WS No. 19 Mark III

PULSE CODE MODULATION

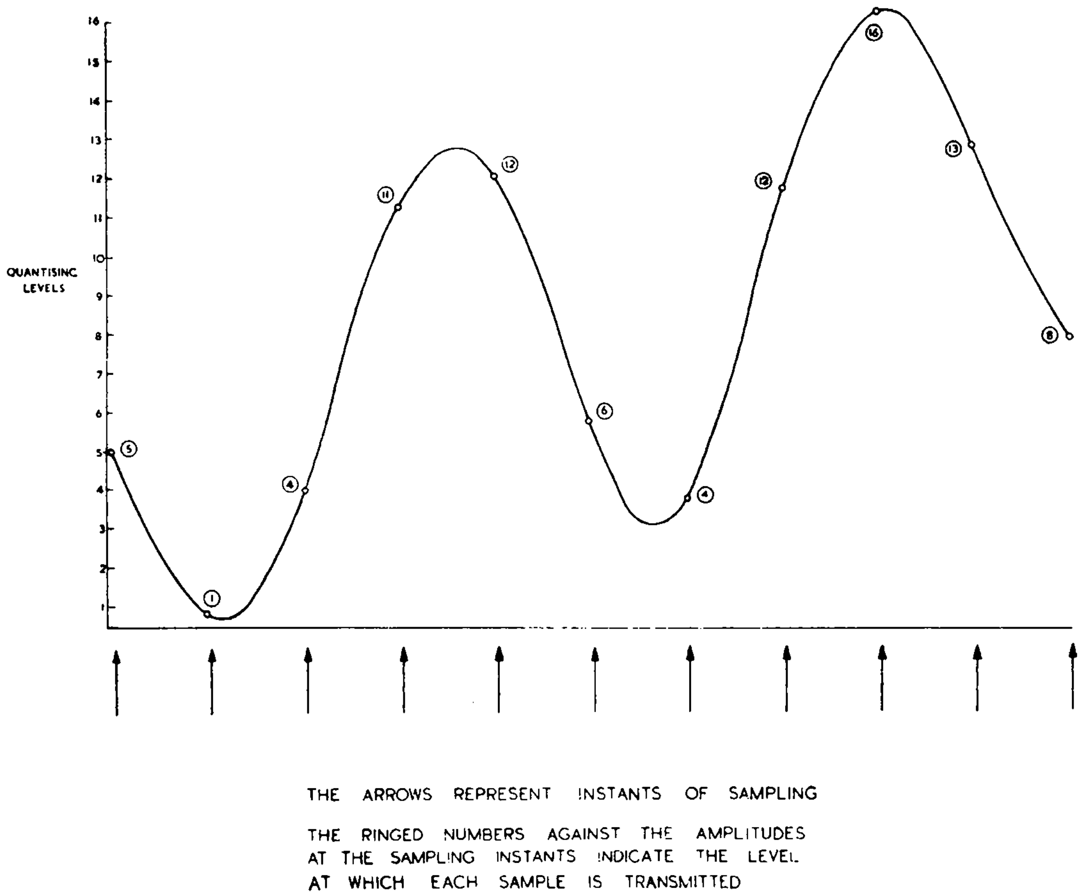
INTRODUCTION

1. In normal pulse modulation systems (e.g. pulse amplitude, pulse width, pulse frequency types) the train of pulses transmitted between the sending and receiving terminals carries the full information of the original modulation waveform. In pulse code modulation the transmitted code pulses are in the nature of control signals, in that they determine the manner in which each sample is reconstructed at the receiver by selecting the appropriate locally-generated signal increments for each pulse position in a code group. It has been established that in order to transmit the essential information contained in a continuously varying complex waveform, such as speech, it is sufficient if the waveform is "sampled" (i.e. the instantaneous amplitude regularly measured and transmitted) at a rate which is at least twice as great as the highest component frequency which it is desired to transmit. For example, in order to transmit the information contained in a speech waveform covering the frequency range up to 3000c/s, the samples must be taken at a rate of at least 6000 per second; in practice a margin would be allowed on this minimum value, and the sampling rate would be chosen to be, say, 6500 or 7000 per second.

2. In the pulse code modulation system, in addition to the speech waveform being sampled at some appropriate sampling rate, the instantaneous amplitude obtained at each sample is also "quantised". This additional process of quantising can best be explained by considering an example.

3. In Figure 1 the full-line curve represents a portion of a signal waveform to be transmitted, and the horizontal solid lines divide up the total range of signal amplitude, between the maximum and minimum values which it can assume, into a number of equal portions (sixteen in the example chosen); the mean amplitude of each portion is referred to as a "level", and the sixteen corresponding levels are shown by dotted lines and numbers in the Figure. Each amplitude sample of the signal waveform will, therefore, fall within one of the sixteen portions, and in the subsequent operations of transmission, instead of identifying the sample by its exact amplitude, it is merely described by the level number corresponding to the portion within which it falls.

