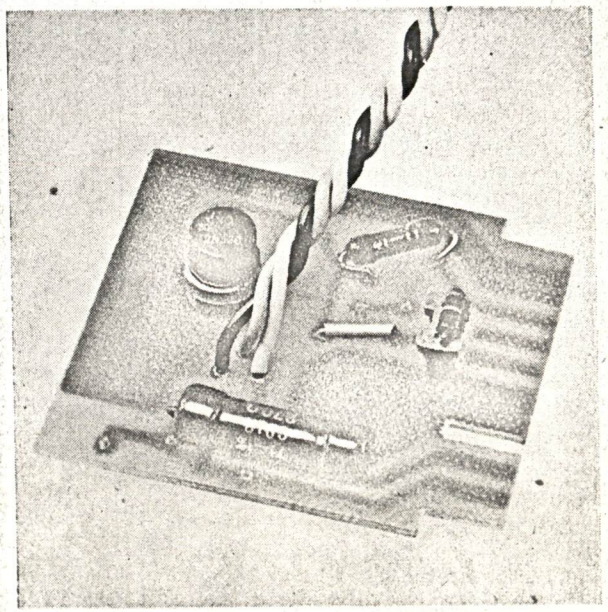


Fig. 2. For high-current modules main transistor is cable-mounted from heat sink.

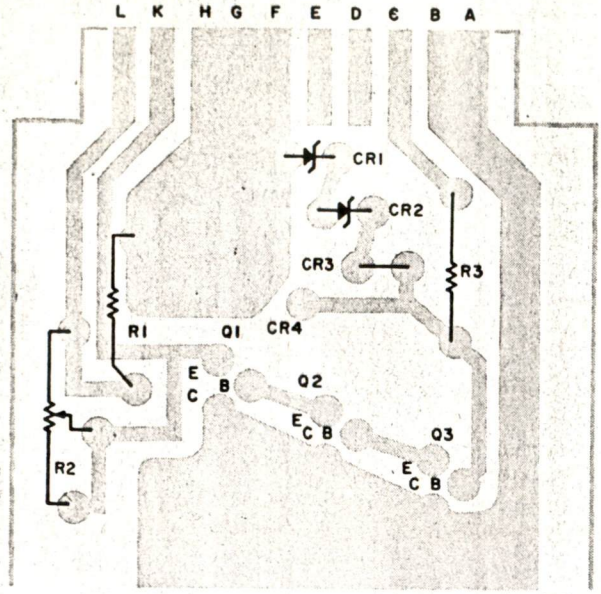


Fig. 3. Circuit board has mounting spaces for up to three transistors and three diodes.

$$P_1 = V_{CE1} \times I_4,$$

$$\text{let } V_{CE1} = E_{max} \qquad (3)$$

For stable operation, the dissipation of Q_1 should be approximately a third of the total dissipation given for the devices. If this requires a device where the I_{CBO} is greater than 1/10 I_4, then Q_1 should be selected for lower maximum heating dissipation qualities and better I_{CBO} characteristics. In this case, the device must be mounted on a heat sink. Current I_2 should be a negligible portion of the current flowing in R_3

$$I_3 = I_2 + I_1 \qquad (4)$$

if I_1 is very much greater than I_2 then

$$I_1 = I_3$$

The value of I_1 should be large enough so that it corresponds to a point far down on the break of the Zener-diode voltage characteristic curve. This minimizes the effects of change in temperature and changes in I_2.

The value of R_3 is determined as follows:

Since E_{in} is approximately $3E_{max}$, all of the input voltage will appear across voltage divider CR_1, CR_2 and R_3. The voltage across R_3 should be approximately twice output voltage. The value of R_3 is:

$$E_3 = E_{in} - V_{z1} - V_{z2} \qquad (5)$$

where V_{z1} and V_{z2} are the reference voltages of CR_1 and CR_2.

Let $V_{z1} + V_{z2} = V_z'$

from Eq. 4, $I_1 = I_3$

then $R_3 = V_3/I_3 = V_{in} - V_z'/I_1 \qquad (6)$

DESIGN DECISIONS

If Q_1 and Q_2 are needed, R_1 and R_2 are determined as follows:

Since $h_{FE} = I_C/I_B$ with V_{CE} constant and

if $I_4 = I_C$ for Q_1

then $I_{B1} = I_4/h_{FE1}$ (7)

where I_{B1} is the base current of Q_1 and h_{FE1} is the beta of Q_1.

If I_{B1} is less than one-tenth of I_4 then Q_2 is not needed. If it is greater than 1/10 I_4 then:

Let $I_{B1} = I_{E2}$

where I_{E2} is emitter current of Q_2

then $I_{E2} = I_{C2} + I_{B2}$ (8)

since for Q_2, I_B is much less than I_C.

