The Bell System Technical Journal

Vol. XXVI | *July, 1947* | *No. 3*

Telephony By Pulse Code Modulation*

By W. M. Goodall

An experiment in transmitting speech by Pulse Code Modulation, or PCM, is described in this paper. Each sample amplitude of a pulse amplitude modulation or PAM signal is transmitted by a code group of ON-OFF pulses. 2^n amplitude values can be represented by an n-digit binary number code. For a nominal 4 kc. speech band these n ON-OFF pulses are transmitted 8000 times a second. Experimental equipment for coding the PAM pulses at the transmitter and decoding the PCM pulses at the receiver is described. Experiments with this equipment indicate that a three-unit code appears to be necessary for a minimum grade of circuit, while a six- or seven-unit code will provide good quality.

Introduction

THIS paper describes an experiment in transmitting speech by PCM, or pulse code modulation. The writer is indebted to his colleagues in the Research Department, C. E. Shannon, J. R. Pierce and B. M. Oliver, for several interesting suggestions in connection with the basic principles of PCM given in this paper. Work on a different PCM system was carried on simultaneously in the Systems Development Department of the Bell Laboratories by H S. Black. This in turn led to the development of an 8-channel portable system for a particular application. This system is being described in a forthcoming paper by H. S. Black and J. O. Edson.[1] A method for pulse code modulation is proposed in a U. S. Patent issued to A. H. Reeves.[2]

The material now presented is composed of three parts. The first deals with basic principles, the second describes the experimental PCM system, while the last discusses the results obtained.

Basic Principles

PCM involves the application of two basic concepts. These concepts are namely, the time-division principle and the amplitude quantization

* Paper presented in part at joint meeting of International Scientific Radio Union and Inst. Radio Engineers on May 5, 1947 at Washington, D. C.

[1] Paper presented on June 11, 1947 at A. I. E. E. Summer General Meeting, Montreal, Canada. Accepted for publication in forthcoming issue of *A. I. E. E. Transactions*.

[2] A. H. Reeves, *U. S. Patent* #2,272,070, Feb. 3, 1942, assigned to International Standard Electric Corp.; also, French patent #852,183, October 23, 1939.

principle. The essence of the time-division principle is that any input wave can be represented by a series of regularly occurring instantaneous samples, provided that the sampling rate is at least twice the highest frequency in the input wave.[3] For present purposes the amplitude quantization principle states that a complex wave can be approximated by a wave having a finite number of amplitude levels, each differing by one quantum, the size of the quantum jumps being determined by the degree of approximation desired.

Although other arrangements are possible, in this paper we will consider the application of these two basic principles in the following order. First the input wave is sampled on a time-division basis. Then each of the samples so obtained is represented by a quantized amplitude or integer number. Each of these integer numbers is represented as a binary number of n digits, the binary number system being chosen because it can readily be

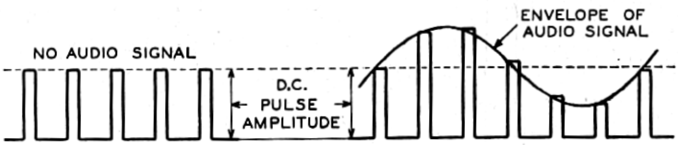

Fig. 1—Pulses in a PAM System.

represented by ON-OFF or two-position pulses. 2^n discrete levels can be represented by a binary number of n digits.[4] Thus, PCM represents each quantized amplitude of a time-division sampling process by a group of ON-OFF pulses, where these pulses represent the quantized amplitude in a binary number system.

The discussion so far has been in general terms. The principles just discussed will now be illustrated by examples.

Multiplex transmission of speech channels by sending short pulses selected sequentially from the respective speech channels, is now well known in the telephone art and is called time-division multiplex. When the pulses consist simply of short samples of the speech waves, their varying amplitudes directly represent the speech waves and the system is called pulse amplitude modulation or PAM. In PAM the instantaneous amplitude of the speech wave is sampled at regular intervals. The amplitude so obtained is trans-

[3] This is because the DC, fundamental and harmonics of the wave at the left in Fig. 1 all become modulated in the wave at the right, and if the highest modulating frequency exceeds half the sampling rate, the lower sideband of the fundamental will fall in the range of the modulating frequency and will not be excluded by the low-pass filter. The result is distortion.

[4] In a decimal system the digits can have any one of 10 values, 0 to 9 inclusive. In a binary system, the digits can have only two values, either 0 or 1.

mitted as a pulse of corresponding amplitude. In order to transmit both positive and negative values a constant or d-c value of pulse amplitude can be added. (See Fig. 1.) When this is done positive values of the information wave correspond to pulse amplitudes greater than the constant value while negative values correspond to pulse amplitudes less than the constant value. At the receiver a reproduction of the original speech wave will be obtained at the output of a low-pass filter.