4. The nett effect of this process, carried out on each successive sample, is that in order to transmit the original signal waveform it is necessary to pass to the receiver only such information as will specify to which level each sample of signal must be built up; thus in the example chosen, it is only necessary to provide means for indicating any one of sixteen discrete values to the receiver for each sample, instead of requiring to transmit an exact amplitude from an infinite range of values as is done in other methods of pulse modulation. In this respect the system resembles a telegraphic method of signalling, and in fact certain advantages which telegraph methods possess in comparison with normal telephony can be exploited by the use of pulse code modulation.



A-020 1-1

Fig. 1 - Sampling operation

5. It will be apparent that by transmitting the original signal waveform in terms of a finite number of amplitude levels a certain degree of distortion is introduced right at the beginning of the system, since all signal amplitudes falling within a predetermined portion of the total amplitude range are transmitted as being of equal value. By increasing the number of levels this distortion can be reduced as required, but as will be explained in para. 22, only at the expense of providing a circuit (radio or cable link) of greater bandwidth between the sender modulator and the receiver demodulator. For any particular application it is therefore necessary to determine the minimum number of levels which will give a satisfactory reproduction of the original signal waveform at the receiver.

6. It has been established that speech of good telephone quality can be transmitted on a system using thirty-two levels; a satisfactory reproduction of music would require a system using upwards of one hundred levels.

METHOD OF PULSE CODE TRANSMISSION

7. As an example, it will be taken that sufficient accuracy of transmission of a specified signal can be obtained if its total range of amplitude is divided into

sixteen levels. The transmission problem is then that of indicating to the receiver without ambiguity, to which of the sixteen possible levels it has to adjust the output on receipt of each successive sample. In a very simple case we may assume that a group of pulses, up to a maximum of sixteen, can be transmitted for each sample taken by the modulator; in order to indicate the N^{th} level the first N pulses of the group would be transmitted and the remainder suppressed. At the receiver a device would be required to count the number of pulses received in each group in order to determine to which level the output amplitude should be raised for each sample.

8. Such a simple system as described above would be wasteful of pulses, and it is possible to devise more economical methods of transmission. The binary scale can be employed, resulting in effective indication of the quantised signal level of any sample, demanding the use of only a comparatively small group of pulses, transmitted or suppressed as required.

9. For example, any number between 0 and 15 inclusive (16 in all) can be represented by four digits, each a choice of 0 or 1. Each digit, from right to left will represent the multiplying factor associated with 2^0, 2^1, 2^2, 2^3, etc. such that the total sum indicates the number.

For example, 14 is represented by 1110 in the scale of two

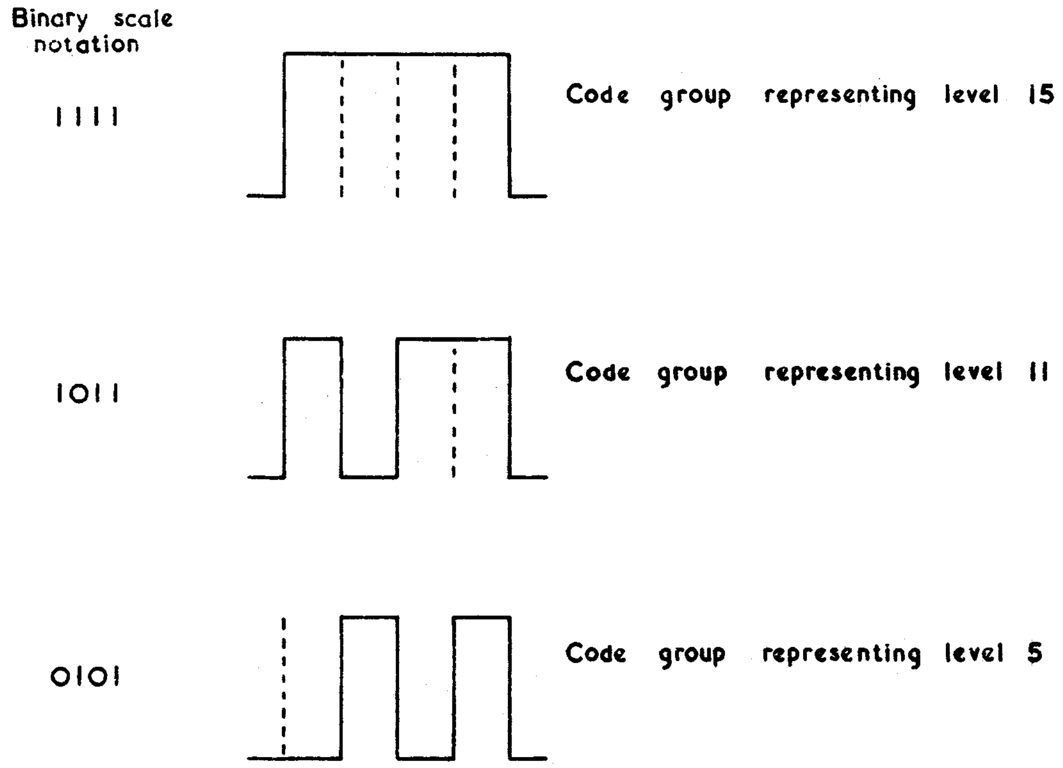
$$\equiv 1 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$$

Further examples are given in Fig. 2.