Therefore $I_{E2} = I_{C2}$

Then $I_{B2} = I_{C2}/h_{FE2}$. (9)

Assuming I_B of Q_2 is very much less than I_1, Q_3 would not be needed. Using the base currents obtained for Q_1 and Q_2, determine the average V_{BE} per unit from the input characteristic curve.

Then $R_1 + R_2 = V_{in} - V_3 - V_{BE1}/I_4 + I_{BE}$ (10)

Most of the calculated resistance should be in R_1, then R_2 can be used to trim out difference caused by the actual voltage drops.

Design Procedure For Typical Application

For a practical example, a current of 0.5 ma is predetermined; the worst possible load is 6.0 v at 0.5 ma.

$$E_{in} = 3E_{max}$$

$$= 3 \times 6 = 18 \text{ v}$$

The power dissipation in Q_1 would be:

$$P_1 = V_{CE1} \times I_4$$

$$E_{max} = V_{CE1} \qquad I_4 = 0.5 \text{ ma}$$

Then $P_1 = 6.0 \times 0.5 \text{ ma} = 3.0 \text{ mw}$

A good low-leakage transistor with reasonable h_{FE} and breakdown voltage is the silicon 2N338 having the following characteristics: $h_{FE} = 45-150$; $P_{max} = 125$ mw; $BV_{CBO} = 45$ v; I_{CBO} = less than 1 μa

To find the reference point, the 1N700 Zener-diode characteristic curve shows that 10 ma is well down the voltage curve.

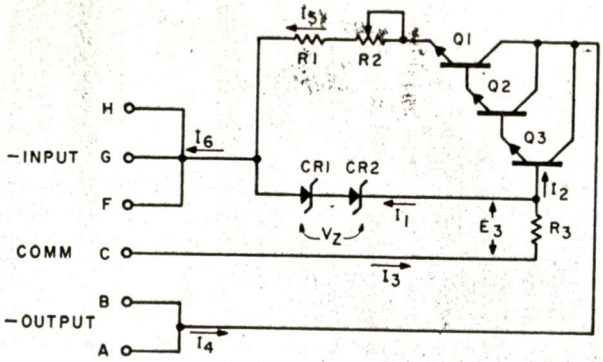


Fig. 4. Circuit arrangement for negative constant-current source.

This current also is well below maximum dissipation for the diode.

Since $E_{in} = 18$ v

then $V_{z\ total} = 1/3\ E_{in} = 6.0$ v

The 1N700 series data sheets show 6.2 v to be closest to the 6.0-v point, therefore, only one diode is needed for the reference.

Therefore $E_3 = E_{in} - V_z$

$= 18$ v $- 6.2$ v $= 11.8$ v

then $R_3 = V_3/I_3 = 11.8/10 = 1.18$ K

The closest standard value of resistance is 1.2 K;

therefore $I_3 = V_3/I_3 = 11.8$ v/1.2 K .

$= 9.84$ ma.

The new value for R_3 will not affect the reference voltage of the diode.

If the design is made for the lowest possible h_{FE} of 45

then $I_{B1} = I_4/h_{FE1}$

$= 0.5$ ma/45

$= 11\ \mu$a.

This current is well above the worst possible I_{CBO} given on the data sheet. Even with the highest possible h_{FE} of 150, 3.5 μa is still far above the worst leakage given. The worst case of base drive needed is 11 μa, which is close to 900 times less than I_3 of 9.84 ma, and satisfies the condition set up in Eq. 4; where I_B of $Q_1 = I_2$ which is very much less than I_3. Therefore, neither Q_2 nor Q_3 are needed.

Finding R_1 and R_2;

From Eq. (10) where $R_1 + R_2 = E_{in} - V_3 - V_{BE1}/I_4 + I_{BE1}$.

From the input characteristic curve for the 2N338, $V_{BE1} = 0.64$ v.

Since I_{B1} is very much less than I_4

then $R_1 + R_2 = 18$ v $- 11.8$ v $-$ (0.64 v/0.5 ma)

$= 18 - (12.44/0.5)$

$= 11.12$ K

therefore let $R_1 = 11$ K a standard value, and $R_2 = 0.5$ K pot for compensation.

The printed circuit board was designed by Electronic Fabricators Inc. of Dallas, Tex. ■ ■

RAPID CONSTANT-CURRENT DESIGN USING NOMOGRAPHS

RONALD M. MANN, *Engr., Texas Instruments Incorporated, Dallas, Tex.*

Active constant-current sources have many uses in industry, and are desirable over the high-voltage, large-resistance variety, in that a stable constant current can be achieved from short circuit to a large finite impedance provided E_o does not approach closely the voltage E1 on the emitter of the series regulator. A few examples are: forcing a constant collector or emitter current in transistor testing, forcing a constant charging current in a battery where there are charging current limits, feeding emitters of differential amplifiers to minimize drift, and use as ultra-high-impedance loads for transistor amplifiers to make high-gain voltage amplifiers.