The PCM system considered in this paper starts with a PAM system and adds equipment at the terminals to enable the transmission of a group of ON-OFF pulses or binary digits to represent each instantaneous pulse amplitude of the PAM system. Representation of the amplitude of a single PAM pulse by a finite group of ON-OFF pulses or binary digits requires quantization of the audio wave. In other words, we cannot represent the actual amplitude closer than $\frac{1}{2}$ "quantum". The number of amplitude levels required depends upon the grade of circuit desired. The disturbance which results from the quantization process has been termed quantizing noise. For this type of noise a signal-to-noise ratio of 33 db would be obtained for 32 amplitude levels and this grade of circuit was deemed sufficiently good for a preliminary study. These 32 amplitude levels can be obtained with 5 binary digits, since $32 = 2^5$.

Figure 2 shows how several values of PAM pulse amplitude can be represented by this binary code. The first column gives the digit pulses which are sent between the transmitter and receiver while the second column shows the same pulse pattern with each pulse weighted according to its assigned value, and the final column shows the sum of the weighted values. The sum, of course, represents the PAM pulse to the nearest lower amplitude unit. The top row where all the digits are present shows, in the middle wave form, the weighted equivalent of each digit pulse. By taking different combinations of the five digits all integer amplitudes between 31 and 0 can be represented. The examples shown are for 31, 18, 3, and 0.

Referring to Fig. 3 sampling of the audio wave (a) yields the PAM wave (b). The PAM pulses are coded to produce the code groups or PCM signal (c). The PCM pulses are the ones sent over the transmission medium. For a sampling rate of 8000 per second, there would be 8000 PAM pulses per second for a single channel. The digit pulse rate would be 40,000 pps for a five-digit code. For a time-division multiplex of N channels both of these pulse rates would be multiplied by N.

Wave form (d) shows the decoded PAM pulses where the amplitudes are shown under the pulses. The original audio wave is repeated as wave form (e). It will be noted that the received signal is delayed by one PAM pulse interval. It is also seen that the decoded pulses do not fit exactly on this curve. This is the result of quantization and the output of the low-pass filter will contain a quantizing disturbance not shown in (e) which was not present in the input signal.

A signal that uses regularly occurring ON-OFF pulses can be "regenerated" and repeated indefinitely without degradation. A pulse can be "regenerated" by equipment which transmits an undistorted pulse provided a somewhat distorted pulse is received, and transmits nothing otherwise.

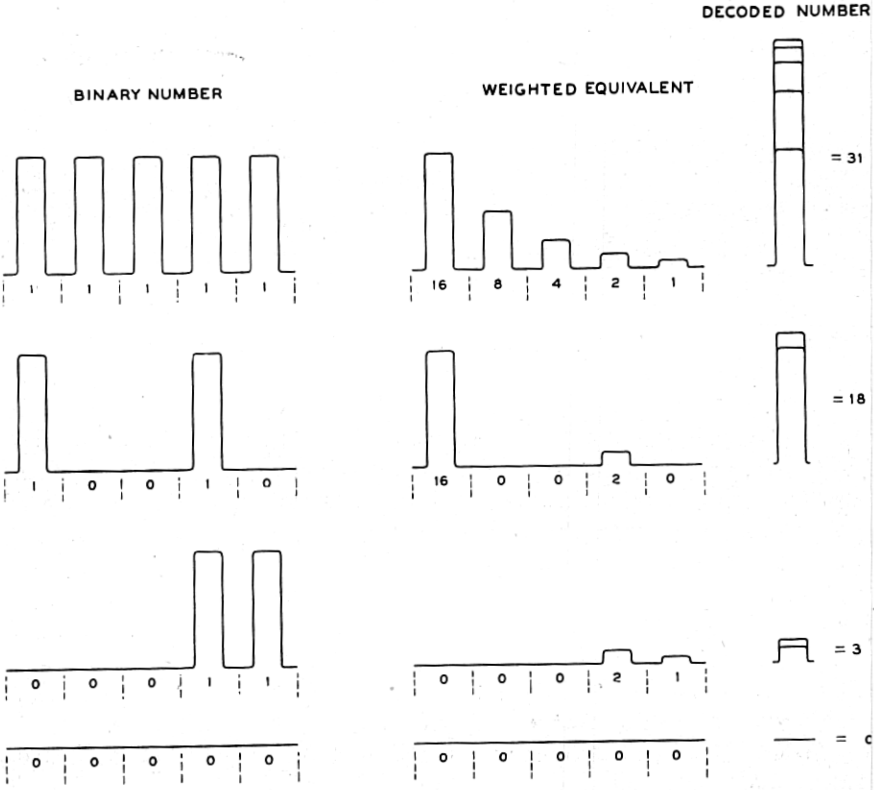

Fig. 2—Binary and decimal equivalents.