10. Applying this notation to the transmission problem, the figures 1 and 0 can be associated with transmission and suppression respectively of a pulse. In the succeeding text, the digits 1 and 0 will so be taken to represent the transmission or suppression of an associated pulse. Thus all level numbers between 0 and 15 inclusive can be transmitted by the various combinations of four pulses either present or suppressed.

11. So far the number of different levels to be transmitted has been chosen as sixteen, this requiring four pulses in each code group with each pulse being capable of two conditions, viz. present or suppressed. In general, if there are X pulses in each group, each capable of assuming one of two possible conditions, the number of levels which can be transmitted is 2^X, this being the maximum number of unique combinations in the binary scale obtainable with X digits.

12. It is not essential that each pulse should be limited to two possible conditions; it would be quite practicable for each pulse to be capable of assuming, say, three possible conditions differing in either amplitude or width, or both, in which case the number of levels which could be transmitted with X pulses per group would be 3^X. (The digital analogy in this case is the ternary scale). However, the system in which each pulse in a group has only two conditions, present or suppressed (i.e. mark or space), has many advantages. The detector at the receiving station will be dealing with groups of pulses which may be of distorted shape and of amplitude comparable with the noise level, but provided that it can definitely detect whether the successive pulses in each group are present or suppressed, no matter of what relative amplitude or width they may be, it will have its duty (and hence the possibility of error) reduced to the very minimum, since a new and clean-cut train of pulses can be regenerated from the information given by such a detector.



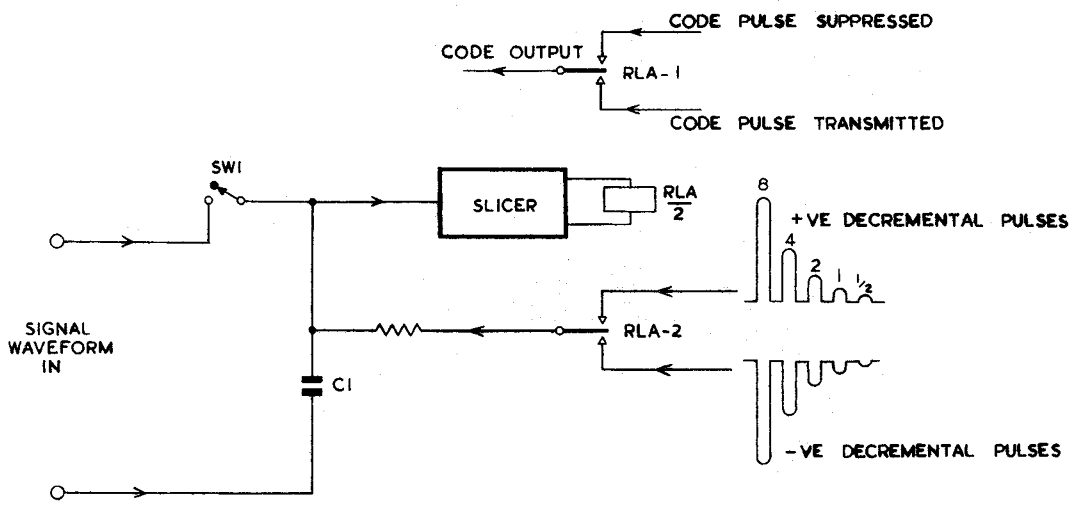
A-020
1-2

Fig. 2 - Binary code pulse

13. This regenerated pulse train can then be passed on to the circuits which adjust the signal output to the correct level for each received code group. Similar advantages are gained when a chain of repeater stations operating between two terminals is considered; provided that the receiving detector at each repeater can distinguish the presence or absence of each pulse, then regenerating circuits can supply a completely new, clean-cut, and noise-free train of pulses for onward transmission to the next repeater, and which will contain all the information carried by the original pulse train without any added distortion or noise. It is in this respect that telegraphic signalling methods can be exploited with success by means of pulse code modulation.

DESCRIPTION OF A SIMPLE PULSE CODE MODULATION SYSTEM

14. The problem to be solved by the pulse code modulator is the sampling of the signal waveform at a suitable rate and the translation of successive samples into the appropriate code groups. Many methods of achieving this have been proposed and investigated, and also incorporated into successful operational systems; the particular method described below has the advantage of being carried out with standard electronic valves and components, and of being economical in their use. It is assumed that the system is required to transmit a voltage lying between +16 and -16 volts on thirty-two levels, with two-state pulses in the code groups. Five pulse positions per group will thus be required.