The theory of operation and the design equations for active constant-current sources have been covered by the author in a previous article.[1]

Operation of a series-transistor current source of the type under discussion (Fig. 1) depends on the stable reference voltage created by the divider string made up of the breakdown diode D1 and resistor R3. The regulator strives to maintain E1 a constant. As the load requirements vary, a rapid change in I_o tries to occur, effecting a change in E_{R1+R2}, making a corresponding change in E1. This change is compared to the constant E3 by Q1, Q2 and Q3 and they are biased accordingly by the constant E3 to maintain E1. This is done by changing the V_{CE} of Q1, thereby changing E_o to maintain I_o as a constant. The conditions will remain static until E_o closely approaches E1, at which point I_o ceases to be constant and E_o remains a constant approximately equal to E1.

Requirements that must be known before starting a design are: (1) the current I_o to be supplied, (2) E_o maximum required by the load or (3) R_L maximum that will be attained. With the first requirement and either (2) or (3) design may proceed. If there are limits as to the input voltage, E_{in}, and current, I_E, these also need to be known; if not, they are determined in the design as follows:

Example:

(1) Let I_o = 10 ma
(2) Let R_L (max) = 1k, then
(3) E_o (max) = $(1 \times 10^3)(1 \times 10^{-2})$ = 10v

Step 1. Start with nomograph No. 1. Draw a line from reference point 1 to line 2, E_o maximum of 10v. Where the line crosses line 1, E_{in}, is the required input voltage to regulator. In this case E_{in} = 30v.

Step 2. Determine the power requirements of Q1 under operating conditions. Draw a line from line 2, E_o maximum, to line 4, I_o; where it crosses line 3, P1, is the power requirement. In this case P1 = 100 mw.

Step 3. Note values of P1 and E_o (max); these will be used later to help determine Q1. E_{in} = 30v P1 = 100 mw

Step 4. Determine a value for E_B by looking at the typical breakdown diode characteristics with

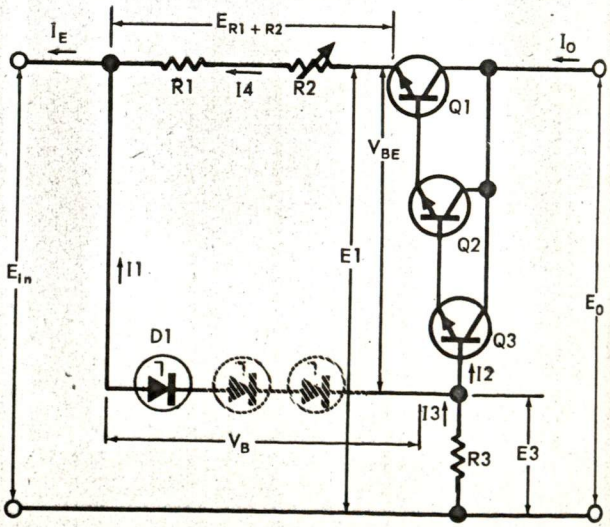


Fig. 1—SERIES-TRANSISTOR CURRENT SOURCE

1. Mann, Ronald M., "Constant Current Modules Use One Basic Board", Electronic Design; June 21, 1962; Vol. 10, No. 13; pp. 88-91.

