From HISTORY OF TECHNOLOGY, ed. T. I. Williams,
Oxford, 1978, Vol. 7, pp. 1220–1267

50

ELECTRICAL COMMUNICATION

D. G. TUCKER

I. TELEGRAPHY

THE nineteenth-century developments in telegraphy have been previously described in this work (Vol. IV, Ch. 22; Vol. V, Ch. 10) and there is not a great deal to add now. By the end of the nineteenth century there was an efficient inland telegraph system in all developed countries, together with a world-wide network of submarine cables and other international lines. The only subsequent developments we have space to mention here, other than radio which is dealt with separately, are the general adoption of the teleprinter (a telegraphic typewriter) in place of other telegraph instruments; the spread during the last two decades before 1950, and even more since then, of the telex service whereby teleprinters operate over ordinary local telephone lines and through an international system of switching centres; and the general adoption of carrier transmission whereby telegraph signals are modulated on to carrier tones for transmission over telephone channels.

In the following account of the history of electrical communication we shall have occasion to refer to certain basic devices—such as valves, capacitors, resistors, and inductors—which are normal components of the circuits involved. The history of the development of these devices has already been discussed (Ch. 46).

II. THE BEGINNINGS OF TELEPHONY (1876–*c.* 1915)

Origins. The invention of the telephone is generally credited to the Scottish-born inventor Alexander Graham Bell who emigrated to the U.S.A. in 1871. It was on 10 March 1876 that he first obtained good clear articulation from his experimental system. There had been much earlier attempts at electric telephony but they were hardly successful. More importantly, there was an almost simultaneous rival inventor, Elisha Gray (U.S.A. [1]), who filed a caveat in the U.S. Patent Office on his design of a telephone system on the same day (but later in the day) that Bell had filed a patent application for his system. Both Bell and Gray had been working on harmonic telegraphs which could transmit and respond to a number of simultaneous tones, and an extension of this idea to the transmission of the human voice was a logical step. Gray did not think telephony was important and did little more work on it; Bell followed it up for a while and his sponsors made a commercial success of it. Bell was in advance of Gray experimentally, but used at first an electromagnetic transmitter, rather similar to his receiver, the general principle of which still forms the basis of modern telephone receivers. This transmitter caused a piece of iron to vibrate under the influence of the sound waves in the field of a magnet with coils on it, and thus to induce voltages in the coils and corresponding currents in the line extending from them to the receiver. It was Gray, however, who invented the variable-resistance transmitter, in which the voice waves, through the medium of a wire attached to the diaphragm and dipping into a conducting solution, varied the resistance and thus the current (established by a battery) in the line and receiver. This was (and is) a far more powerful method and, in the form of a cell of carbon granules where the pressure of the diaphragm varies the resistance through the cell, is the basis of the modern telephone transmitter. Here the power is provided by the current obtained from a battery, not from the voice, which merely modulates the current.

Bell was undoubtedly aware of the importance of having in his instruments a relatively strong unidirectional magnetic field (produced either by a battery giving a standing current in the coils, or by the use of a permanent magnet as the core), but the fundamental nature of the requirement appears to have been first set out by Oliver Heaviside (Britain), as early as February 1887 [2]. He showed that if the permanent field in a receiver is H, and this varies from $H+h$ to $H-h$ under the influence of the speech currents, then the force on the diaphragm, which is proportional to th square of the field, varies from $(H+h)^2$ to $(H-h)^2$, a variation of $4Hh$, ignoring the constant of proportionality. That is to say, the variation is proportional to the permanent field H provided H is larger than h. If there is no permanent field, then $H = 0$ and the force on the diaphragm does not reproduce correctly the current waveform.

Late in 1876 Bell offered his patent to the Western Union Telegraph Company for \$100 000. Western Union then saw no future in the telephone and refused the offer. Bell continued his developments with support from the company he himself founded, and a year later the obvious success of the telephone showed Western Union the greatness of their error of judgement. They entered the telephone field on the basis of Gray's patent (Gray was

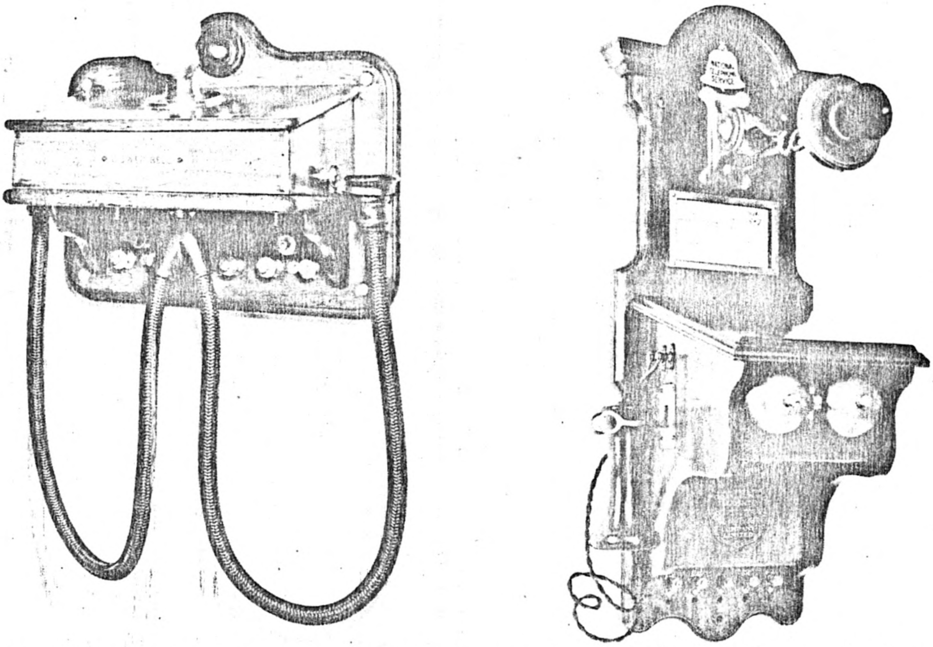

FIG. 50.1(a). The Gower–Bell telephone, 1880–2.

(b). National Telephone Company's telephone instrument, type No. 1 for Central Battery Exchanges, *c.* 1910.

associated with them through their subsidiary Western Electric Manufacturing Company, which Gray had virtually founded) and with commissioned development by Thomas A. Edison (U.S.A.). Edison developed an improved variable-resistance transmitter in which the speech waves acted upon a loose contact between pieces of carbon, and a rather unsatisfactory receiver in which the speech current produced a varying drag on a stylus conducting it into a rotating chalk cylinder moistened with potassium iodide; this varying drag produced vibrations in a diaphragm coupled to the stylus. Western Union went ahead commercially, and fought the Bell Company over patents, but eventually sold out to Bell in return for 20 per cent royalties for the duration of the patents.

The beginning of telephone service [3]. Commercial development of the telephone began quite soon after the initial inventions of Bell and Gray in 1876. Bell's company started in 1877, changing its form rapidly at first, and became well established in the U.S.A., with subsidiary companies in Britain and elsewhere, by 1879. Independent telephone companies which Edison had set up in London and elsewhere amalgamated within a further year or two with the

Bell or other suitable companies in their locality. Nevertheless, there were numerous inventions of new or modified forms of telephone transmitter and receiver, and numerous companies were set up to exploit them (Fig. 50.1).

The concept of the telephone providing a public service on a basis rather different from that provided by the telegraph sprang up quickly. Individual telephone renters, or 'subscribers', would require to be connected to another subscriber on demand, and to provide this service telephone 'exchanges' were introduced; by Bell from 1878 and by Edison from 1879. At first, of course, the service was entirely local, but as exchanges became more numerous, the need for interconnecting them became apparent. Thus an inter-urban telephone network started to grow; its growth and usefulness were considerably restricted by the fact that in many areas each exchange, or small group of exchanges, belonged to a separate company using different technical and commercial methods. Since also many governments actively discouraged the development of inter-urban links from fear of competition with state-owned telegraphs, it is not surprising that it was only in the U.S.A. that a significant telephone network had developed by 1884. Statistics published in 1885 showed the U.S.A. to have 140 000 subscribers with 800 exchanges, while for the rest of the world the list was topped by Britain with a mere 10 000 telephones; inter-urban lines in the U.S.A. numbered over 800, while there were only about 80 in Britain. It is probable that the U.S.A. had more than twice as many telephones and inter-urban lines as the rest of the world put together. In spite of disparities in telephone densities in different countries, proportional growth rates represented by a doubling of numbers and wire mileages every two or three years were maintained almost everywhere.

Telephone exchanges. The idea of connecting one subscriber, on request, to another by means of a flexible switching system had, on a very small scale, been applied to local telegraph subscribers in the U.S.A. and Britain before 1877. During 1877 one or two very small private telephone exchanges had been tried in the U.S.A., but the first commercial telephone exchange was that at New Haven, Connecticut, opened on 28 January 1878. In this, eight lines (with 21 telephones connected) could be connected in pairs by means of a direct cord with a plug at each end; calling 'annunciators' were provided, operated by d.c. from the subscriber's battery. A 20-line exchange was fitted six months later at Bridgeport, Connecticut. During the same year, the American District Telegraph Company opened an Edison exchange in Chicago, also using direct interconnection by a single cord. Calling for sub-

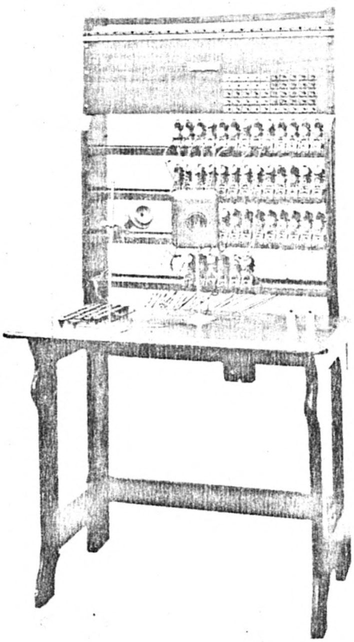

FIG. 50.2. The Jones telephone switchboard, 1879, made by C. E. Jones & Brother, Electricians, Cincinnati, Ohio, U.S.A.