A-020
1-3

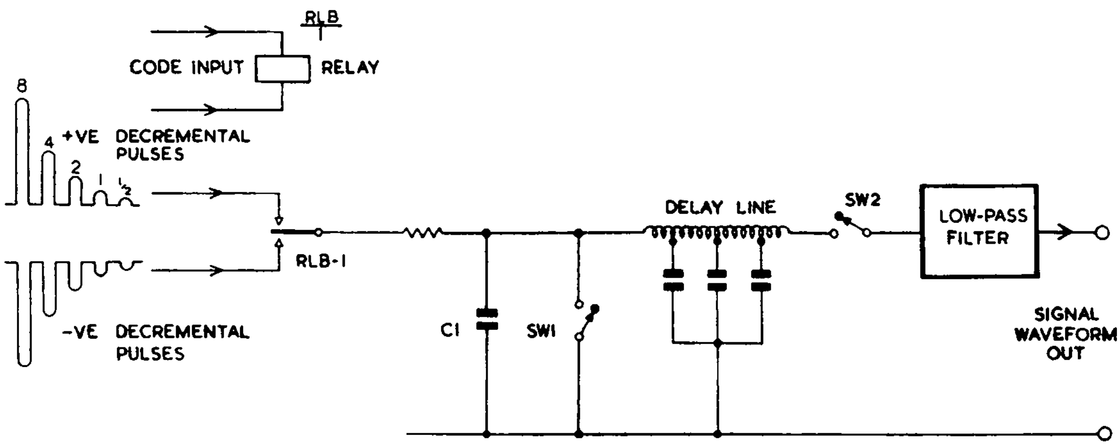
Fig. 3 - Basic coding circuit

15. A simplified block diagram of the proposed method is shown in Fig. 3, where mechanical devices are shown as replacing the high-speed electronic switches which would normally be used, in order to make the method of operation more obvious. The "slicer" is a high-gain amplifier which is biased to the centre of the linear portion of its characteristic with zero input signal; a small change of input in either direction is sufficient to cause the centre-stable relay at the output to close to one side or the other. The "slicer" is operated by the voltage built up on the storage capacitor C1, and momentary closure of switch SW1 at the sampling rate allows a voltage to be stored on C1 which is equal to the instantaneous sampled value of the signal waveform; between sampling instants the switch remains open. Two similar series of decremental pulses, one positive-going and the other negative-going, are generated immediately after each operation of SW1, and are applied to the fixed contacts of the output relay; the frequency of repetition of the series will be that of the sampling rate; in the present system each series contains five pulses, and each pulse is exactly one-half of the amplitude of the preceding one, the first pulse being equal to one-quarter of the total amplitude range of the input signal waveform. Thus the amplitudes of the five pulses correspond to 8, 4, 2, 1 and $\frac{1}{2}$ levels respectively. The "slicer" circuit is permitted to examine the sign of the voltage on condenser C1 five separate times, thus operating the relay five times and generating the five code pulses. The "slicer" circuit is so arranged that a positive signal on C1 causes the output relay to close on to the side connected to the negative-going decremental pulses, at the same time transmitting an output pulse, while a negative signal feeds back a positive-going decremental pulse and suppresses the output pulse. The selected pulses are fed back via the centre contact of the relay to the storage capacitor C1 immediately after the operation of the relay, and so cause a change in voltage level on this capacitor ready for the next examination. In actual practice no examination is made after the fifth pulse is fed back; instead, a new sample is impressed on capacitor C1 and the process repeated. The two series of decremental pulse trains can be formed quite simply by impulsing a tuned circuit and allowing it to ring with the appropriate decrement so that each positive half cycle is one-half the amplitude of

the previous one. This waveform is then rectified and applied to both cathode and anode followers to produce a pair of low impedance paraphrased outputs.

16. To make the operation described in para. 15 more obvious an example is given as follows: assume the incoming level to be sampled and coded is +3.2; this level lies within the 19th quantised level if the range -16 to +16 is divided up in 32 quantised levels from 0-31 inclusive. The binary number equivalent to 19 is 10011. Now the circuit of Fig. 3 charges condenser C1 to the value +3.2, and the slicer circuit will then transmit a pulse (this will be digit 1 from para. 10) and feed back a negative pulse of amplitude 8. The resultant amplitude (-4.8) will cause the slicer circuit to suppress the next pulse (digit 0), and feed back a positive pulse of amplitude 4. The resultant amplitude (-0.8) will result in another suppressed pulse (digit 0), and a positive feedback of +2 will give a resultant amplitude of +1.2. The slicer circuit reacts to give a transmitted output pulse (digit 1), and a negative feedback of amplitude 1. Adding +1.2 to -1, we have a positive sum, and so that final digit transmitted to complete the definition of the particular sample under consideration is digit 1. The stage-by-stage formation of digits 10011 is thus completed. A precise interpretation of the above operations as a binary scale digit formation is given in para. 29. It will be appreciated after study of para. 29, that the circuit takes the incoming signal, quantises it and forms a binary scale 5 digit code in one operation. It is important to emphasise that the binary scale code so formed ranges between 0 and 32; the actual originating quantised level will be less than the binary scale code by the quantity 16.