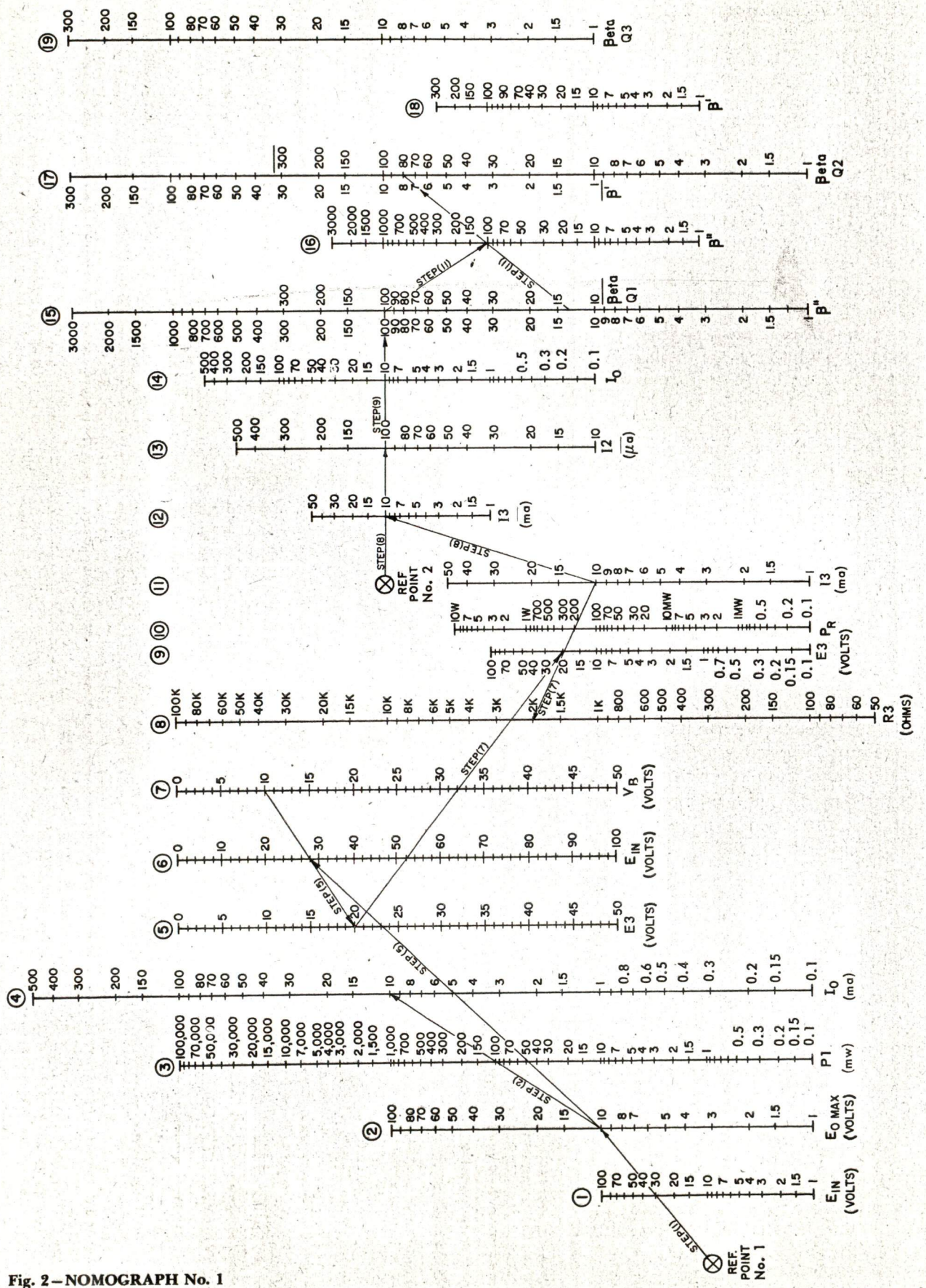


Fig. 2—NOMOGRAPH No. 1

(Continued on next page)

Circle Service No. 722 for free reprints of above article.

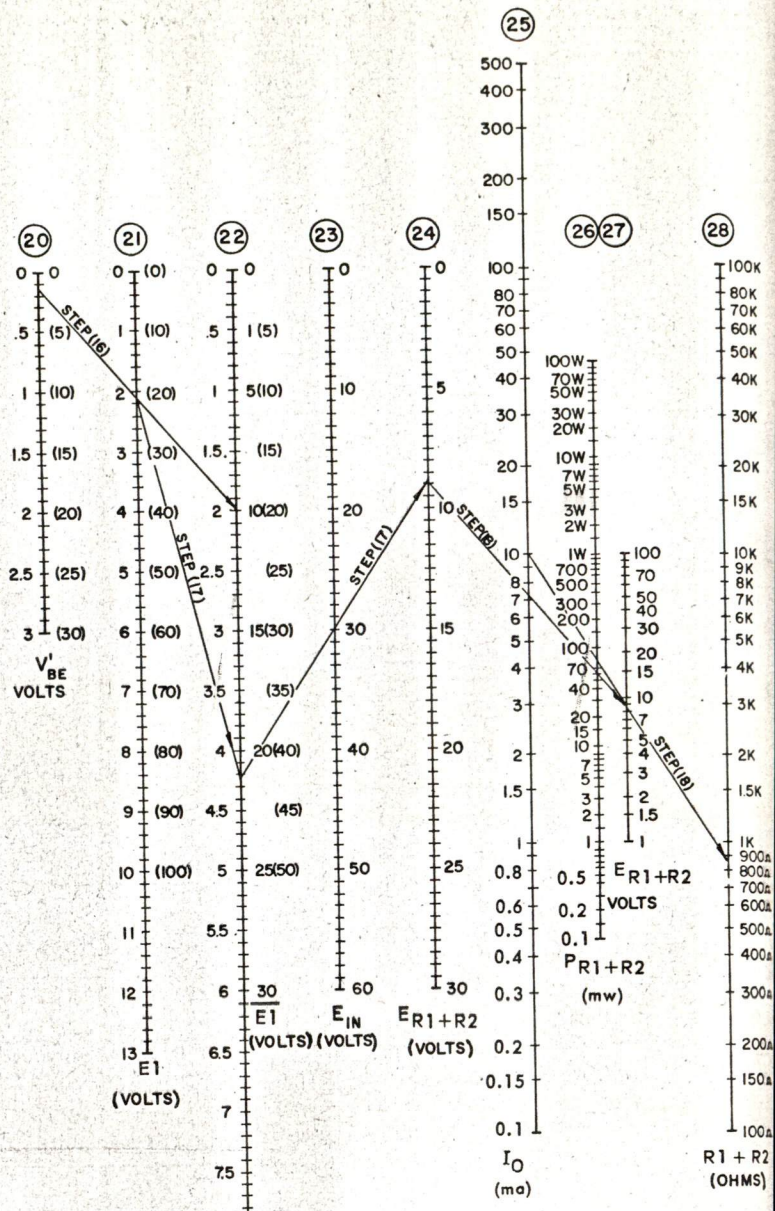
nomograph No. 2. Choose a diode or diodes whose total breakdown voltages approximate E_O (max). This gives approximately a 2 to 1 safety factor. In this case a single 1N758 was chosen with E_B = 10v.