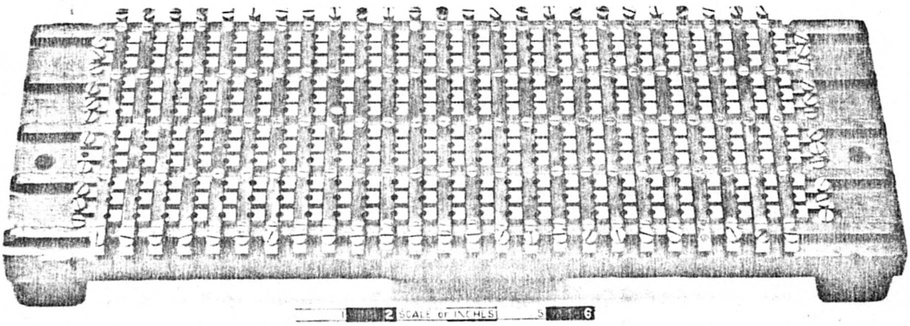

FIG. 50.3. The Edison crossbar telephone exchange unit, 1879.

scribers who already had a telegraph instrument was by telegraph; for others, calling relays were used. This switchboard grew rapidly and several boards were needed, with interboard connections and two operators involved in most calls. It was here that the 'jack-knife' switch, forerunner of the well-known 'jack', was introduced by C. E. Scribner, engineer of the Western Electric Manufacturing Company.

The year 1879 might well be regarded as the year of the telephone exchange, for it saw the introduction of double-cord interconnection via intermediate connecting bars which could run the length of a whole suite of 25-line boards; of the 'multiple' whereby outgoing access to subscribers' lines was repeated throughout the suite so that each operator had direct access to every subscriber; of several line-engaged testing arrangements; and of the first exchange in Britain. This last was opened by the (Bell) Telephone Company in August 1879 at Coleman Street in London, and had calling indicators, jacks, cords, and connecting bars.

It could fairly be said that the manual telephone exchange of the form which became so ubiquitous in the first half of the present century arose directly out of the developments just described. Many technical and operating improvements were introduced, of course, such as making the engaged signal a click when the operator touched the tip of her calling plug on the outer ring of the jack, the use of small lamps as calling and supervisory signals, etc. But this type of switchboard was not the only one. The National Bell Telephone Company in the U.S.A. in 1880 introduced a board using an array of horizontal and vertical bars—one of the latter for each line—so that interconnections could be made by inserting plugs at each crossing of any one horizontal bar with the two lines to be connected together. In Britain, the Edison Telephone Company used a similar arrangement (Fig. 50.3). It was, however, not suitable for expansion by the 'multiple' method (Fig. 50.4), and so died out.

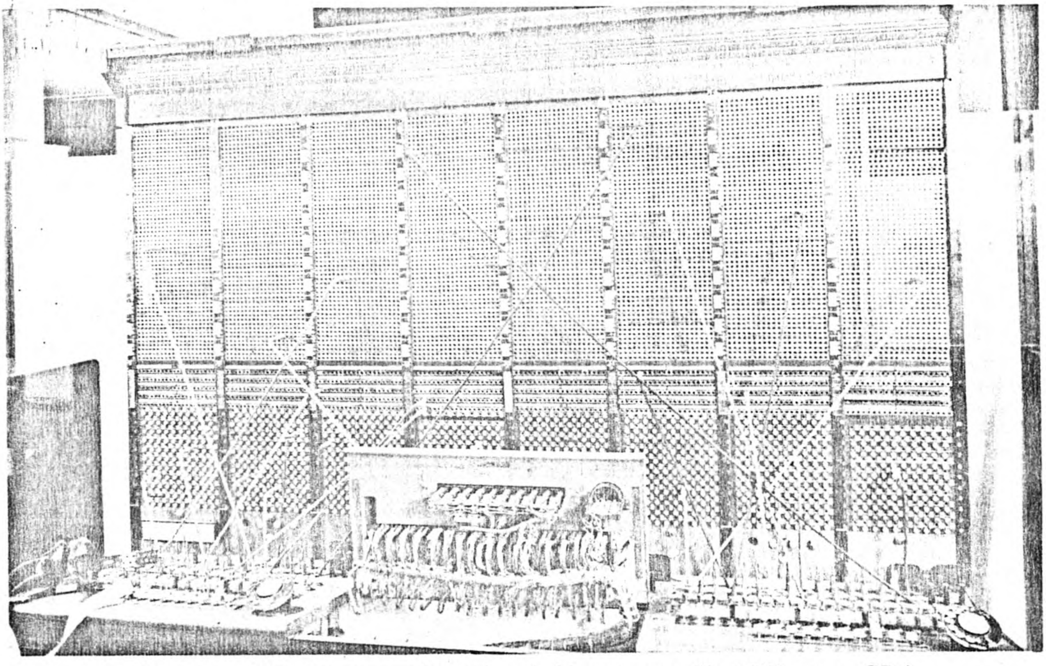

FIG. 50.4. A multiple-type telephone switchboard, 1925 (Post Office type CB1).

The reliance on batteries at each subscriber's premises was obviously undesirable. Common battery signalling, with a central battery at the exchange, was introduced at Boston as early as 1880, and the use of a central battery for both speaking and signalling first came into commercial service at Lexington, Mass., in 1893.

For many decades there was no economic incentive to try to use automatic exchanges, for operators were cheap and equipment was expensive. Nevertheless, A. B. Strowger of Kansas City began the development of his automatic switching system in 1889 (Fig. 50.5); the first commercial automatic exchange was working by 1897; and by 1898 the U.S.A. had 22 such exchanges. The first one in Britain was at Epsom in 1912.

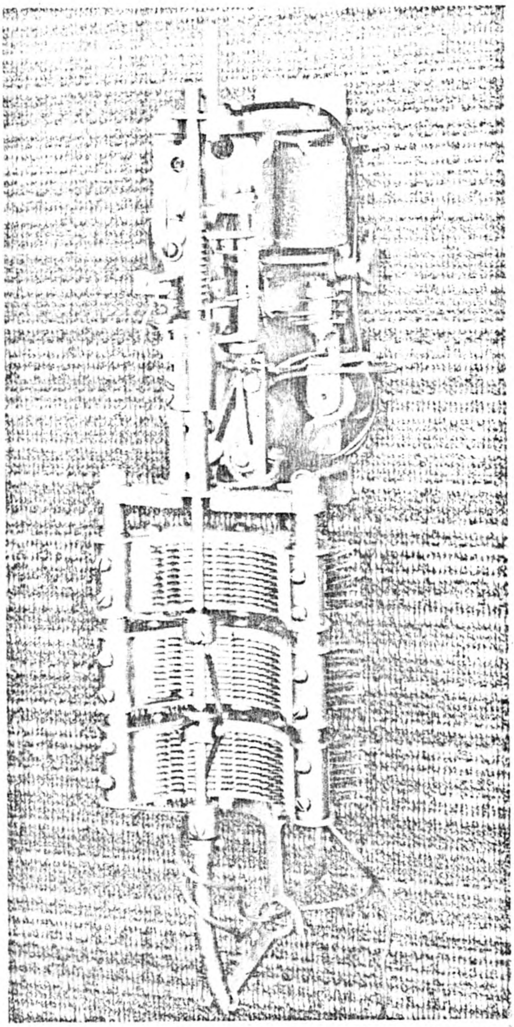

FIG. 50.5. The Strowger automatic telephone switch of 1897–8.

The growth of inter-urban telephone networks. As the number of telephone exchanges grew, so did the demand for links between the exchanges in neighbouring places. Subscribers' lines were usually of the single-wire type, with earth return, so there was no difficulty in linking them by means of single-wire inter-exchange lines. In the U.S.A. there was no obstacle other than transmission and commercial considerations to the growth of inter-exchange networks. In Britain, however, there were legal difficulties imposed by the licensing system. In granting a licence to a company, the Post Office restricted the area to be served under the licence to a radius of four or five miles. It was thus impossible for a company to develop a network which could be classed as inter-urban. This disability was removed in 1884, and thereafter there were a few notable inter-urban networks built up, an outstanding example being that of the Lancashire and Cheshire Company which by 1886 had about 3000 km of inter-urban lines.

Long-distance telephony. The main obstacle to the achievement of effective long-distance telephone communication was the interference due to telegraph signals on adjacent wires. This 'induction', as it was called, could seriously interfere with telephony even over very short distances, and it entirely prevented telephone communication from taking place over distances greater than about 50 km except at night when the telegraphs were quiet. It was appreciated that the trouble was largely due to the use of single-wire lines with earth return, and as early as 1877 the use of 'metallic', that is two-wire or loop, circuits without an earth connection had been suggested. In such circuits the lateral induction would be opposite in the two wires and would therefore tend to cancel out, especially if the wires were run on a twist system so that each wire had the same average spacing from the source of interference. In spite of this early knowledge of the technical cure for induction, however, the use of single wires continued because of the great expense of providing metallic circuits. It was not until the turn of the century that metallic circuits became more or less universal, although there were some notable examples of their use in the early 1880s—for example, the line between Boston and Providence, U.S.A., about 80 km, opened in 1882 between special switchboards at each end, which had metallic-loop extensions to special subscribers. Another major exception was the whole of the telephone system operated by the Post Office in Britain from its very beginning in 1881, including trunk lines in South Wales extending to 130 km, over which speech was said to be good.

Induction was not the only problem in long-distance telephony; another was the poor quality of speech transmitted over iron wires. The use of iron wire in telephony was only slowly abandoned, but most long-distance lines were of copper or phosphor-bronze from the middle 1880s.