17. Thus the block diagram of Figure 3 represents a basic circuit which carries out the required quantising process to the nearest level, and which also produces the required groups of code pulses for transmission; its mode of operation is that of a servo-mechanism which, on the injection of a disturbing signal, always tries to reset itself to its zero condition in a series of predetermined steps of logarithmically decreasing amplitude.



A-020 1-4

Fig. 4 - Basic de-coding circuit

18. A simplified block diagram of the demodulator is shown in Fig. 4. Two series of decremental pulses, one positive-going and the other negative-going, exactly like those used in the modulator, are generated locally and can be selected by the input

relay which is controlled by the incoming mark/space code groups; for each group a voltage is built up on the capacitor C1, the code pulses operating the relay so that an increment of correct sign is added according to the presence or absence of each pulse. At the end of each code group the switch SW2 closes momentarily to pass the final sample to the output, and switch SW1 also closes to discharge the capacitor ready for building the next sample from zero. The delay line is inserted so that as the signal voltage is built up it is stored in the delay line and simultaneous operation of switches SW1 and SW2 does then not result in any loss of signal level. The low-pass filter "smooths out" the sudden transition of the output signal from one level to another as successive samples are received.

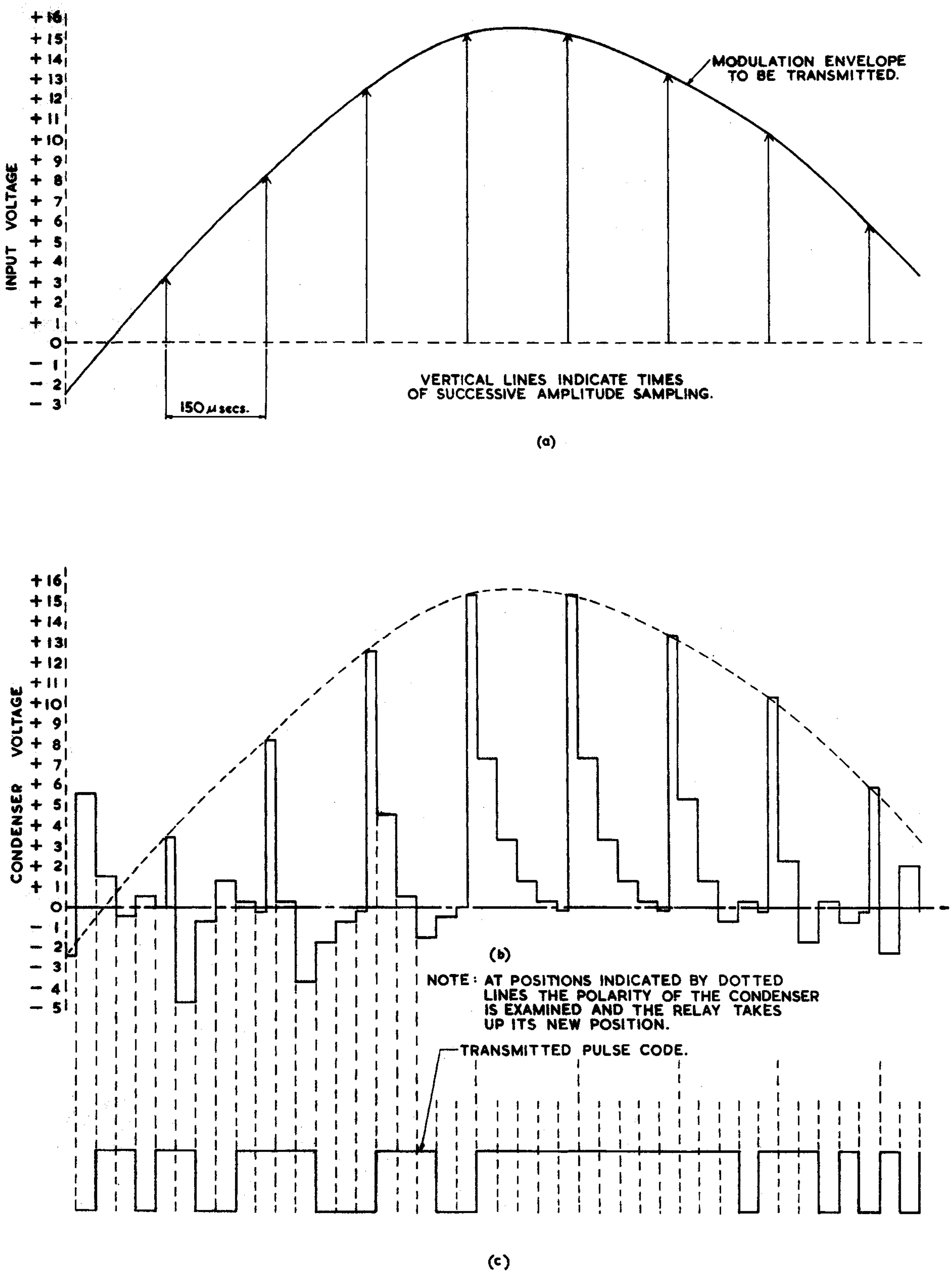
19. The interpretation of the above operations as a reconstitution of the original signal from a five digit scale of two code, bearing in mind the degree of inaccuracy mentioned in para. 5, is given in para. 30.