Thus, the received signal at the output of the final decoder is of the same quality as one produced by a local monitoring decoder. To accomplish this result, it is necessary, of course, to regenerate the digit pulses before they have been too badly mutilated by noise or distortion in the transmission medium.

The regenerative property of a quantized signal can be of great importance in a long repeated system. For example, with a conventional system each repeater link of a 100-link system must have a signal-to-noise ratio 20 db

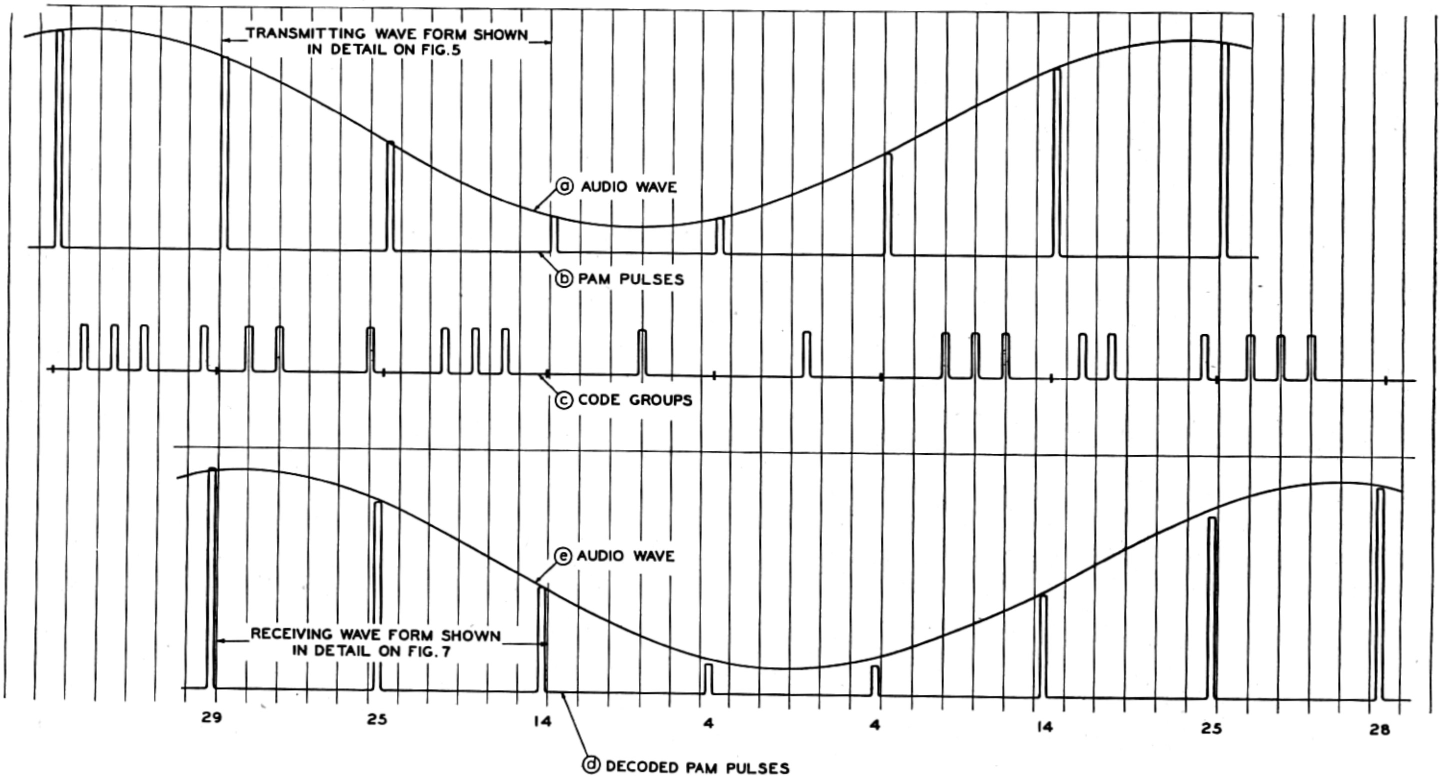

Fig. 3—PAM and PCM transmitting and receiving wave forms (amplitude vs. time).

better than the complete system. For PCM, however, with regenerative repeaters the required signal-to-noise ratio in the radio part of the system is independent of the number of links. Hence, we have a method of transmission that is ideally suited to long repeated systems.