To enable single-wire telephone lines to operate in the vicinity of telegraph wires, François van Rysselberghe (Belgium) introduced some special anti-interference methods. Realizing that the interference from telegraph to telephone circuits was largely due to the transients caused by the rapid rise and fall of the telegraph impulses, he showed early in 1882 that suitable chokes connected in the telegraph circuits could reduce interference to a satisfactory level on adjacent telephone lines. From there it was a short step to showing that the telephone circuits could actually be superimposed on the telegraph

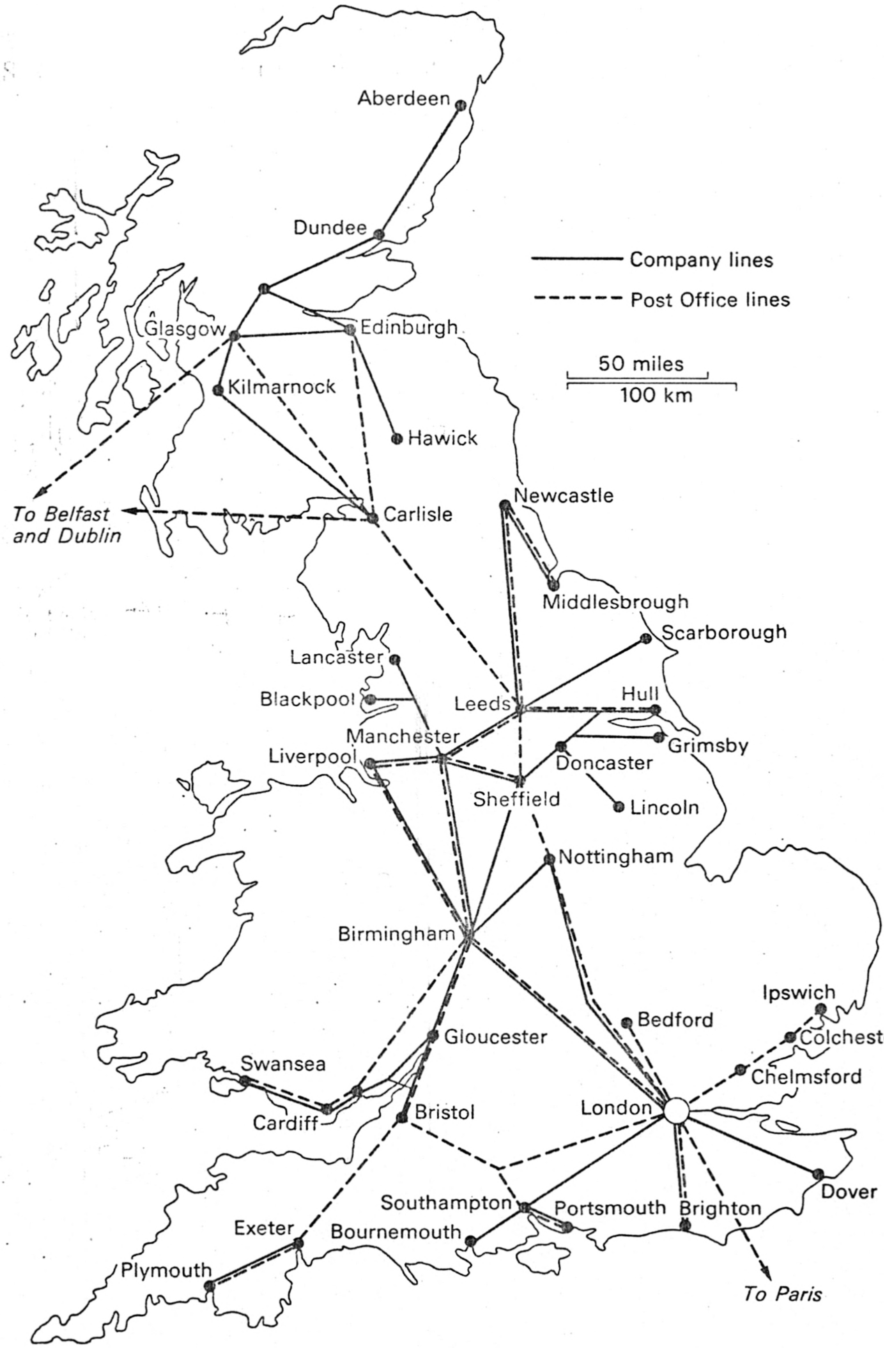

FIG. 50.6. The trunk telephone network in Britain in 1896, distinguishing between the lines constructed by the Post Office and those taken over from the National Telephone Company.

lines by means of capacitance couplings. Thus a technical and economic solution of the long-distance telephony problem was available, and was immediately taken up by the Belgian authorities and, before long, by many others. Telegraph lines existed over Europe and in other countries on an extensive scale, and it was attractive to be able to work telephony over them at low cost. By mid-1887 there was a long-distance telephone network in Europe amounting to over 17000 km, all on the van Rysselberghe system; there were many similar lines in many other parts of the world, including South America, China, and Japan.

Britain was markedly behind in the provision of trunk lines owing to the Government's anti-telephone policy, but the amalgamation of the numerous telephone companies into the National Telephone Company in 1889 led to a considerable improvement. Mounting public pressure led to the establishment of a proper trunk telephone network (see Fig. 50.6) by the Post Office in 1896, incorporating also the trunk lines of the National Telephone Company which had been compulsorily purchased (leaving the Company with only its local systems and subscribers). The network of trunk lines on the Continent of Europe had by then grown into a largely interconnected system (Fig. 50.7). With the availability of lines, demand grew quickly, and these

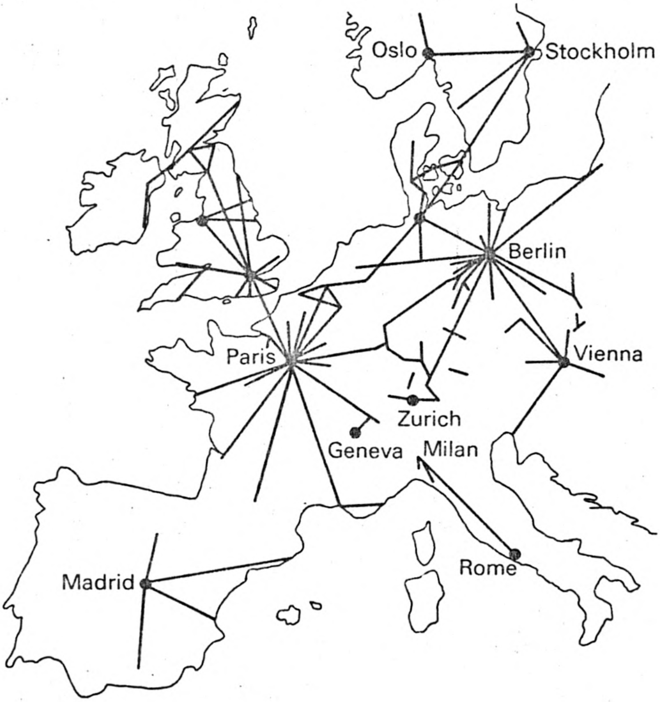

FIG. 50.7. The trunk telephone network in Europe in 1896. N.B. this map is a modern compilation and may well have inadvertently omitted some important links of which records have not been found.

networks were rapidly extended and consolidated. By this time, metallic-loop working of independent telephone lines was becoming normal everywhere, and the van Rysselberghe system gradually went out of use except for some special cases.

In the U.S.A., interest in long-distance telephony arose even earlier than in Europe. There were a number of experiments in long-distance working from 1879 onwards. The Southern New England Telephone Company opened in 1884, for public service, a metallic-loop circuit (claimed to be the first metallic long-distance telephone line ever built) between New York and Boston, approximately 450 km in length. This proved very satisfactory. In 1886 the newly formed American Telephone and Telegraph Company (A.T. and T.), generally referred to as the 'Long Distance Telephone Company' and a subsidiary of the Bell Company, embarked on a programme of long-distance lines, several routes being opened in 1887–9 over distances up to about 700 km. The success of these led to the construction in 1892 of a through metallic-loop circuit from New York to Chicago with copper conductors of No. 8 gauge (121 kg/km), a substantial increase in conductor size over the previously-used No. 12 gauge (48 kg/km). The lines could be extended at each end to give through speaking between Boston and Milwaukee, about 2100 km, but this was considered rather beyond the limit of satisfactory transmission. Normally special soundproof booths were used at New York and Chicago.

By the early years of the twentieth century, the A.T. and T.'s network had linked all major towns in the eastern half of the U.S.A. The first transcontinental line, from New York to San Francisco, was opened in 1915. This, however, depended for its operation on the use of coil loading and, most important of all, telephone repeaters or amplifiers.

III. THE BEGINNINGS OF RADIO (NINETEENTH CENTURY–*c.* 1910)

Origins. There were, from 1838 until the end of the nineteenth century, numerous attempts at communication by electrical signals without the use of wires to connect the two communicants—thus introducing the concept of 'wireless' communication. These mostly relied on conduction, generally through water, or on inductive coupling between what were effectively coils of wire with transmitted signals in the same form as on lines; they gave some useful practical results in communicating across estuaries or relatively narrow sea-channels. Their potentiality for communication over long distances was, however, negligible; inductive coupling gives, for a constant transmitted power at the frequencies of telegraph or telephone signals where the wavelength is several hundred kilometres, a received signal power inversely proportional to the sixth power of the distance, provided the distance is large compared with the dimensions of the coils and small compared with the wavelength. They were thus necessarily short-range systems and almost irrelevant to the modern 'radio' method of communication, which owes its importance to the use of radiated electromagnetic waves whose power at a distant point is inversely proportional to only the square of the distance.

The concept of electromagnetic waves was due to James Clerk Maxwell who, in 1864, set out a mathematical statement of their properties; showed that light was an example of an electromagnetic wave; and predicted that waves could be radiated at longer wavelengths with the same velocity of propagation as light. Electromagnetic waves (in free space, remote from their source) are characterized by electric and magnetic fields of equal energy density, with the vibratory electric and magnetic fields mutually at right angles and also at right angles to the direction of propagation. Following Maxwell, many people made experiments and put forward ideas on the subject [4], but it was the work of Heinrich Hertz which in 1887–8 demonstrated clearly that electromagnetic waves could be produced by a sufficiently rapid disturbance (that is, a spark discharge) in an electrical system; that the wavelength (in his experiments between 0·5 and 10 m) was a function of the electrical circuit parameters; that resonance response was exhibited; that the waves were polarized; and that they could be reflected by plane and curved mirrors made of conducting material, and refracted by a dielectric (non-conducting) material. The foundation for radio communication had been laid, but the potentiality for communication was largely overlooked by scientists who followed Hertz, until it was brought to notice by the work of Guglielmo Marconi.[1]

Another important development, made between the Hertz and Marconi periods, was the coherer, a device for the detection of electromagnetic waves. Hertz had used for detecting the reception of a wave a rather insensitive device, namely the observation of a small spark between small spheres fixed to the ends of an incomplete loop of wire. The coherer was in essence a non-conducting tube containing a large number of very small conducting particles (typically metal filings) between two metal contact wires or plates. Normally the resistance of the device was high, because of the loose contact between

---

[1] In passing, it is interesting to note that A. G. Bell in 1878, and others soon afterwards, were able to demonstrate telephony over a modulated-light link, using selenium as the detector; this was a form of electromagnetic-wave wireless communication, but was, of course, irrelevant to the development of radio.

particles, but when an electrical discharge occurred in its vicinity, or a voltage of the high frequency associated with electromagnetic waves was applied to the wires at its ends, the particles to some extent stuck together or cohered, and its resistance fell dramatically. This fall in resistance could be detected by a local battery circuit with some sort of indicator in it. The coherer had then to be tapped mechanically to loosen the particles and thus restore the high resistance.