20. Figure 5 represents the above process carried out in the modulator on a number of samples of a signal waveform to be transmitted. Fig. 5(a) shows the amplitude samples as arrowed vertical lines; Fig. 5(b) shows the changes in capacitor voltage as the five selected positive or negative decremental pulses are applied to the capacitor via the feedback path. Fig. 5(c) shows the corresponding groups of code pulses, and it will be noted that they form a continuous train without any gap between consecutive pulses or groups of pulses.

21. Figure 6 illustrates the action of the demodulator in dealing with the received and regenerated pulse train. For each group of five pulse positions, the incremental voltage built up on the capacitor will follow the waveform of the Fig. 6(b) diagram, starting from zero level in each case, with the large dot representing the final signal level reached. In Fig. 5(c) the successive final signal levels, as passed out by the delay line, are again shown, together with the smoothed output from the low-pass filter; the resultant waveform then represents the received version of the original signal waveform of Figure 5(a).

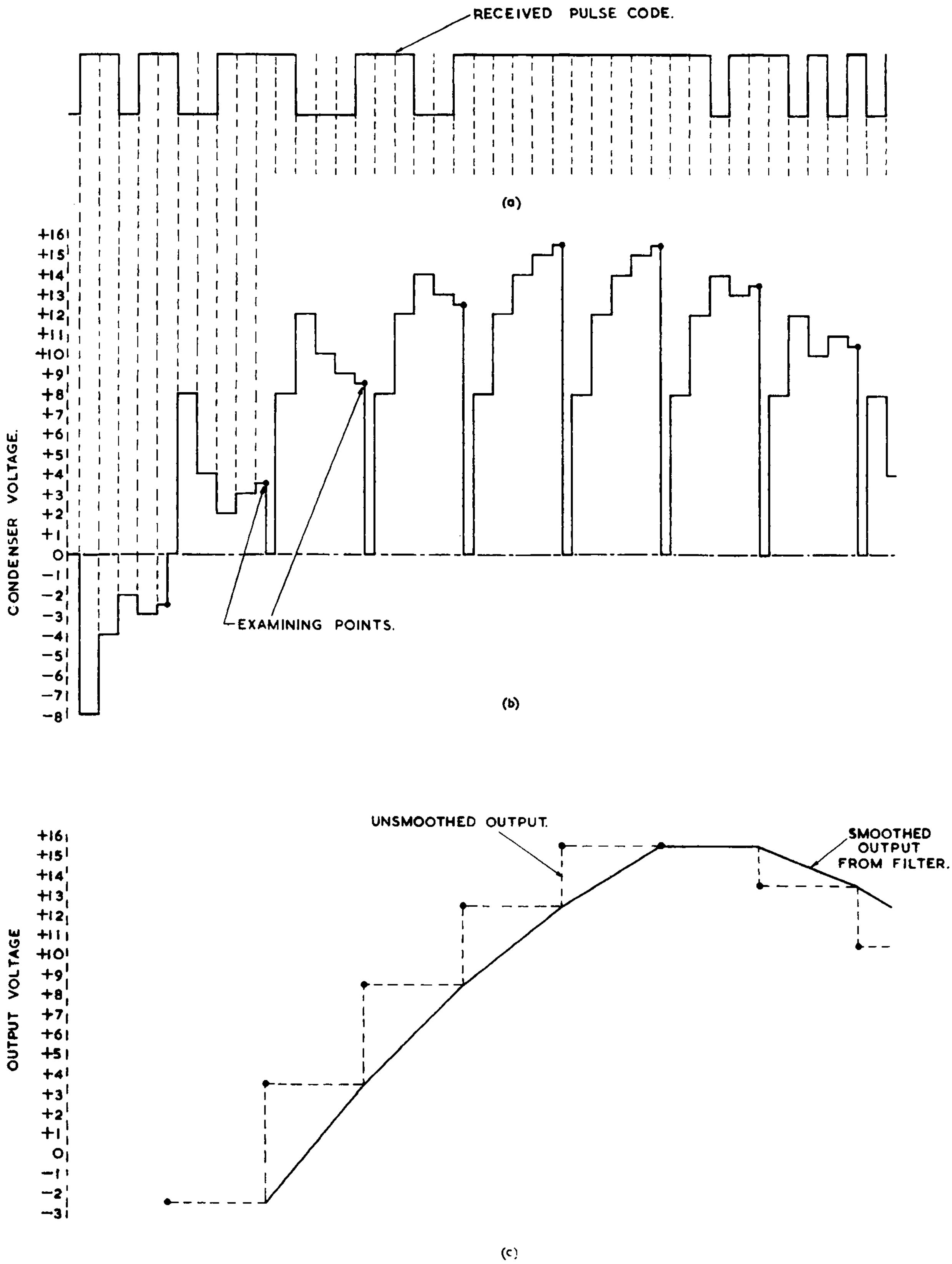
BANDWIDTH REQUIREMENTS FOR THE TRANSMISSION LINK

22. In para. 1 it was stated that the sampling rate must be at least twice as great as the highest component frequency it is desired to transmit; furthermore, to transmit speech of good telephone quality, at least thirty-two quantising levels are necessary. Assuming speech modulation to lie in the frequency band up to 3000c/s, this implies a sampling rate of at least 6000 samples per second: for each sample five code pulses have to be transmitted, so that the minimum basic rate of pulse transmission between sender and receiver is 30,000 pulses per second (i.e. a baud speed of 30,000). Such a rate of transmission requires a minimum bandwidth in the transmission link of $\pm$15kc/s (assuming double sideband methods). This is based on the necessity to transmit the equivalent fundamental frequency of a baud speed of 30,000, and in practice a bandwidth of up to twice this minimum value would normally be specified to allow a margin of safety against further pulse distortion. Thus the use of pulse code modulation for speech transmission requires a channel of between five and ten times the bandwidth which would be required if the speech were transmitted by normal amplitude modulation methods; this increase in bandwidth is the price to be paid for providing a system in which distortion never exceeds a predetermined amount, and the effects of noise in the transmission link can be entirely eliminated, irrespective of the overall distance between terminals of the system, if suitably spaced repeaters be provided for long-distance circuits.



A-020
1-5

Fig. 5 - Code formation from typical waveform



A-020
1-6

Fig. 6 - Reconstitution of original waveform

On account of this bandwidth requirement, the application of pulse code modulation is likely to be confined to v.h.f., u.h.f., and centimetric radio equipments. It is a particularly useful method of modulation for multi-channel radio telephony, operating on a time-division multiplexing basis, since such circuits will generally be required to provide a high order of intelligibility and reliability; in such applications the same modulator can be used in time sequence for each of the separate audio channels, provided that it is designed to operate at the higher pulse repetition rate so required, and this can lead to an economical use of equipment in relation to the facilities provided.