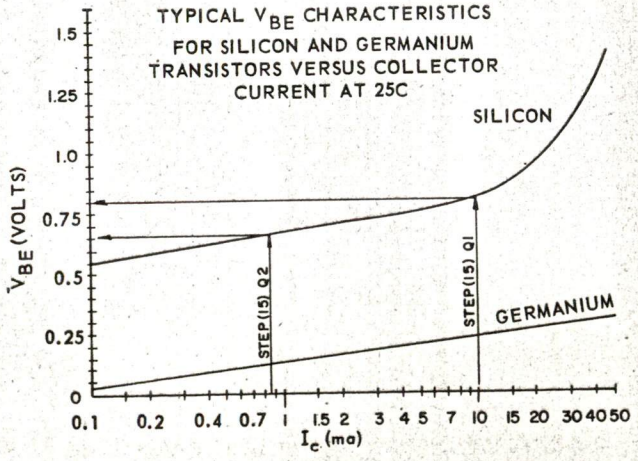
Step 5. Transpose the value for E_{in} found on line 1 to the proper point on line 6. On line 7, locate the E_B found in step 4. A line connecting these two points crossing line 5 gives a value for E3. In this case E3 = 20v.

Step 6. Go to typical breakdown diode curves with nomograph No. 2 and choose a current for I1 far down on the linear portion of the curve that does not exceed the 100-mw curve. (This gives another safety factor of 2 to 1.) I1 = 10 ma was chosen.

Step 7. Since, in final analysis, I1 will be found to be approximately equal to I3, I2 will be very small. Mark the point on line 11 corresponding to I1. Transfer the value found for E3 on line 5 to the corresponding value on line 9. Draw a line between the points; where they cross line 8 and 10 will be the resistance and the power rating of R3, which in this case are R3 = 2k, P_{R3} = 200 mw.

Step 8. Find value for I2. Transfer the value for I3 on line 11 to the corresponding value on line 12. Draw a line from reference point (2) through the point on line 12 to line 13. This gives a value for I2 (in this case) = 100 μa.





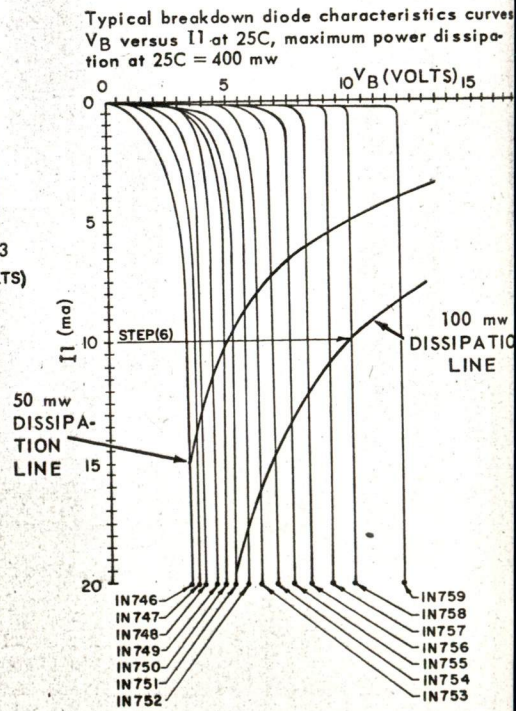


Fig. 3 – NOMOGRAPH No. 2

Step 9. Find a value for β'' (combined d-c beta needed, minimum). Draw line from value found for I2 on line 13 through point on line 14 corresponding to value of I_o (given initially) to point on line 15, giving a value for β''. In this case $\beta'' = 100$.