The coherer had a long prehistory, but in its practical form was due largely to Edouard Branly (France) in 1890–1, although Oliver Lodge also claimed to have developed it, and had undoubtedly done much along similar lines.

Marconi and other pioneers. As early as 1894 the young Marconi, who had little scientific training, had seen the potentiality of Hertz's work for commercial telegraphy, and started experimenting in his home at the Villa Griffone near Bologna in Italy. He had soon tried an elevated aerial with its lower end connected to an earth plate; he used an induction coil to generate sparks under the control of a telegraph key; and he used a coherer for detection, improving its design by the use of nickel filings with an admixture of 4 per cent of silver filings and by exhausting the air from the tube. He was able to achieve a range of 2·4 km.

The Italian Government was not interested in taking up Marconi's ideas, and so in February 1896 he took his apparatus to Britain. Before long, and with the support of the Post Office, he was able to demonstrate radio telegraphy over several kilometres on Salisbury Plain and across the Bristol Channel. He also had valuable contacts with the Admiralty and War Office, and a friendly relationship with H. B. Jackson, a senior naval officer who was himself also a pioneer of radio, particularly for naval use. The friendly relationship with the Post Office terminated, however, in 1897 as a result of the formation by Marconi of a commercial company to exploit his patents, which the British Government had declined to pay for. His experiments continued; in May 1898 he demonstrated telegraphy over a sea path of 23 km from Alum Bay to Bournemouth, and in March 1899 over 50 km across the English Channel from Wimereux near Boulogne to the South Foreland lighthouse near Dover.

On 12 December 1901 Marconi's long-planned attempt to transmit a radio message across the Atlantic Ocean resulted in partial success; he claimed to have heard at his station at St. John's, Newfoundland, signals comprising a series of the letter *S* in Morse code—each letter being three dots—transmitted by prior arrangement from the station in Cornwall, over 3500 km away. There was no instrumental record to support his claim, which has thus always been suspect; however, a recent reassessment in the light of modern knowledge shows that there is now no scientific case against it [5]. At the time nobody suspected the existence of reflecting layers in the earth's upper atmosphere which would divert back to earth some of the energy in the transmitted signal, which probably encompassed the range of wavelengths from 600 m down to 30 m (the frequency range 0·5–10 MHz). However, whatever the real facts of the Newfoundland experiment, there was no disputing Marconi's achievement of a range of about 2500 km between Poldhu and the ship *Philadelphia* in February 1902, for on this occasion, with much better equipment, he obtained inked-paper records of genuine telegraph messages. Only six years later Marconi was able to establish a commercial radio telegraph service across the Atlantic.

There were undoubtedly several quite independent workers in the practical application to radio communication of the scientific matters demonstrated by Hertz, Lodge, Branly, and others. Priorities have been difficult to determine, but it now seems generally agreed that Marconi was slightly ahead of—and was certainly more successful than—Jackson (already mentioned), A. C. H. Slaby (Germany), and A. S. Popov (Russia); these three were his closest rivals.

The beginning of radio technology and radio service. Marconi's early experiments in Italy and Britain were made with transmitters which were still recognizably based on those used by Hertz; Fig. 50.8(a) shows the arrangement. When the telegraph key K is depressed to make a 'mark' signal, the break-and-make contacts of the induction coil IC vibrate. Each time the contacts break or make there is a sudden change of current in the primary winding and of magnetic flux in the iron core; there is thus induced a large electromotive force (e.m.f.) in the secondary winding which has many times the number of turns of the primary. Now the aerial A_1 forms, with the earth, effectively a capacitor (then called a 'condenser'), and the e.m.f. from the coil causes electric charge to flow into it. As it charges up, a voltage develops between A_1 and earth. Eventually (after a time-interval, short compared with the period of vibration of the contacts of IC) this voltage is sufficient to cause a spark between the metal balls of the spark gap G, and the charge stored in the aerial capacitance is suddenly discharged through the conducting path of ionized gas produced by the spark. This provides the energy for the production and radiation of electromagnetic waves in a short decaying burst, of

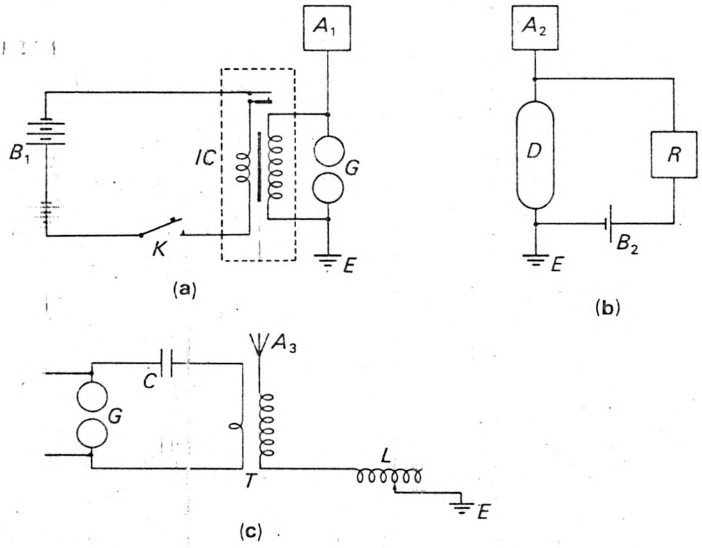

FIG. 50.8. Early radio circuits used by Marconi.
(a). Transmitter, 1896–7.
(b). Receiver, 1896–7.
(c). Transmitter tuning arrangement, 1900; a similar circuit was used in the receiver.

A_1 = plate-type transmitting aerial; A_2 = plate-type receiving aerial; A_3 = antenna-type aerial; B_1, B_2 = batteries; C = capacitor (condenser); D = detector, i.e. coherer; E = earth connection; G = spark gap; IC = induction coil with iron core and magnetic break-and-make; K = telegraph key; L = tuning inductance coil with adjustment; R = receiving instrument (Morse sounder, inker, or telephone receiver); T = transformer, or coupled coils.

duration much less than the period of vibration of the contacts of *IC*, and of frequency and wavelength determined by the capacitance and inductance of the aerial–gap circuit.

The receiver in these early experiments was typically as shown in Fig. 50.8 (b). A_2 is an aerial similar to A_1. When the burst of waves is received, the resistance of the coherer *D* falls, and the battery current through the receiving indicator *R* rises and operates it. The resistance is then automatically restored to its high value by a mechanical tapper on the coherer. The telegraph key *K* is kept depressed for the duration of a mark signal, and this greatly exceeds the period of vibration of the contacts of *IC*; the telegraph signal thus comprises a train of short bursts of electromagnetic waves.

In these simple circuits there is no very definite natural frequency, and so, in Marconi's early work, the transmitted waves covered quite a wide spectrum. This meant that neighbouring radio links would interfere with one another; the difficulty was particularly marked when trying to operate radio telegraphy between ships of a naval fleet. The potential solution of this problem was to apply the principle of tuning (then called 'syntony' (p. 1113)) which had been developed by Lodge during several years before 1897; this principle involved resonating both transmitter and receiver circuits to the same frequency, so that a more limited band of frequencies was transmitted and the receiver was much more sensitive to this band than to any other. Marconi could not use Lodge's patent, but eventually designed a circuit arrangement, shown for a transmitter in Fig. 50.8(c), which achieved the desired result without infringing Lodge's patent—although there was a dispute over this. Here the aerial A_3 (by now a wire antenna replacing the old plate aerials) was tuned to the frequency of the tuned spark-gap circuit by means of the adjustable inductance L, the relatively loosely coupled transformer T making the aerial tuning and the spark-gap tuning largely independent. The receiver had a similar double tuning. This system proved successful.

Many improvements in transmitters and receivers were made in very rapid succession in the years following 1900. We can here mention only a few. Although spark transmitters continued to be used, particularly in marine equipment, for several decades, they had serious disadvantages, and for point-to-point communication were largely displaced by either arc-generators or high-frequency alternators. The former was first used by W. Duddell (Britain) in 1900 as a simple direct-current arc with a resonant circuit connected across it. The negative-resistance characteristic of the arc (that is, the negative slope of the voltage–current curve) enabled an oscillation to be generated at the frequency of resonance of the circuit, the power being drawn from the direct current. The improvements made by V. Poulsen (Denmark) in 1903, which included the use of a water-cooled copper anode and an enclosed atmosphere of hydrogen, together with a strong applied magnetic field to extinguish the arc for a part of each cycle, resulted in a much higher efficiency, and such devices were used for important radio stations for some decades. The high-frequency alternator was a special design of rotating machine which could generate an alternating current of frequency up to about 100 kHz, and is particularly associated with E. F. W. Alexanderson (U.S.A.), who built a successful one in 1907. Both the Poulsen arc and high-frequency alternator eventually provided transmitter powers up to about 500 kW.

In reception, the coherer tended to be displaced at first by the electromagnetic detector of E. Rutherford (Britain). In 1896 he demonstrated the effect of an electromagnetic wave on the magnetism in an iron wire, and Marconi developed the idea in 1902 into a device (Fig. 50.9) with a clockwork-driven soft-iron wire or band which passed continuously through a permanent magnetic field and a coil carrying the aerial current to be detected. The receipt of a wave demagnetized the wire suddenly, thus inducing a click in a telephone earpiece connected in a secondary winding on the coil. In the long-term

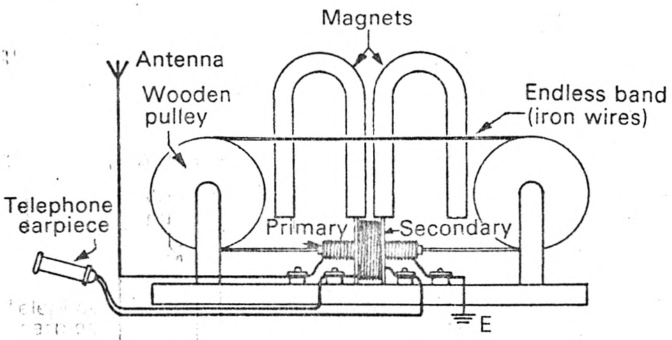

FIG. 50.9. Marconi's second electromagnetic detector, 1902.

aspect, the most important innovation of the early years was the crystal detector or receiver discovered by H. H. C. Dunwoody (U.S.A.) in 1906, but the idea of using a non-linear circuit element as a detector seems to have been due to M. I. Pupin (U.S.A.) in 1898; he used an electrolytic rectifier. Possibly the greatest importance of the rectifier-type of detector was that it could be used as well for radio-telephony as for radio-telegraphy.