At this point we might consider the bandwidth required to send this type of signal. For a 5-digit code the required band is somewhat less than 5 times that required for a PAM system. It is somewhat less than 5 times because in a multiplex system crosstalk becomes a serious problem. In a PAM system this crosstalk would add up on a long system in somewhat the same manner as noise. In order to reduce the crosstalk it would probably be necessary to use a wider band for the PAM repeater system than would be required for a single-link system. For PCM, on the other hand, by using regeneration the whole system requirement for crosstalk can be used for each link. In addition, a relatively greater amount of crosstalk can be tolerated since only the presence or absence of a pulse needs to be determined. Both of these factors favor PCM. This is a big subject and for the present we need only conclude that from considerations of the type just given the bandwidth penalty of PCM is not nearly as great as might first be expected.

The same two factors that were mentioned in connection with crosstalk also apply to noise, and a PCM signal can be transmitted over a circuit which has a much lower signal-to-noise ratio than would be required to transmit a PAM signal, for example.

Hence, we conclude that PCM for a long repeated system has some powerful arguments on its side because of its superior performance even though it may require somewhat greater bandwidth. There are other factors where PCM differs from more conventional systems but a discussion of these factors is beyond the scope of this paper.

The previous discussion may be summarized as follows: One begins with a pulse amplitude modulation system in which the pulse amplitude is modulated above and below a mean or d-c value as indicated in Fig. 1. It is assumed that it will be satisfactory to limit the amplitude range to be transmitted to a definite number of amplitude levels. This enables each PAM pulse to be represented by a code group of ON-OFF pulses, where the number of amplitude levels is given by 2^n, n being the number of elements in each code group. With this system the digit pulses can be "regenerated" and the quality of the overall transmission system can be made to depend upon the terminal equipment alone.

Experimental PCM Equipment

The experimental coder used in these studies might be designated as one of the "feedback subtraction type". It functions as follows: Each PAM pulse is stored as a charge on a condenser in a storage circuit. (See Fig. 4.)

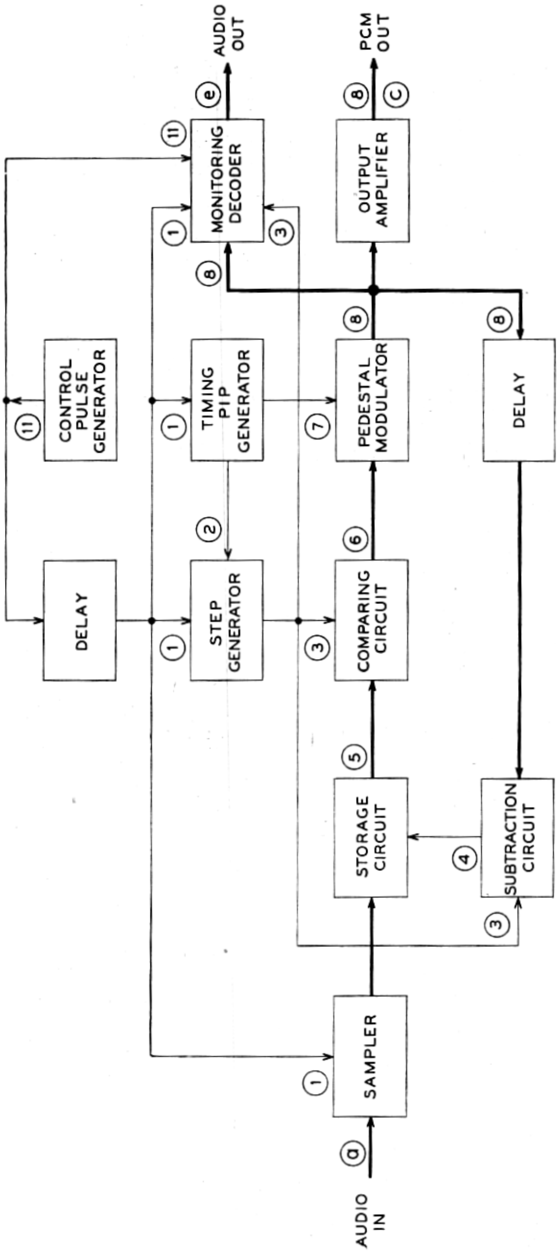

Fig. 4—PCM Transmitter block schematic.

The voltage across this condenser is compared with a reference voltage. The magnitude of this reference voltage corresponds to the d-c pulse amplitude of Fig. 1. The voltage has a magnitude of 16 units. If the magnitude of the condenser voltage exceeds the magnitude of the 16-unit voltage, a positive pedestal voltage is obtained in the output of the comparing circuit. This pedestal voltage is amplified, limited and applied to the pedestal modulator. The pedestal modulator serves as a gate for timing pulses from the timing pip generator. If the pedestal voltage and timing pulse are applied simultaneously to the pedestal modulator, a pulse is obtained in the output. In the present case this pulse corresponds to the presence of the 16-unit digit in the code group which represents this PAM pulse. This digit pulse after amplification and limiting is (1) sent out over the line (PCM out) and (2) fed back through a suitable delay circuit to a subtraction circuit. The function of the subtraction circuit is to subtract a charge from the condenser corresponding to the 16-unit digit. The charge remaining on the condenser is now compared with a new reference voltage which is $\frac{1}{2}$ the magnitude of the first reference voltage or 8 units. If the magnitude of the voltage across the condenser exceeds this new reference voltage the above process is repeated and the second digit pulse is transmitted and another charge, this time corresponding to the 8-unit digit, is subtracted from the remaining charge upon the condenser.