Step 10. With values in step (3) noted, choose a transistor for Q1, on the basis of a BV_{CEO} equal to or greater than E_{in} and a power dissipation of several times P1. Note the device chosen and the minimum and maximum beta given for the device. In this case a 2N497 was chosen for Q1. This is a Texas Instruments NPN silicon power transistor. (Transistors are NPN for negative input and output currents and voltages, and PNP for positive input and output voltages.)

$$\text{Beta (min)} \approx 13 = \beta_{Q1}$$
$$\text{Beta (max)} \approx 36$$

Step 11. Find a value for β'. Transfer value for β'' on the left side of line 15 to corresponding point on line 16. Mark value for β_{Q1} (min) on right side of line 15. Connect and draw line through to left side of line 17. This gives a value for β' of 8.

Step 12. Find minimum beta required for Q3. Since in this case β' was 8 and a 2N337 chosen for Q2 has a minimum beta of 20, there is no need for Q3. If Q3 were needed, β' would be transferred to line 18 and minimum beta for Q2 indicated on the right side of line 17. A line connecting these points intersecting line 19 gives minimum beta for Q3.

Step 13. If the exact maximum value for I2 is needed to be known for worst case conditions, insert minimum beta for Q2 from the data sheet on left side of line 17, connect with data sheet minimum value for beta for Q1 on right side of line 15. Line 16 now gives new value for β''. Transfer this value to left side of line 15. Draw line from it through value for I_o. New value on line 13 is maximum value for I2 in worst case conditions.

Step 14. Determine, in the case where both Q1 and Q2 are at maximum data sheet value of beta, if the leakage current will saturate Q1. Place point on left side of line 17 corresponding to maximum beta for Q2 and place point on right side of line 15 for maximum beta of Q1. Connect; new maximum value for β'' is given on line 16. Transfer to left side of line 15. On line 13 mark point for maximum given I_{CBO} for Q1; connect points; this gives value for I_o with no bias. If this value for I_o is not small compared to the I_o desired, then transistors with lower maximum betas must be chosen and/or leakage current in Q1 must be lowered. Since in this case both transistors are silicon, this step can be neglected since the leakage of Q1 is less than 1 μa.

Step 15. Determine minimum power rating of Q2. Place value for minimum beta for Q1 on left side of line 15; draw from this point through I_o point and at the intersection of line 13 is the current in Q2. Mark this current on line 4; connect to E_o maximum point on line 2 and the minimum power for Q2 is shown on line 3 in milliwatts. Usually this step may be omitted except when I_o is large and/or minimum beta of Q1 is very small.

Step 16. Determine V_{BE} of Q1 and Q2. For V_{BE} approximation go to typical V_{BE} curve shown; draw line from 10-ma point vertically to intersect silicon line. Use minimum data sheet beta of Q1 and divide into I_o to get collector current of Q2. Draw line from this value to silicon curve. Take the two V_{BE} values and add, giving V_{BE}'. In this case $V_{BE}' = 0.65 + 0.79 = 1.44$v.

Step 17. Find value for E1. Place value for V_{BE}' on line 20 and value for E3 from line 5 on line 22 and connect; the intersection of line 21 is the value for E1. In this case E1 = 21.44v. (Values in parentheses were used since E3 was larger than 10v.)

Step 18. Find voltage drop across R1 + R2 (E_{R1+R2}). Place value for E_{in} found on line 1 on line 23; transfer value of E1 found on line 22; connect points, and where line crosses line 24 is the value for E_{R1+R2}. In this case $E_{R1+R2} = 8.56$v.

Step 19. Find total resistance of R1 + R2 and power dissipation. Transfer value for E_{R1+R2} found on line 24 to corresponding point on line 27. Mark value for I_o on line 25, connect lines and where it intersects line 28 is the total resistance; intersection of line 26 is minimum allowable power rating of R1 + R2. In this case R1 + R2 = 856 ohms, $P_{R1+R2} = 85.6$ mw. Choose R1 as a standard value lower than the figure given and use a variable resistor for R2 to make up the difference to obtain exactly the current wanted.

Step 20. (Optional.) Use this step to obtain a variable constant-current supply. Use I_o as the maximum value of current and find R1 + R2 as in step 19. Choose a minimum value for I_o and find R1 + R2 as in step 19. Use this value for R1 + R2 and subtract from this maximum value the value for R1 + R2 at I_o maximum (R1), giving the value of the variable resistor R2.

Thus, by use of the nomographs and curves, a condensed transistor catalog, pencil and paper and a straightedge, constant-current sources can be designed for a wide variety of uses with reliable results.

Circle Service No. 722 for free reprints of above article.

Feedback-Stabilized

By DEAN W. SLAUGHTER
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ATTEMPTS to build a precision transistor amplifier with gain accurately stabilized by inverse feedback have generally failed because the low input impedance of the transistors loads down the feedback network. When the transistors are replaced or when the transistor parameters vary because of aging or changes in ambient temperature, the feedback factor may vary.