The idea that speech might be transmitted by radio had occurred to several of the early workers and ideas of modulating a high-frequency current by speech signals had been put forward as early as 1886 by M. Leblanc (France). However, it was R. A. Fessenden (U.S.A.) who first demonstrated the transmission of speech by the modulation of a radio wave in 1902. For this purpose he used a continuous alternating wave generated by an alternator, and although he developed several methods of modulation, the most successful seems to have been the simplest, namely the direct modulation of the transmitting aerial current by a microphone in the same way that a direct current is modulated in a normal telephone transmitter. Speech could also be transmitted on a wave generated by a Poulsen arc, or even by a spark. However, radio-telephony made only slow progress before the general use of the thermionic valve.

Radio-telegraphy entered successfully into commercial service after 1900. We shall consider its world-wide applications later.

IV. TECHNOLOGICAL DEVELOPMENTS IN ELECTRICAL COMMUNICATIONS

Having in previous sections described the origins of telephony and radio communication and the main inventions and developments which initially established them as vital means of rapid communication, we now turn to the major technological developments which led electrical communications into the modern era.

The impact of the thermionic valve. There is no doubt that the biggest influence in the development of electrical communications into the modern pattern was the introduction of the thermionic valve. Yet its effect was at first very small and slow to emerge.

The history of the thermionic valve (or 'tube' in the U.S.A.) is outlined elsewhere (Ch. 46). Although originating as a diode rectifier, it had little impact in that form. The triode valve or 'audion' invented by Lee de Forest (U.S.A.) in 1906 was a crude device, little understood—even by its inventor—and therefore not very effectively used for some years. It was used to some extent as an amplifier, but the real breakthrough came in the years 1912–15 when three important steps were made. First, the use of the triode as an oscillator, or generator of radio-frequency oscillations, using positive feedback from the output to the input circuit. Secondly, the use of the triode as an amplifier of very high amplification and high tuned-selectivity, also by means of positive feedback, carefully adjusted to avoid the oscillatory condition. Thirdly, the understanding of triode valve operation and the quantitative design of valve circuits through the use of characteristic curves relating anode–cathode current to anode–cathode and grid–cathode voltages. E. H. Armstrong (U.S.A.) contributed more than anyone else to these matters, but the invention of positive feedback was strongly disputed [6]. It now became possible to obtain relatively pure radio-frequency power at frequencies which were readily adjustable and controllable, thus making radio-telephony satisfactory for the first time; it became possible to make extremely sensitive receivers with high selectivity. Because of these factors, together with the hastening effect of the First World War, radio communication—and soon radio broadcasting—made a rapid leap forward.

Many other inventions based on the thermionic valve led to significant advances in radio. One of the most important was the superheterodyne principle, which improved receiver design greatly. The idea of forming a low-frequency 'beat' note between the incoming radio signal and a local oscillation-generator of frequency near to that of the signal was due to Fessenden in 1902, but its development as the superheterodyne, based on valves, was due to Armstrong in 1919. The advantages of the system were that, since the incoming signal was effectively changed to a much lower frequency on reception, the amplification and selectivity could be readily provided at the lower frequency, thus permitting the use of radio frequencies above the range at which effective amplification could then be provided directly. Moreover, selectivity could be better when carried out at the fixed intermediate frequency; this was particularly important when radio broadcasting got under way in the mid-1920s.

Even though the application of the thermionic valve to radio was surprisingly slow, its application to line telephony was quite remarkably laggardly. The problems of long-distance telephony in the U.S.A. had led to the idea of using amplifiers in telephone lines as early as 1900, and electromechanical 'repeaters' based essentially on the coupling of a telephone receiver to a microphone were developed by 1903 [7]. Yet such amplifiers were never introduced on any large scale. They were used initially on the first transcontinental telephone line, opened between New York and San Francisco in 1915. However, the telephone engineers had at last caught up with the thermionic valve, and valve amplifiers replaced the electromechanical type almost immediately, giving much better results. Thereafter, long-distance telephone systems were planned on the basis of valve amplifiers, which permitted the use of much lighter wires, underground cable, and much better quality of transmission. After some years of experience of trying to operate amplified circuits using the same pair of wires for both directions of transmission, it became standard practice to use a separate pair of wires for each direction.

Modulation, sidebands, carrier telephony. The rapid growth of the long-distance telephone network in the last two decades before 1950 was not only responsible for the development of new ways of providing trunk circuits but was to a large extent created by the availability of new methods and the economies achieved by them. Although, as we shall see later, the invention of negative feedback was the main trigger for the trunk explosion, the basic concepts had been laid down during the preceding half-century.

The main feature of the 1930s development was the use of a wide spectrum of frequencies on a line, this spectrum being divided up into bands typically of about 4 kHz, each to carry a speech channel. This system is called 'frequency-division-multiplex' (f.d.m.). All the incoming speech signals from subscribers have to be translated in frequency to the bands allotted. This is done by a process of modulation with, or of, a 'carrier tone'. The use of the radio spectrum had, of course, been similar from the early years of the century, and had involved the same basic process of modulation when speech or programme material was transmitted. This process originated at least as early as 1886 [8]. In that year, M. Leblanc (France) described a way of transmitting speech from a microphone which was combined with a self-maintained tuning fork in such a way that the sound waves impinging on a diaphragm controlled the amplitude of vibration, and therefore of the electrical oscillation sent to a telephone line from a pick-up coil on the fork. The envelope of the waveform sent on the line thus reproduced the sound waves. Leblanc's receiving device was not so clear in its operation. However, about five years later Leblanc, in conjunction with M. Hutin (France), put forward a system using electrical resonance and a dynamometer type of receiver, with different carrier frequencies allocated to different telephones, so that a proper f.d.m. system was produced. It is believed that no practical use was made of the system at that time, but it was undoubtedly the forerunner of modern systems.

Amplitude modulation (A.M.) for radio seems to have been developed quite separately around 1900 by R. A. Fessenden (U.S.A.). In neither line nor radio work, however, was there any understanding of sidebands until around 1915, although Lord Rayleigh had fully discussed them in the context of acoustics in 1894. Yet an understanding of sidebands was essential for progress in f.d.m. telephony, whether on line or radio. If the speech signal is represented by a single tone $e_s \cos \omega_s t$, and the carrier tone (that is, the tuning fork note in Leblanc's system) as $e_c \cos \omega_c t$, then the modulated signal is

$$e_c \cos \omega_c t\ (1 + m\, e_s \cos \omega_s t),$$

where m is the depth of modulation; this can be expanded as

$$\begin{aligned} e_c \cos \omega_c t\ [\text{carrier tone}] &+ \tfrac{1}{2}\, m\, e_c\, e_s \cos (\omega_c + \omega_s)t\ [\text{upper sideband}] \\ &+ \tfrac{1}{2}\, m\, e_c\, e_s \cos (\omega_c - \omega_s)t\ [\text{lower sideband}]. \end{aligned}$$

No speech information is contained in the first, carrier, term, and all speech information is contained in each of the sidebands. There is, therefore, no point in transmitting anything more than one of the sidebands without the carrier. If only one sideband is transmitted for each speech channel, the maximum use is made of the available power and the available frequency spectrum. The problem was, of course, to separate the sidebands. This was solved from 1915 by the invention of electrical filters (Ch. 46). There were many early f.d.m. line telephone systems using one carrier channel in addition to the normal direct speech ('audio') channel on each line. Their importance was small, however, until the use of negative feedback made multi-channel working possible.

Towards the end of our period two other kinds of modulation attracted interest and have become important. First, frequency modulation (F.M.) invented in 1902 but first seriously developed in the early 1930s, in which the information is carried by variations of the frequency and not the amplitude of the carrier [9]. Secondly, pulse modulation, in which information is carried by variations in amplitude, timing, or coding of sequences of short pulses. F.M., although extravagant in spectrum width, has the advantages of being relatively unaffected by fading in radio systems and in usually giving good discrimination against noise; it is, therefore, much used for important high-quality radio work. Pulse modulation has the attraction that it can be digital in form and thus compatible with data-transmission requirements, and it lends itself to another kind of multiplexing—time-division-multiplex (t.d.m.)—where signals from different channels are sampled in sequence. T.d.m. for speech was first demonstrated by W. M. Miner (U.S.A.) in 1903, but for practical application had to await the modern electronic era; the principle had been used in telegraphy from the mid-1870s.

The impact of negative feedback on telephony. We have pointed out earlier how ideas on modulation and frequency-division-multiplex in relation to the trunk telephone network had origins in the nineteenth century, and how greatly the introduction of the thermionic valve had speeded progress in electrical communication generally. Techniques of 'carrier-telephony', as the f.d.m. system was usually called, had been developed, and during the 1920s and early 1930s had been applied to open-wire trunk lines without intermediate amplifiers, and also as two-channel systems on amplified lines. In the latter case, one channel was the normal audio channel, and one carrier channel was placed above it in the frequency range of, roughly, 3–6 kHz. Further progress was limited by the difficulty of intermodulation in amplifiers. Owing to the fact that the instantaneous output voltage of a thermionic-valve amplifier was not exactly proportional to the instantaneous input voltage, but had a non-linear relationship, there was a distortion of any signal waveform applied to the amplifier. For a single speech channel this was generally not serious. However, when two or more channels were worked through the same amplifier, this distortion involved the formation of new frequency-components produced by the interaction of the signals in the different channels, and thus elements of the speech in one channel were transferred to another. Although this 'intermodulation' crosstalk was generally unintelligible, it formed a serious limitation to the extension of the system.

The breakthrough came with the invention of negative feedback by H. S. Black (U.S.A.), and its general application following his publication in 1934 [10]. The term 'negative' is not now generally retained, and indeed it is not strictly quite correct, but it is useful to use it here to avoid confusion with the other type of feedback—positive feedback—which was so extensively used in the early valve period. What negative feedback does, in its simplest conception, is to feed some of the output signal back to the input in opposition to the input signal. If the voltage-amplification ratio of the amplifier is A, and the feedback ratio, determined for instance by a network of resistors, is B, then the overall amplification with the feedback is reduced from A to $A/(1+AB)$; if AB is much greater than 1, this amplification is approximately $1/B$. In other words, the overall amplification becomes practically independent of A. Thus, if the original amplification was non-linear, the new amplification is practically linear; and if the original amplification tended to fluctuate, the new amplification is practically constant. Thus negative feedback could reduce intermodulation crosstalk between channels to any desired level, and by stabilizing amplification could permit much more amplification to be provided in any long trunk route.