SYNCHRONISATION

23. From the description given in paras. 14-21 it will be clear that it is necessary to provide two forms of synchronisation between the sending and receiving equipment of a pulse code modulator system. Firstly, the basic pulse repetition rate (or digit frequency), which was 30,000 bauds in the example given in para. 22, must be recovered in order to generate correctly-timed pulses for regenerating and demodulating circuits; secondly, group synchronisation must be provided so that the first of the series of decremental pulses generated at the receiver will always coincide with the first pulse in each incoming code group.

24. Various methods of achieving each form of synchronisation have been devised. Digit synchronism can be obtained by differentiating the changeovers of the received pulse train, and using the short-duration pulses so obtained to lock a local oscillator operating at approximately the correct digit frequency.

25. Group synchronism requires that some special signal be impressed at regular intervals upon the transmitted pulse train, which can be recognised and extracted at the receiver and used to control the timing of the decremental pulses and the operation of the electronic switches. In one practical system this special signal consists of ten consecutive marks in the pulse train, occupying two complete code pulse groups injected once every five seconds; the stability of the local circuits is sufficient to maintain synchronism during these periods. Since such combination of marks could occur at intervals as a normal output from the modulator, means are provided for suppressing any such output and for impressing the special signal on the transmitted pulse train only when required for synchronisation purposes. A device at the receiver is responsive only to the special signal, and when this is received a locking pulse is generated to bring into step the circuits generating the decremental pulses. In the Wireless set D.70 the special signal consists of 60 alternative on/off pulses transmitted 4 times a second occupying 12 successive pulse groups in this 12 channel system.

USE OF A COMPANDOR WITH PULSE CODE MODULATION

26. The method used for code modulation described in paras. 14-21 has an intrinsic limitation in that the quantising is linear. For a radio telephone system, which may be used at many different speaking levels, it is essential to ensure that the fullest use is made of the limited amplitude range of the modulator when handling low-level signals, while at the same time preserving a reasonably high overload limit for high-level signals. To achieve this it is common practice to include some form of compandor in the system, in an attempt to approach logarithmic quantisation.

27. The difference between linear and logarithmic quantisation is brought out in Fig. 1. While linear quantisation will transmit a small % change in the larger amplitude region, it will completely ignore such a % change in the smaller amplitude region. Briefly, the comandor consists of a non-linear amplifier inserted before the modulator, and having a characteristic giving greater amplification to low-amplitude signals than to large-amplitude ones; this characteristic usually approximates to a logarithmic law. At the receiver an amplifier having an inverse characteristic is included after the demodulator. The overall amplitude response of the system thus remains linear, but low-level signals are amplified sufficiently to ensure the use of a reasonably large number of the available quantising levels and so avoid undue distortion.