A transistor differential amplifier[1] employed in the first stage of a feedback amplifier provides a convenient terminal for feedback which does not load down the feedback network. In Fig. 1, e_1 is the equivalent input signal of source resistance R_{g1} and e_2 is the equivalent feedback voltage of source resistance R_{g2}. The stage gain is

$$e_o/e_2 = -e_o/e_1 \approx R_L\alpha_2/[r_{e1}+r_{e2}+(r_{b1}+R_{g1})(1-\alpha_1)+(r_{b2}+R_{g2})(1-\alpha_2)] \quad (1)$$

if $R_e >> r_e$; R_g, r_b and $R_L << r_c$.

Subscripts differentiate between the two transistors. Equation 1 shows that the amplified signal developed across R_L is proportional to the difference between e_1 and e_2. The presence of a small amplified signal voltage across R_L when $e_1 = e_2$ is due to common-mode effect, usually defined[2] as the ratio of this signal voltage to that which is present when either e_1 or e_2 alone is amplified. Its magnitude increases as the inequalities decrease. Common-mode effect due to signal-source resistance and common emitter resistance is

$$\delta \approx \frac{R_{g1}}{r_{e1}} - \frac{R_{g2}}{r_{e2}} + \frac{r_{e1}}{R_e} \quad (2)$$

Typical values of r_c and r_e for a junction transistor are 1 megohm and 40 ohms. A source resistance R_{g1} of 10,000 ohms results in a common-mode effect of about 1 percent.

The overall amplifier gain (Fig. 1) as determined by the feedback network is

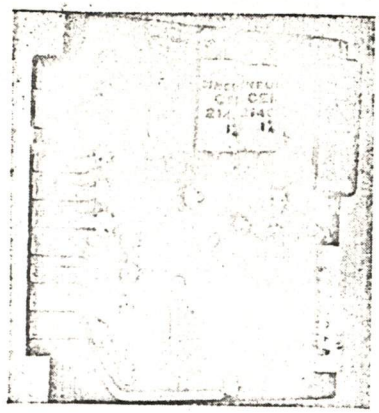
$$A = (R_1 + R_2)/R_1 \quad (3)$$

subject to two conditions. The first requires an amplifier internal gain without feedback A' much greater than gain with feedback A. The variation in A denoted as ΔA, resulting from variations in A' denoted at $\Delta A'$, is

$$\Delta A = \Delta A'A/A' \quad (4)$$

The second condition states that gain A varies in direct proportion to the amount of common-mode effect present in the differential-input stage, that is $\Delta A = \delta$. Variation in gain due to signal-source resistance R_{g1} or feedback circuit-source resistance R_{g2} can be computed directly from Eq. 2. Increasing R_{g1} or R_{g2} also reduces the differential-amplifier gain with a corresponding reduction in internal gain A'. Variation in gain resulting from the factor r_{e1}/R_e can be reduced to negligible proportions by replacing R_e with a constant-current source, as shown in Fig. 2.

Figure 3 is a feedback-stabilized amplifier employing germanium junction transistors. The gain A' without feedback is about 10,000. If R_1 and R_2 are selected for a gain of 10, it is possible to replace either of the transistors in the input circuit with an unselected transistor. The gain of 10 does not change by more than a few tenths of one percent.



Complete amplifier, shown actual size

Low-Drift Amplifier

Variations in base-to-emitter bias and collector cutoff current due to changes in ambient temperature may cause the output of transistorized d-c amplifiers to drift. However, when a differential circuit is employed in the first stage of these amplifiers, equal variations in the parameters of both transistors are not amplified because of the rejection of common-mode signals. For maximum reduction of d-c drift, matched transistors may be selected. They should be mounted in close proximity on a common heat sink.

The use of germanium transistors above room temperature in d-c amplifiers is not recommended. Figure 4 shows a d-c amplifier employing *npn* silicon transistors which perform satisfactory at 200 F or higher. By employing a differential circuit in the second as well as the first stage, a worthwhile increase in common-mode rejection and a decrease in overall amplifier drift is obtained. Power-supply regulation of from 2 to 5 percent is adequate. Amplifier internal gain is about 2,500.

Table I compares drift versus temperature of a differential amplifier employing silicon transistors with other amplifier configurations. Drift is given in terms of equiv-

Table I—Amplifier Drift

Input Stage	D-C Offset*	
	120 F	200 F
Germanium differential amplifier.	10 to 25 mv	
Single-ended silicon amplifier..	25 mv	120 mv
Silicon differential amplifier		
Pair 1.....	3 mv	10 mv
Pair 2.....	negligible	1 mv
Pair 3.....	2 mv	8 mv
Pair 4.....	negligible	1 mv
Pair 5.....	1 mv	3 mv

* Based on initial zero at 70 F

Transistor Amplifier

SUMMARY —— Gain of transistor amplifier is accurately stabilized by inverse feedback to avoid drift, from aging or temperature change. Amplifier can be built with silicon junction transistors for stable operation despite high ambient temperatures

alent input drift, that is that voltage which must be applied to the amplifier input terminal to return the output terminal to zero. These data assume that all the drift is generated by the transistors and none by the circuit resistances. Amplifier zero does not offset by more than a few millivolts over a period of hours or weeks if the amplifier is permitted a 20-minute warmup.