So from 1934 there was no further fundamental obstacle to the extension of the trunk telephone system by the use of very wide spectra on the lines, with numerous channels on each pair of conductors. It became common practice to use basic groups of 12 speech channels in the frequency range 60–108 kHz, thus allowing 4 kHz spacing of channels. Several, or many, such groups could be assembled in a wide frequency range as super-groups, and so on. By thus reducing the line costs per speech channel, it became practicable to offer cheap trunk calls; to meet an ever-increasing demand for long-distance telephone service; and eventually to provide an automatically switched, on demand, subscriber-dialling service.

Not only have special types of cable been introduced to transmit the very wide frequency bands involved, but 'microwave' radio links are also commonly used, with wavelengths of the order of 10–30 cm. Wideband cables do not naturally have a loss which is constant over the frequency band, and it is necessary to provide 'equalizing' circuits to correct this.

One could justifiably refer to the trunk 'explosion', for the number of channels provided between important towns, once numbered in single figures, had by 1950 reached thousands. Although, as shown above, the most obvious result of the invention of negative feedback was the revolution in telephone trunk circuits, yet numerous other important improvements in electrical communications—and indeed in electronic technology generally—also arose from the use of negative feedback. Precise amplification, good frequency response, low distortion, etc. became not only available, but calculable and designable.

Cables. Telegraphs had been worked over long submarines cables, and over short lengths of land cable, for some decades before telephony began (Vol. V, Ch. 10). It was, however, appreciated from the beginning that telephony had requirements for transmission rather different from those of telegraph signals, but the nature of the requirements was not understood. Early telephone cable schemes, such as that at Newcastle upon Tyne, England, in 1882, where the City Corporation would not permit overhead wires, demonstrated clearly that telephone transmission on cable was difficult. With the gutta-percha-insulated cable then available, 6–7 km was the maximum attainable distance over which intelligible speech was possible. It was appreciated that, whatever other requirements there might be, it was essential to reduce the capacitance of the cable. Lead-covered cable with cotton insulation was made in Philadelphia in 1884, and in 1889 the use of paper insulation started the kind of cable that remained standard throughout our period. For local telephone distribution, such cables with relatively fine wire (eventually, for example, 5–10 kg/km) were very satisfactory. For trunk lines, however, cables became possible at first only with the use of very heavy conductors. An important early example was the first Anglo-French telephone cable of 1891 (still gutta-percha insulated) which provided two telephone circuits between London and Paris. The land extensions used open wires weighing about 150–200 kg/km, and the 35-km cable itself had conductors weighing about 50 kg/km.

The understanding of the principles of telephone transmission, which were clearly set out by Oliver Heaviside in 1887 but not appreciated for another decade, led to the introduction of loading coils from about 1901 [11]. These made an immense improvement in transmission, and made it possible to use cable for trunk lines up to perhaps 200–300 km. The principle involved was that the natural electrical constants of the cable—resistance (R), capacitance (C), leakage (G), and inductance (L), per unit length—did not have the right relationships for good transmission. Heaviside had shown that for distortionless transmission it was necessary to have $R/L = G/C$, but in cables L was far too low. A good cable would have low R, G, and C in order to keep the attenuation low, but even so it was necessary to add inductance to achieve the relationship specified. Coils were added in series with the conductors at intervals calculated from the new formulae of Pupin. (Open lines normally met the Heaviside condition fairly well, the wide spacing of the conductors giving a lower capacitance and higher inductance than in cables; but even so, some American open lines had coil loading to improve their transmission.)

The real development of telephone trunk networks based on cable came with the introduction of the thermionic-valve amplifier. It then became possible to use comparatively light and cheap cable, with 'repeaters' (amplifiers) inserted at regular intervals, to achieve very long lengths of route, eventually providing a global network. As modulation and carrier techniques developed, enabling many channels to be operated on each pair of conductors, inductance loading became a restriction, for it had bought improved audio transmission at the expense of a rapid cut-off above the audio range. Thus by about 1950, coil-loading had disappeared from the trunk network.

Crosstalk between pairs of conductors in cables has always been a problem. Twisting of the wires in various ways helped, but for modern requirements, special crosstalk-cancelling or balancing networks have proved useful. One of the most important cable developments has been the use, since about 1935, of 'coaxial' cables in which one conductor of each circuit is encircled by the other. This construction is particularly suitable for use with very wide frequency bands, as in modern carrier-telephone or television transmission. One of the most exciting cable developments is the use of trans-ocean telephone cables with built-in amplifiers and other equipment. These are discussed below.

Radio aerials or antennas. The theory of radio aerials (usually called antennas in the U.S.A. and elsewhere) is very complex, and it is understandable that early aerials were designed by trial and error. As we have seen, the early radio workers appreciated that the wavelengths radiated depended on the resonant tuning (if any) of the transmitting aerial, and Marconi found that larger, higher aerials led to radiation at longer wavelengths, and that these led to greater ranges of reception. As early as 1898, A. Blondel (France) had produced a theoretical treatment of the vertical wire aerial, laying the basis of the long-wave aerial theory using the concept that the earth produces an image of the aerial. The concept of 'effective height' (h_e) of a simple wire aerial (for example, straight vertical wire, inverted L, or T configurations) came into use, the effective transmitted power being proportional to $h_e^2 i_A^2/\lambda^2$ where i_A is the aerial current and λ is the wavelength of the radiated wave. It was realized that aerials of this type used at wavelengths of about 15000 m, even when made several kilometres long and some hundreds of metres high, were very inefficient, the power radiated being only a few per cent of the total power fed to the aerial. Much effort was made to improve this efficiency by devices such as multiple tuning and extensive ground-wire systems which by the early 1920s brought efficiencies up to 20–30 per cent. The aerials radiated more or less uniformly in all horizontal directions.

In the late 1910s Marconi and his colleagues, C. S. Franklin and H. J. Round, were experimenting with short waves, with wavelengths in the range 2–15 m. It was known (as Hertz had known) that such waves could be propagated directionally by means of reflectors, made of a network of wires, which could be large compared with the wavelength and thus give a radiation in the preferred direction some hundred times greater than if the power had been uniformly radiated in all directions. The actual radiating element could be half-a-wavelength in principal dimension and much more efficient than a long-wave aerial. Such a system was used successfully by the Marconi group over relatively short paths of about 100 km and then in 1923–4 with a wavelength of the order of 100 m over paths up to nearly 20000 km. The story of the exploitation of this 'beam' system is told elsewhere in this chapter. Franklin went on to design short-wave aerials on the array system, where a large number of individual half-wave aerial elements were suspended in front of reflecting wires in a huge assembly, separated horizontally and vertically by the right distance to keep all elements in phase. Such a system radiates effectively only in a limited angular cone; the larger the array for a given wavelength, the smaller the angle of the cone. The diffraction theory on which it is based was already well known, especially in optics, and the system had been put forward for radio telegraphy as early as 1903 by A. Blondel [12].

Numerous other kinds of aerial were developed over the years and there were many significant improvements of detail. Moreover, whereas the aerials of 1895–1920 were designed largely empirically, a much greater extent of theoretical design became apparent in later years; the introduction of the electronic computer led to aerials being designed theoretically almost down to the smallest detail. However, the principles discussed above remained the foundation of aerial operation, which extended during the 1940s into the range of wavelengths between 1 m and 3 cm. At the lower end of this range the feeders to the aerial were waveguides in place of the wire-line and coaxial-cable feeds used at the longer wavelengths, and reflectors and horns of metal sheet were used in place of wire structures.

For broadcast receivers and mobile-communication radio, special small aerials have been used, often without directivity, although small directional arrays are normally used for reception of VHF and UHF sound and television broadcasts (wavelengths in the range 0·1–10 m).

Telephone exchanges; automatic switching; trunk dialling [13, 14]. We have earlier mentioned that automatic telephone exchanges originated in the

U.S.A.; Strowger's system was the first to be put into practice, in 1897. In 1900 an exchange with a capacity of 10000 lines, on this system, was brought into service at New Bedford, Mass. Subscribers still had local batteries for speaking, but common-battery working became normal within a few years. The subscriber's dial, for generating the impulses which govern the switch movements, was introduced in 1896. Each train of impulses, corresponding to each digit of the telephone number dialled, stepped a selector switch at the exchange. If the subscriber dialled 763, for example, the switch connected to his own line would step to the seventh set of contacts, which would lead to another selector switch which would step to its sixth set of contacts, which would lead to a final selector switch which would step to its third set of contacts: that is, the required subscriber's line. This became known as the step-by-step system.

These early Strowger exchanges were uneconomic in having to allocate an expensive selector switch to each subscriber's line. A much cheaper method was introduced; this was the line switch, where a contact arm would rotate and search a bank of contacts, each set of contacts being the termination of a subscriber's line, to find the set corresponding to a calling subscriber. By this system, only as many selector switches needed to be provided as the traffic density warranted. Thus arose the concept of telephone traffic planning. This had been dealt with in a very empirical way in manual exchanges, but now started to become a mathematical study of some precision. By using the methods of mathematical statistics, coupled with actual observations of the incidence of calls in exchanges, it became possible to calculate the number of switches to be provided at each stage in order to give a specified grade of service, defined in terms of the probability of an incoming call finding a free selector switch for each digit dialled. This flexible approach greatly reduced the cost of automatic exchanges. The mathematical methods improved over the years, and did not reach some sort of finality until around 1950.

Whether telephone switching should be manual or automatic was essentially a matter of economics. In places where the average number of calls per subscriber per day was high it might be cheaper to use an automatic exchange. And since in early systems—and, indeed, in some cases, throughout our period—it was necessary to route calls for other exchanges through a manual switchboard, another criterion was the proportion of calls which could be handled automatically, that is, which were within the one exchange area. These considerations may account for the relatively rapid introduction of automatic exchanges in the U.S.A., compared with a much slower development in Britain and Europe. There had been demonstrations of automatic exchange systems in Britain before 1900, but the first operational installations were in 1912 at Epsom and at the Post Office Headquarters in London. Other systems for automatic exchanges were developed and used elsewhere, but the Strowger system was standardized for Britain.