28. In a multi-channel time-multiplexed pulse code modulation system the compandor amplifiers would be common to all channels, that at the transmitter being situated between the channel-switching equipment and the modulator, and that at the receiver being immediately after the demodulator. In all cases it is necessary that these non-linear amplifiers should be of the "instantaneous response" type, i.e. they should have a time-constant of operation which is small compared with the interval between successive sampling operations in the pulse code modulator.

BINARY ANALYSIS OF CODING OPERATION

29. Let any number between +16 and -16 be represented by

$$A.2^4 + B.2^3 + C.2^2 + D.2^1 + E.2^0 + F - 2^4 \qquad \dots (1)$$

A, B, C, D and E are either 0 or +1
and F is a +ve fraction

(The quantity 2^4 is inserted in the expression to keep A, B, C, D and E +ve)

Examine the sign of the expression (1)

If +ve A must be +1
If -ve A must be 0

(This can be seen by grouping the last six quantities together - whatever the value of B, C, D, E, F within the restrictions specified above this group must always be -ve)

Record the value of A

Add to expression (1) $2^3.(1 - 2A)$

If then becomes $A.2^4 + B.2^3 + C.2^2 + D.2^1 + E.2^0 + F - 2^4 + 2^3 - A.2^4$

$$= B.2^3 + C.2^2 + D.2^1 + E.2^0 + F - 2^3 \qquad \dots (2)$$

Examine the sign of expression (2)

If +ve B must be +1

If -ve B must be 0 with similar reasoning to the first instance

Record the value of B

Add to expression (2) $2^2.(1 - 2B)$

It then becomes $C.2^2 + D.2^1 + E.2^0 + F - 2^2$ ... (3)

In like manner, C, D and E can be recorded.

F can be neglected if a "quantised" result is required, i.e. a single number representing all numbers between adjacent whole numbers.

The electrical circuit described in para. 15 performs the electrical equivalent of the above process - it examines the expression for sign, records and transmits the appropriate pulse and then adds to the signal a voltage represented by $2^3.(1 - 2A)$, $2^2.(1 - 2B)$, etc. in succession until five pulses (A, B, C, D, E) have been transmitted.

The equipment actually subtracts or adds 8, 4, 2, 1 in succession, depending on the sign of the previous expression (+ve or -ve respectively), but it will be seen below that this is an identical operation to that detailed in the mathematical evaluation,

i.e. $-8 \equiv 2^3.(1 - 2A)$ when $A = +1$

$+8 \equiv 2^3.(1 - 2A)$ when $A = 0$

$-4 \equiv 2^2.(1 - 2B)$ when $B = +1$

$+4 \equiv 2^2.(1 - 2B)$ when $B = 0$

$-2 \equiv 2^1.(1 - 2C)$ when $C = +1$

$+2 \equiv 2^1.(1 - 2C)$ when $C = 0$

$-1 \equiv 2^0.(1 - 2D)$ when $D = +1$

$+1 \equiv 2^0.(1 - 2D)$ when $D = 0$

BINARY ANALYSIS OF DECODING OPERATION

30. The transmitted code consists of five pulses representing the five digits A, B, C, D and E. The requirement is to reproduce the originating number, bearing in mind that the original quantising will restrict the accuracy with which the original number can be reconstituted. As all numbers between two adjacent whole numbers were transmitted by one level (the lower of the two numbers concerned), a fair average is achieved by adding the quantity $\frac{1}{2}$ to the quantised level.

Thus the reconstituted number required is

$$A.2^4 + B.2^3 + C.2^2 + D.2^1 + E.2^0 + \frac{1}{2} - 2^4$$

This equals

$$A.2^4 + B.2^3 + C.2^2 + D.2^1 + E.2^0 + \frac{1}{2} - (2^3 + 2^2 + 2^1 + 2^0 + 2^{-1} + \frac{1}{2})$$

$$= (A2^4 - 2^3) + (B2^3 - 2^2) + (C2^2 - 2^1) + (D2^1 - 2^0) + (E2^0 - 2^{-1})$$

$$= 2^3(2A - 1) + 2^2(2B - 1) + 2^1(2C - 1) + 2^0(2D - 1) + 2^{-1}(2E - 1)$$

$2^3(2A - 1) \equiv \pm 8$ depending on the value of A (0 or +1)

Likewise $2^2(2B - 1) \equiv \pm 4$

$2^1(2C - 1) \equiv \pm 2$

$2^0(2D - 1) \equiv \pm 1$

$2^{-1}(2D - 1) \equiv \frac{\pm 1}{2}$

Thus the requirement resolves itself into adding these five quantities together with the sign of each dependent on the value of the received digit.

It will be seen by reference to para. 18 that the equipment performs this addition under control of the value of the received digits.

END