If the magnitude of the voltage across the condenser is less than the reference voltage, in either case above, then no pedestal will be produced and no digit pulse be transmitted. Since no pulse is transmitted, no charge will be subtracted from the condenser. Thus the charge remaining upon the condenser after each operation represents the part of the original PAM pulse remaining to be coded. The reference voltage wave consists of a series of voltages each of which is $\frac{1}{2}$ of the preceeding one. There is one step on the reference voltage function for each digit to be coded.

A better understanding of the coding process can be had by reference to the various wave forms involved. For completeness, wave forms from audio input to the coded pulse signal are shown for the transmitter in Figs. 3 and 5 and from the coded pulse signal to audio output for the receiver in Figs. 7 and 3. In the diagram the abscissas are time and the ordinates are amplitudes. Some of these wave forms have already been discussed in connection with Fig. 3. Since the coder functions in the same manner for each PAM pulse the detailed wave forms of the coding and decoding processes are shown for only two amplitudes. The block schematic for the transmitter is given on Fig. 4, while that for the receiver is given in Fig. 6. The letters on Figs. 4 and 6 refer to the wave forms on Fig. 3, while the numbers refer to the wave forms in Figs. 5 and 7.

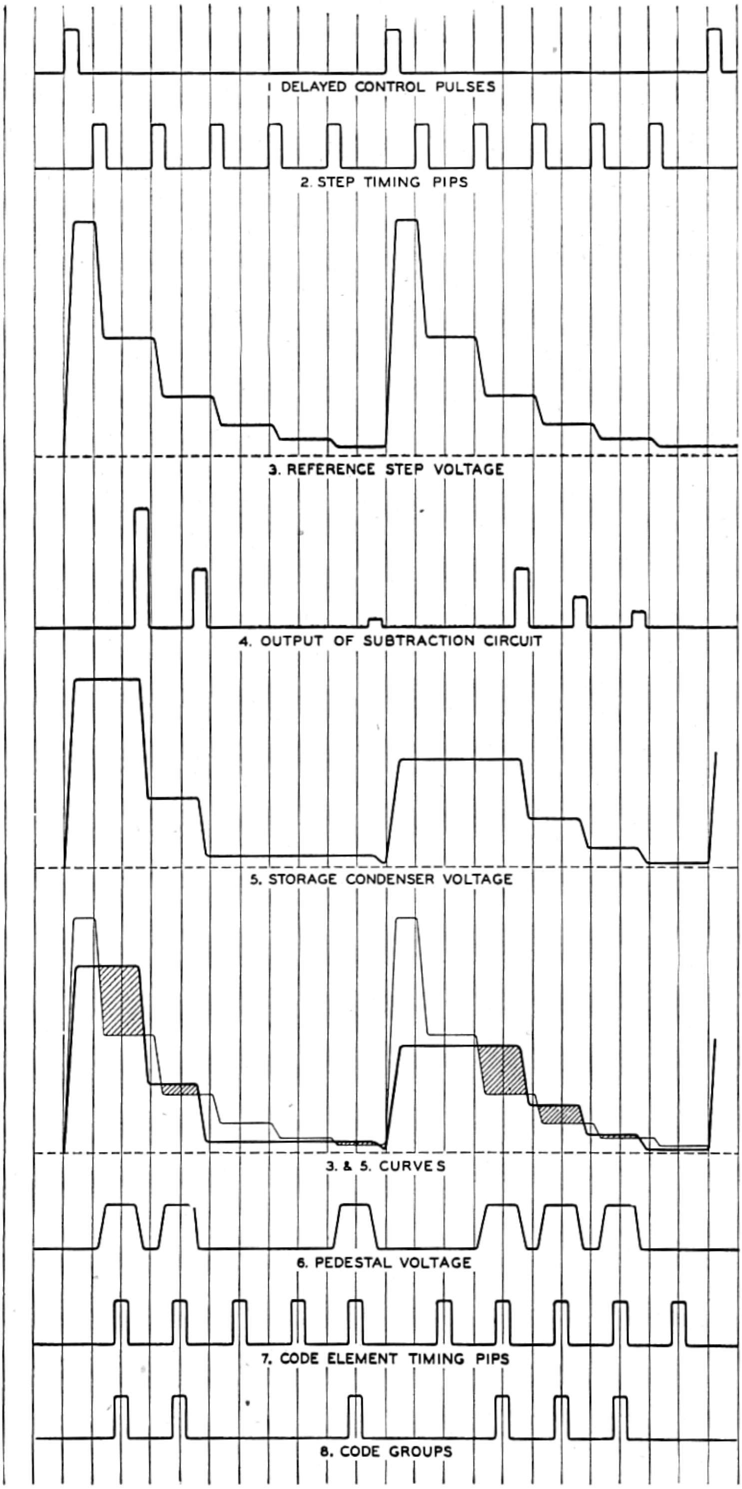

Fig. 5—Detailed wave forms for PCM Transmitter (amplitude vs. time).