The supply voltages applied to the output stage TR_5 permit an output swing of at least ±10 volts. The 10,000-ohm collector load resistor may be safely decreased to 2,500 ohms without exceeding the collector dissipation rating of 150 mw. Because of feedback, the amplifier input impedance may approach but never exceed the collector resistance of TR_1. The output impedance is approximately equal to the output load resistance divided by the amplifier loop gain A'/A.

The resistor and capacitor connected between the collector of TR_4 and ground provide the phase correction necessary to prevent oscillation. Their values must be determined experimentally.

The values shown were suitable for a gain of 10. The resulting frequency-response curve is essentially flat to 10 kc, rises 3 to 6 db at 60 kc and cuts off at 150 kc.

This paper presents the results of one phase of research carried out under Contract No. DA-04-495-Ord 18, sponsored by the Department of the Army, Ordnance Corps.

REFERENCES

(1) D. W. Slaughter, The Transistor Emitter-Coupled Amplifier, WESCON paper, Los Angeles, 1954.

(2) G. Valley and H. Wallman, "Vacuum Tube Amplifiers," McGraw-Hill Book Co., Inc., New York, 1948.

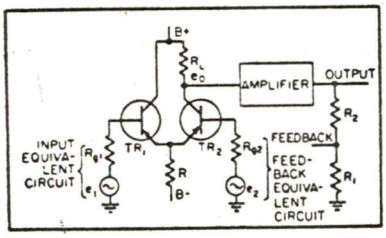


FIG. 1—Transistor differential amplifier

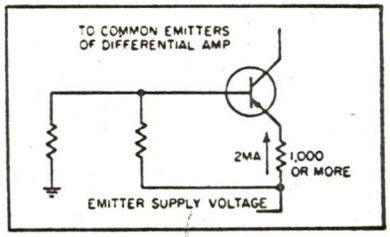


FIG. 2—Emitter constant-current source

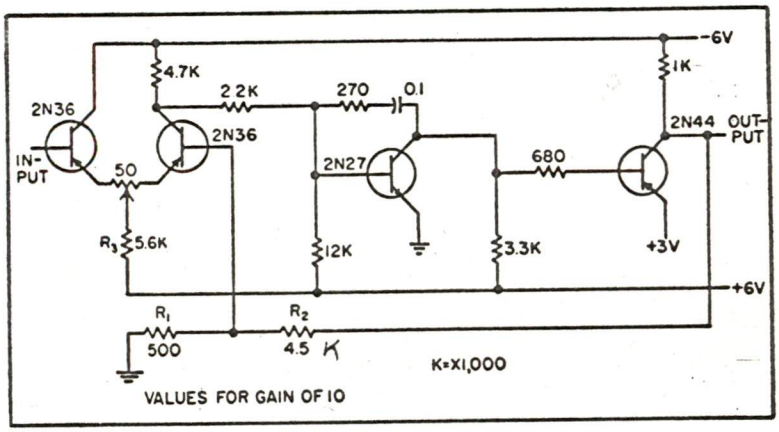


FIG. 3—Feedback-stabilized amplifier using germanium junction transistors

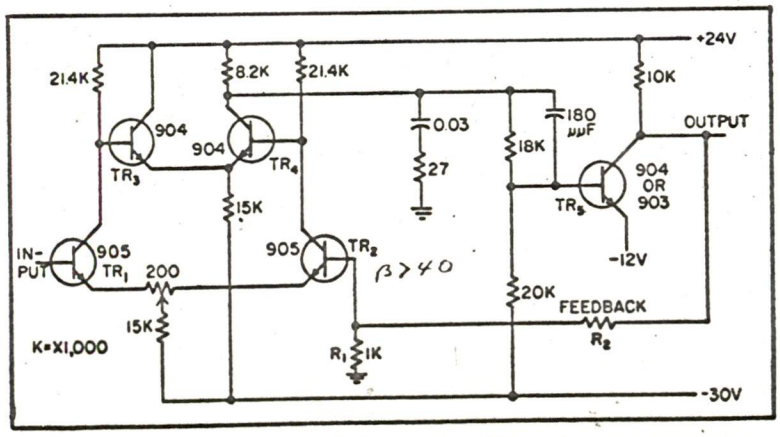


FIG. 4—Low-drift d-c amplifier employing npn transistors