As the number of automatic exchanges increased, the problem arose of how to deal with large areas. The solution adopted was to make each area one large system, often with a central large exchange and outlying satellite exchanges, all on the same numbering scheme. Calls for subscribers on the same satellite exchange as the caller were routed locally; all other calls went through the central exchange, which had the one manual switchboard for the whole area. In very large and dense areas such as New York, London, Birmingham, etc., a more complex system was necessary, with a complicated pattern of inter-exchange links. Here the 'director' system was used. All subscriber's numbers were of seven digits, although the first three were commonly represented by letters related to the name of the local exchange. When a subscriber dialled a number, the impulses were stored in his own exchange; the first three digits, indicating the desired exchange, were then automatically translated into a routing code of up to six digits in the 'director'. This new set of impulses operated selector switches which connected the call either directly or via intermediate exchanges to the called exchange. The last four digits (being the called subscriber's number within his local exchange) were then transmitted to operate selector switches in the normal way. The first director exchange in Britain was that at Holborn, London, in 1927.

In all these automatic exchange areas the system had to allow for some (initially often most) of the exchanges remaining manual for many years, since it was not economic to replace manual exchanges of recent type until some return had been obtained on their capital cost. The solution in smaller areas was that subscribers on automatic exchanges had to call the operator for connection to manual exchanges; but in director areas subscribers on automatic exchanges were allowed to dial any number in the whole area. In this case, when the called exchange was manual, the first three digits took the call to an operator at the called exchange, to whom the final digits were visually displayed so that she could complete the connection without having any verbal contact with either calling or called subscriber.

The direct dialling of calls over trunk lines from one area into the automatic exchange system of another area began only at the end of our period, around 1950. It had to be preceded by a period of development of alternating-current (a.c.) signalling over trunk lines. The reason for this was that, whereas the old trunk lines had used one pair of wires for each speech channel, thus providing an individual conducting circuit for direct-current (d.c.) signalling, the newer trunk lines worked on the frequency-division-multiplex (or carrier) principle, and no longer provided a d.c. path for each channel. Once sufficiently sophisticated a.c. signalling was available, dialling over trunk lines was possible by sending trains of pulses of a.c. tones. Initially only operators dialled calls, but eventually an almost universal and international subscriber-trunk-dialling system came into use; it began, in Britain, at Bristol in 1958.

Since 1950 a great deal of work has been done on the theory and design of electronic automatic exchanges—a natural application of principles developed in electronic computers (Ch. 48), but with some very special and difficult problems, and with solutions which bear little or no relation to methods used in the conventional exchanges. New methods of electromechanical switching have also been introduced, and partially electronic exchanges are satisfactory; but the economic advantages of all-electronic exchanges are not obvious. However, this development was barely started before 1950.

Theoretical and mathematical developments. It must be obvious that electrical communications could not have developed to the state of sophistication attained by 1950 without a very sound background of theory and mathematical methods. There can, indeed, be few branches of technology so much influenced by, and so dependent on, mathematical treatment. Kelvin's theory of telegraph transmission on cables (1855) and of elementary electrical circuits (1853), Clerk Maxwell's theory of electromagnetic waves (1864), Hertz's development of Maxwell's theory, and Heaviside's electromagnetic theory (1887 onwards)—to mention only the most outstanding work—laid a sound theoretical and mathematical foundation for electrical communications. Few of the engineers of those days understood this theory, and its immediate impact was therefore not as great as its importance warranted. Initial progress was undoubtedly retarded by this, but education in electrical engineering was well established on a sound basis by the end of the nineteenth century, and theory took its proper place. Much of the theoretical work was the application of established scientific theory and established mathematics to the engineering problems, but there has been, all through, a strong element of new theory and new mathematics developed for the problems of electrical communications. Heaviside's 'operational calculus' for solving circuit problems was one example. A more recent case has been the development of a general theory of communication by C. E. Shannon [15] (1948 onwards), based on an understanding of the fundamental nature of the background 'noise' in an electrical system as first set out by J. B. Johnson [16] in 1928, and on the limits set by bandwidth as first explored by H. Nyquist [17] and R. V. L. Hartley [18].

Johnson showed that there was in all electrical systems a basic noise composed of random variations in the potential of any point in a conductor due to the random motion of electrons in the conductor. He showed, too, that the mean power was uniformly distributed over the frequency spectrum passed by the circuit, and was proportional to the absolute temperature of the conductor and to the bandwidth (that is, the width of the spectrum passed by the circuit). However well a communication system might be protected from external interference, such as crosstalk from other circuits, its background noise level could never be reduced below this basic level. Communication theory introduced formally the idea of signal-to-noise ratio as a criterion of performance and related to this the rate at which information could be communicated. This involved a basis for the measurement of information; it was postulated that information, however complex, can, fundamentally, be broken down into a series of binary, or yes–no, decisions. Each elementary decision was a binary unit, or 'bit', of information. A complicated speech sentence may involve some thousands of bits, in some ways analogous to the way it could be transmitted by telegraphy using, for example, the Morse code. The rate at which bits could be transmitted through a communication system with a specified proportion of erroneous detections of the 'yes' or 'no' content of the bits could be precisely related to the bandwidth and to the signal-to-noise ratio. The effect of this theory and its ramifications has been profound since 1950, but it was discernable before then, particularly in the new kinds of communication equipment developed during the Second World War, especially those using pulse modulation, and in radar and sonar systems.

V. GLOBAL COMMUNICATION

Radio versus cable. The first commercial transatlantic radio telegraph service was opened by the Marconi Company between Clifden (Ireland) and Glace Bay (Canada) in October 1907, and the charges for messages were substantially lower than those of the cable companies. The Company's experience from this operation enabled them to propose to the British Government, in 1910, a radio-telegraph chain linking up the whole of the then very extensive

British Empire. Unfortunately, the attitude of the British Government to the problem of the monopoly such a scheme would give the Company in terms of traffic and also of available wavelengths, together with the opposition of the cable companies, resulted in the proposal being repeatedly deferred. A hesitant start in 1914, providing the British terminal at Leafield, was nullified by the First World War. Government opposition to the Marconi Company led, after the war, to a new plan for a Post Office radio network with relay stations at intervals of up to 4000 miles, although the Company had already demonstrated its ability to communicate directly, without intermediate relay, between Britain and Australia. There were further delays, and in 1922 the various Dominions made independent arrangements with the Marconi Company for direct links to Britain. But still the British Government could reach no agreement with Marconi.

The breakthrough occurred in 1924. Until then the plans had always been for long-wavelength (that is, wavelength greater than 1 km) systems, which transmitted and received in all directions. But for some years Marconi had been experimenting with 'short waves' (that is, with wavelengths of about 30 m) which permitted the aerials to be made directional, so that transmissions could be 'beamed' towards the equally directional receiving stations. Moreover, very long distances could be obtained by using reflection from the ionized layers in the upper atmosphere. This development was decisive; a contract with the Marconi Company was speedily signed by the British Government, the work went ahead, and by 1928 there was global radio-telegraph communication, linking Canada, South Africa, India, Australia, and other Dominions to Britain and to one another (Fig. 50.10).

The existing global cable telegraph network, which had by then been established for nearly three-quarters of a century (Vol. V, Ch. 10), was severely challenged by the beamed-radio network, which could offer greatly reduced charges for messages. Cable charges fell, but so also did cable traffic. After negotiations, started in 1928, the British long-distance radio and cable interests—that is Marconi's Wireless Telegraph Company, the Eastern Telegraph Company, the Eastern Extension Telegraph Company, and the Western Telegraph Company, including the Marconi Company's very extensive long-wave network—were amalgamated into Cables and Wireless Ltd. in 1929. This was a holding company which had an operating company, Imperial and International Communication Ltd., over which there was a measure of government control through an official Advisory Committee. In 1934 the operating company took the name Cables and Wireless Ltd., the

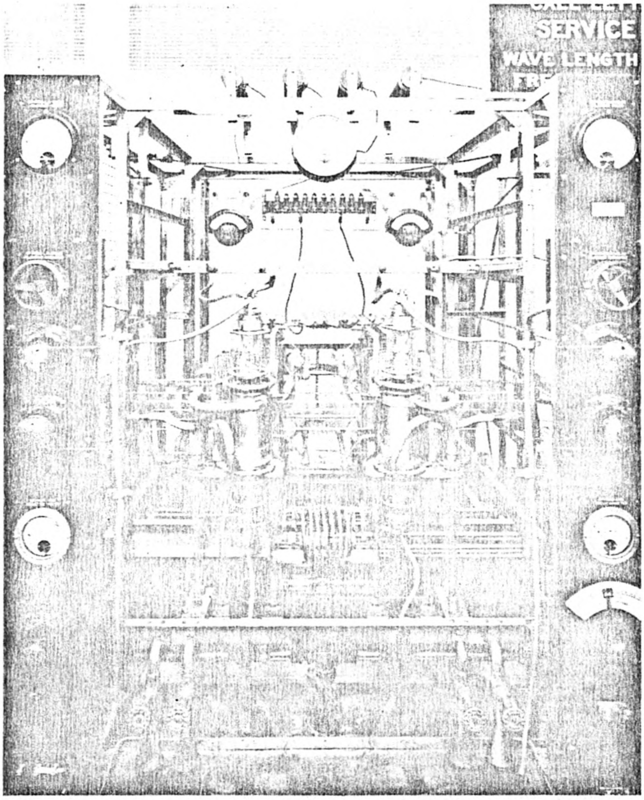

FIG. 50.10. Marconi Company's 20-kW short-wave transmitter Type SWB1, installed at Dorchester in 1927, showing water-cooled valves.

holding company being distinguished by the parenthetic addition of (Holding) to its name. The companies were nationalized after the Second World War.

The ascendancy of telephony over telegraphy. As we have seen, the transmission and reception of telephony on radio links had been established before the general availability of thermionic valves, but it was the valve which made radio telephony suitable for long-distance services. As early as 1915 A.T. and T. had demonstrated the transmission of speech from the U.S.A. to France. By the early 1920s, special high-power triode valves with water-cooling could be used in parallel-connected banks of 20 or 30 to give transmitted powers of up to 300 kW at long wavelengths (5–10 km), that is, at frequencies around 30–60 kHz. An experimental one-way transatlantic telephone link was set up for long-term evaluation by the A.T. and T. in association with the Radio Corporation of America, between Rocky Point, U.S.A., and New Southgate, England, early in 1923 on a carrier frequency of 60 kHz, using single-sideband transmission at 60 kW [19].