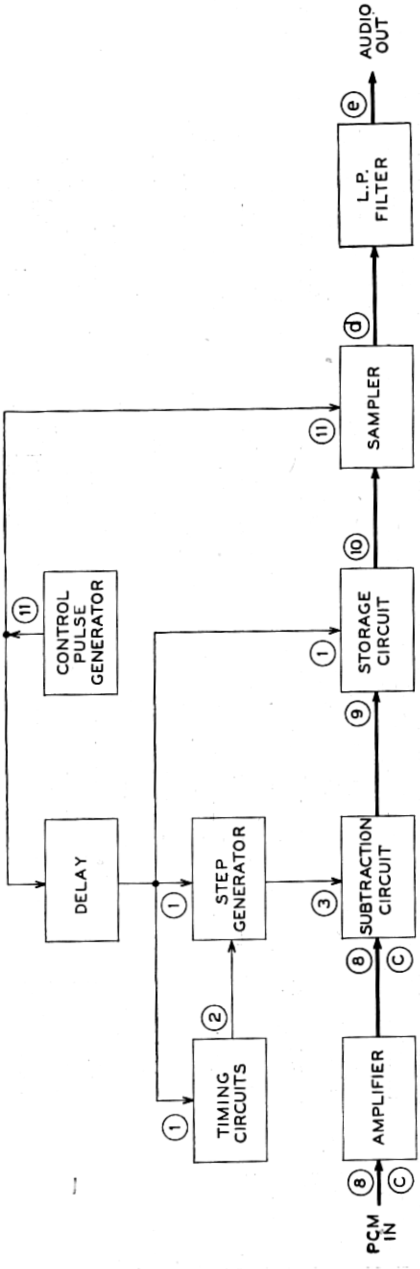

Fig. 6—PCM Receiver block schematic.

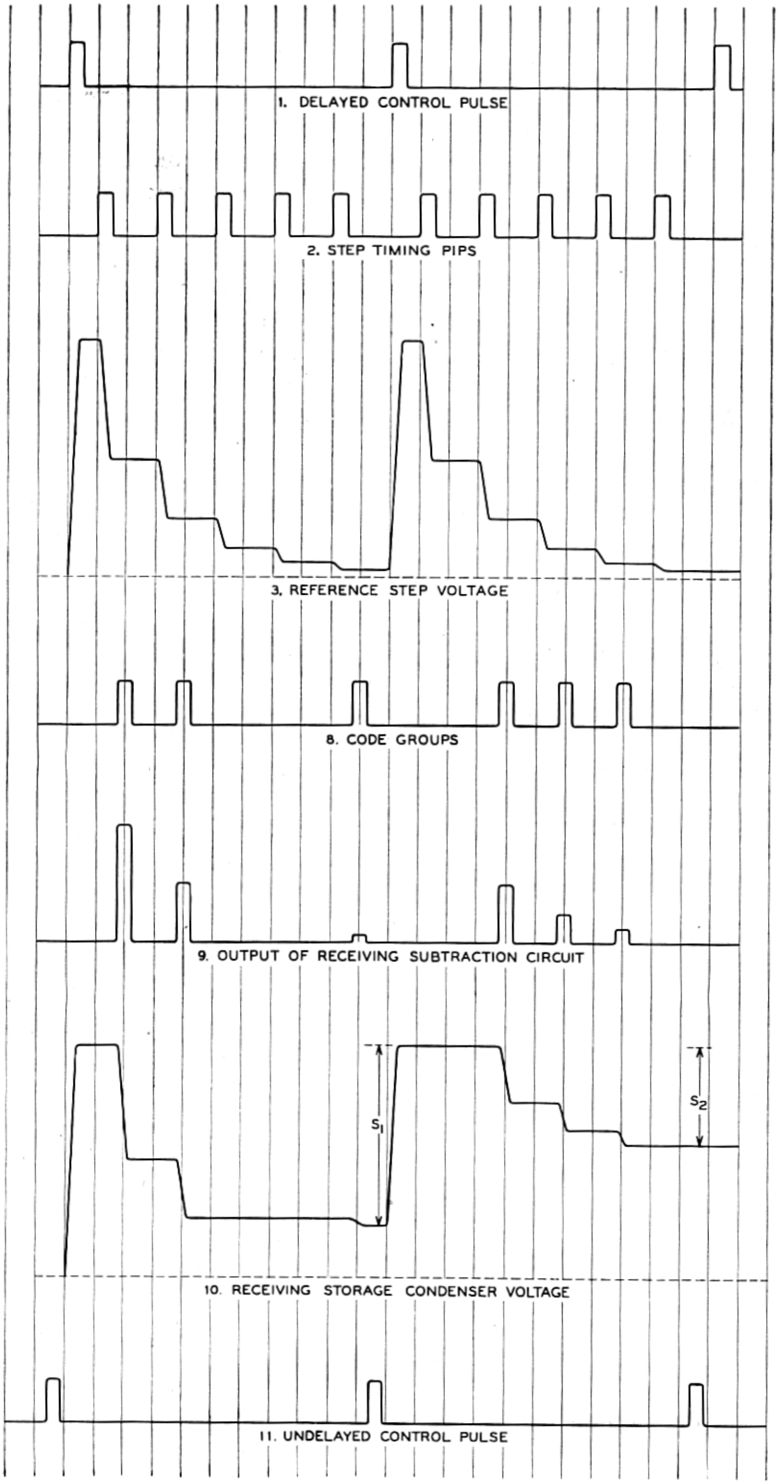

Fig. 7—Detailed wave forms for PCM Receiver (amplitude vs. time).

Referring to Figs. 4 and 5, the "delayed control pulse" Curve 1 is the principal timing pulse for the transmitting coder. It is used to sample the audio wave and to start the step and timing-pip generators. Two sets of timing-pips are produced; one, Curve 2, is used to generate the reference step voltage while the other, Curve 7, is used for timing the digit pulses. The reference step voltage, Curve 3, is used in the comparing circuit and in the subtraction circuit. Curve 4 gives the output of the subtraction circuit, while Curve 5 is the voltage on the storage condenser. The next plot gives Curves 3 and 5 superimposed; the shaded area on this plot corresponds to the time during which a pedestal voltage is generated. The pedestal voltage is given by Curve 6, and the output of the pedestal modulator is given by Curve 8. This last curve is a plot of the two code groups corresponding to the two PAM pulses being coded.

In studying these wave forms it will be noted that the delayed control pulse, the two sets of timing-pips and the reference step voltage curves are the same for each code group. On the other hand the storage condenser voltage, the pedestal voltage, the group of code pulses, and the group of pulses from the subtraction circuit are different for each code group.

It will be recalled that a pedestal voltage is produced during the time that the condenser voltage exceeds the reference step voltage. The leading edge of each pedestal pulse is generated by the falling part of the reference step voltage. The trailing edge of each pedestal pulse is produced by the falling part of the condenser voltage. This drop in condenser voltage is the result of the operation of the subtraction circuit. The output of the subtraction circuit depends upon the delayed digit pulse which has just been passed by the pedestal pulse. Its magnitude depends upon the reference voltage step that applies to the particular digit being transmitted. The function of the delay in the feedback path is to allow the outgoing digit pulse to be completed before the pedestal is terminated.

It is seen that the pedestal voltage contains the same information as the transmitted code groups. Under ideal conditions the use of auxiliary timing pulses would not be required. However, in a practical circuit the leading edge of the pedestal varies, both as to relative timing and as to rate of rise. Under these conditions the auxiliary timing-pips permit accurate timing of the outgoing PCM pulses, as well as constant pulse shape for the input to the subtraction circuit.

Summarizing the foregoing it is seen that in the coder under discussion a comparison is made for each digit between a reference voltage and the voltage across a storage condenser. Initially the voltage across this condenser represents the magnitude of the PAM pulse being coded. After each digit the voltage remaining on the condenser represents the magnitude of the original PAM pulse remaining to be coded. A pedestal voltage is

obtained in the output of the comparing circuit whenever the storage condenser voltage exceeds the reference step voltage.
This pedestal, if present, allows a timing pulse to be sent out as a digit of the code group. This digit pulse is also delayed and fed back to a subtraction circuit which reduces the charge on the condenser by a magnitude corresponding to the digit pulse just transmitted. This process is repeated step by step until the code is completed.

Synchronizing the two control pulse generators, one at the transmitter and one at the receiver, is essential to the proper operation of the equipment. This may be accomplished in a variety of ways. The best method of synchronizing to use would depend upon the application. Although the control could easily be obtained by transmitting a synchronizing pulse over the line, the equipment would have been somewhat more complicated and for these tests a separate channel was used to synchronize the control pulse generators at the terminals.

Having thus established the timing of the receiving control pulse generator shown in Fig. 6 relative to the received code groups, the receiver generates a new set of waves as shown in Fig. 7. Except for delay in the transmission medium, the first three curves are the same as those shown in Fig. 5 for the transmitter. (1) is the delayed control pulse, (2) is the step timing wave, and (3) is the reference step voltage. Curve 8 is the received code group and (9) is the output current of the subtraction circuit. (10) gives the wave form of the voltage across the receiving storage circuit, and (11) gives the curve for the undelayed control pulse.

The receiver functions as follows: The storage condenser is charged to a fixed voltage by each delayed control pulse. The charge on the condenser is reduced by the output of the subtraction circuit. The amount of charge that is subtracted depends upon which digit of the group produces the subtraction pulse. This amount is measured by the reference step voltage. At the end of the code group the voltage remaining on the condenser is sampled by the undelayed control pulse.

It is seen that the storage subtraction circuits in the transmitter and receiver function in similar ways. In the transmitter the original voltage on the condenser depends upon the audio signal, and after the coding process this voltage is substantially zero. The receiver starts with a fixed maximum voltage and after the decoding process the sample that is delivered to the output low-pass filter is given by the voltage reduction of the condenser during the decoding process. Except that the conditions at beginning and end of the coding and decoding periods are different as discussed above, the subtraction process is the same for both units.

The monitoring decoder in the transmitter operates in the same manner described above, except that it employs the various waves already generated for other uses in the transmitter (see Fig. 4).

Experimental Results

An experimental system was set up as shown in Fig. 8. The pulse code modulator, radio transmitter, and antenna comprised the transmitting terminal; while an antenna, radio receiver and pulse code demodulator were used for the receiving terminal. A short air-path separated the terminals. The transmitter used a pulsed magnetron oscillator and the receiver employed a broad-band superheterodyne circuit. The results obtained with this system were similar to those obtained by connecting the pulse code

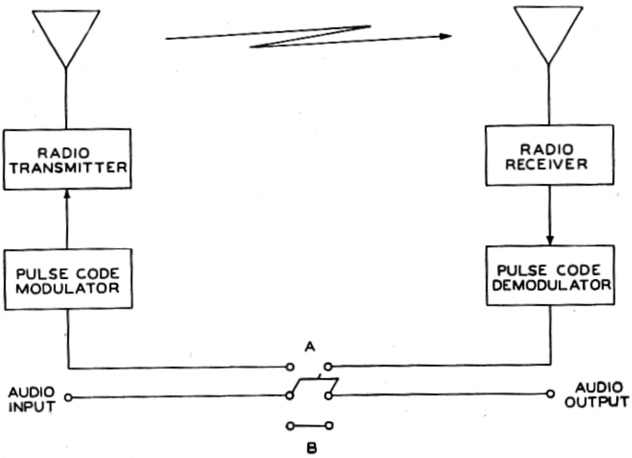

Fig. 8—Block diagram of PCM system.

modulator and demodulator together without the radio equipment. In fact, unless a large amount of attenuation was inserted in the path the presence of the radio circuit could not be detected.

It was possible to adjust the PCM transmitter so that different numbers of digits could be produced. A brief study was made of the number of digits required. It was found that, with regulated volume, a minimum of three or four digits was necessary for good intelligibility for speech though, surprisingly enough, a degree of intelligibility was obtained with a single one. With six digits both speech and music were of good quality when regulated volume was used. Even with six digits, however, it was possible to detect the difference between PCM and direct transmission in A-B tests. This could be done most easily by a comparison of the noise in the two systems. If unregulated volume were used several more digits would probably be desirable for high quality transmission.

In listening to the speech transmitted over the PCM system one obtained the impression that the particular sound patterns of a syllable or a word could be transmitted with three or four digits. If the volume range of the talker varied it would be necessary to add more digits to allow for this variation. Over and above these effects, however, the background noise which is present to a greater or lesser extent in all communication circuits, is quantized by the PCM system. If the size of the quanta or amplitude step is too large the circuit will have a characteristic sound, which can easily be identified. Since the size of the quanta is determined by the number of digits, it is seen that the number of digits required depends not alone upon the speech but also upon the background noise present in the input signal.

Summarizing, experimental results obtained indicate that at least 3 digits are desirable for a minimum grade of circuit and that as many as 6 or more will provide for a good quality circuit. If we wish to transmit a nominal speech band of 4000 cycles, PCM requires the 8000 pulses per second needed by any time-division system, multiplied by the number of digits transmitted. The extra bandwidth required for PCM, however, buys some real advantages including freedom from noise, crosstalk and signal mutilation, and ability to extend the circuit through the use of the regenerative principle.

The writer wishes to acknowledge the assistance of Mr. A. F. Dietrich in the construction and testing of the PCM equipment discussed in this paper.