The British Post Office decided, as a result of these trials, to set up a commercial two-way link between Britain and the U.S.A., using its Rugby longwave radio telegraph station, which was being established to provide a worldwide telegraph service on a wavelength of 40 km as a rather expensive addition to the beam service already discussed. Aerials for telephony could readily be added, and radio-telephone equipment was provided by Standard Telephones and Cables Ltd., the British subsidiary of the A.T. and T. Both telegraph and telephone transmitters at Rugby were completed in 1926 [20]. The telephone link comprised two separate one-way links, well spaced geographically and operating on different frequencies, in the range 50–65 kHz. The service was successful, although because of the propagation fluctuations and interference it was necessary to have technical operators manually regulating the link all the time speech was in progress. The charge was £15 for a three-minute (effective) call.

Similar telephone services were subsequently provided to Australia and South America, and many other places. The short-wave beam system, so successful with telegraphy, was also exploited for telephony, and it could be said that from the 1930s there was global telephony as well as telegraphy.

As with inland communications, telephone business increased while telegraph business declined. It was evident that provision must be made for the expansion and improvement of telephony on inter-continental services. The concept of the oceanic telephone cable with electronic amplifiers embedded in it arose in the U.S.A. in the 1930s and was developed after the Second World War, leading to a limited installation between Key West and Havana in 1950. From this experience was gained of a system using separate cables for each direction of transmission, each with three built-in repeaters operating in depths up to 1700 m; 24 speech circuits were carried. The repeaters were mechanically flexible, electrically simple, fed with power over the cable. In the meantime, experience had been gained in Britain with submerged repeaters of more complex type. These gave operation in both directions on a single cable, attached to, rather than embedded in, the cables, operating in relatively shallow water on the Continental Shelf. The first such repeater was laid on the route between Holyhead and the Isle of Man in 1943, providing 48 speech circuits. Techniques for making reliable, long-life thermionic valves were successfully developed. This successful experience on both sides of the Atlantic led to the planning and provision of a transatlantic cable telephone system which was successfully inaugurated in 1956 and provided 35 high-quality telephone circuits between Europe and America. Subsequently the demand grew so enormously that by the mid-1970s several thousand telephone circuits were available, by cable and by satellite-radio links, not only across the Atlantic, but all round the world.

Although there is no doubt that Britain played the leading part for more than a century in global communication, and the role of the U.S.A. has latterly been dominant, nevertheless many other countries, and particularly Germany, have been very active.

VI. COMMUNICATION WITH SHIPS AND VEHICLES

Marine radio. Several of the early Marconi experiments involved lighthouses and lightships, and it seemed clear that radio offered a solution to the problem of communicating with ships at sea in all weathers. Marconi and Jackson were much concerned in the years before 1900 in trials which showed the value of radio communication in the Royal Navy, which by the end of 1900 had 51 sets installed. The Marconi Marine Company was formed in 1900 to exploit the Marconi patents in marine applications. It quickly became apparent that the Company itself would have to build the shore stations, hire out the ship-board apparatus, and control the marine communications service. There was soon much competition from other firms, and in Britain also from the Post Office, which in 1909 took over all the coast stations in Britain. The situation in the world generally was far from satisfactory, but eventually all organizations agreed to handle messages from whatever equipment they originated. By March 1915 there were altogether 706 coast stations and 4846 ship-board installations.

The loss of the liner *Titanic* in the Atlantic in 1912 stimulated a demand for compulsory provision of radio on ships. It gradually became obligatory for ships to carry radio; legislation to this effect was introduced for the different classes of ships in the U.S.A., Britain, and other countries from 1914 onwards. The early ship-board transmitters all used spark generators. As valve transmitters became available they were commonly adopted in both naval and merchant ships. In 1927 it was internationally agreed that spark transmitters over 300 watts should be abolished by 1940 because their impure transmissions caused interference problems.

Other important marine radio developments were direction-finders as an aid to navigation (Ch. 34), mainly after the First World War; ship-to-shore telephony from the early 1920s; and automatic distress-call alarms, also in the 1920s. Although single-sideband working was well established in fixed point-to-point radio systems during the last two decades before 1950, it had little influence on marine systems, which remained double-sideband until after 1950.

Vehicular and pedestrian communication [21, 22]. Radio communication with land vehicles lagged far behind marine radio. There had been early attempts to provide communication with railway trains by inductive methods, but these did not come into general use. Radio was the only feasible method for road vehicles, but the problems of effective service with very small aerials and very compact equipment, and with difficult propagation conditions, held back development until the 1930s.

Early systems used frequencies around 2 MHz, mainly with Morse signalling. In the late 1930s police systems in Britain moved to the frequency range 30–100 MHz, with fixed station power of 100 W and mobile station power of 10 W; good telephone communication was generally possible up to about 25 km. Over the world the use of mobile radio in this frequency range, and later, after 1950, in the range around 400 MHz, spread to many private as well as official uses, and equipment became very small and convenient.

The main fundamental problem was, and remains, that of fading owing to the phasing in and out of signals travelling to or from a moving vehicle by different paths involving reflection from buildings and other objects, fixed or moving. These fluctuations are different for different frequencies and for different positions on the vehicle, and 'diversity' methods have often been used in which the outputs from transmissions at different frequencies or from several aerials have been combined to give less serious fading.

Communication with pedestrians, which also started in the 1930s, followed the same lines as for vehicles, but with greater emphasis on compactness and convenience.

The early development of radio communication between the ground and an aircraft followed the same line as marine radio, using medium frequencies (300–3000 kHz). The Second World War led to the use of much higher frequencies (30–300 MHz) as for land mobile radio, giving effective communication over line-of-sight paths.

VII. RADIO BROADCASTING

The beginnings. Radio broadcasting, in the present-day sense of the broadcast transmission of radio sound programmes and messages for reception by anyone who has access to a suitable receiver, had its origins in many experiments in several countries before the First World War, and in several local broadcasting programmes during the years 1914–20 in Belgium, Holland, and Germany. Its genesis as a regular public service may be said to date from 1920. From 23 February 1920 until the licence from the Postmaster-General was withdrawn on 23 November that year, the Marconi Company transmitted regular news broadcasts from its 15 kW transmitter at Chelmsford, England, on a wavelength of 2·8 km. From April 1920, Frank Conrad of the Westinghouse Company broadcast entertainment privately from his home in Pittsburgh, U.S.A., and the Company's official broadcasting station opened in Pittsburgh on 2 November 1920. Two years later there were an estimated million listeners and nearly 600 broadcasting stations in the U.S.A. In Britain, the Marconi Company engaged in further experimental broadcasts early in 1922, and opened the famous 2LO broadcasting station in London on 11 May that year, operating on a wavelength of 360 m. By the end of the year the British Broadcasting Company had been formed and represented the end of commercial competition in the broadcasting field in Britain; the Company became a public corporation in 1927, by which time there were over two million receiving sets in Britain. In the U.S.A., broadcasting has remained commercial, with some Federal supervision; and in the world generally some degree of government control or supervision can be found in every country.

The provision of receivers for radio broadcasting has been a competitive commercial matter in almost all countries, and led to much patent activity and an emphasis on well-engineered, cheap, mass production. There has, however, always been a keen amateur interest in receiver construction; this was particularly marked in the early days when a simple receiver with no thermionic valve and only a 'catswhisker' (crystal) detector was often adequate for use with headphones. A typical circuit arrangement of such a type is shown in Fig. 50.11. The addition of even one valve permitted a loudspeaker to be used. Loudspeakers using a conical metal diaphragm, vibrated by a small magnetic

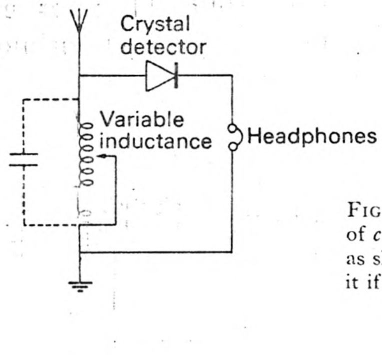

FIG. 50.11. A typical circuit arrangement of a simple 'wireless set' of *c.* 1923. It was generally unnecessary to add a physical capacitor as shown; self-capacitance of the inductor was sufficient to resonate it if suitably designed.

reed in the field of a coil through which passed the audio current from the radio receiver, became available from 1920; the horn loudspeaker followed very quickly. The performance of studio microphones, headphones, and loudspeakers, in terms of frequency response and non-linear distortion, was rather poor at first; constant research and development led to the condenser microphone and moving-coil speaker and eventually to the superb quality required by 'hi-fi' (high-fidelity) enthusiasts.

Developments in broadcasting. The range of the medium-wave transmitters which were first set up, with wavelengths in the range 300–500 m and powers of about 3 kW, was not reliably more than about 40 km. In Britain, 'relay stations' were therefore set up to transmit programmes, received from the main stations, in limited local areas outside the effective range of the main stations. Many countries, including the U.S.A., relied entirely on medium-wave broadcasting. but in some European countries long-wave transmitters were set up with higher powers and the ability to give good coverage of a whole country. The British long-wave station at Daventry, with a wavelength of about 1·5 km and a power of 25 kW, opened in 1925, was the first of these.

The use of relay stations in Britain, and the frequent need for all stations to transmit the same programme, led to the development of a line network to carry programmes from one centre to another. At first, ordinary telephone lines were used, but to ensure acceptable technical quality only open-wire lines could be used, the frequency response of loaded cable being too restricted. Special amplifiers were used, and national networks grew up. As the use of cables with amplifiers became more general for telephony, the provision of special circuits for programme material within telephone cables became common. This started, it is believed, in Denmark and some other European countries about 1927, and in Britain in 1931, when four pairs in the centre of a new London–Birmingham cable were separately screened to prevent interference, and very lightly loaded (15 mH coils at intervals of about 3 km) to give a good frequency response up to almost 8000 Hz. Various means of providing even higher quality were introduced in later years.

An important development in broadcasting, introduced in the mid-1920s, was the use of short-wave transmission (wavelengths in the range 10–100 m) which could be directional if desired, and with quite modest powers of perhaps 10 kW could attain ranges of many thousands of kilometres, relying on reflection from ionized layers in the upper atmosphere.

The development of broadcasting on even shorter wavelengths of 1–10 m (the so-called VHF or very-high-frequency range, 30–300 MHz) started after the Second World War. The object here was not to transmit over a long range, but to provide a better coverage of a region with high-quality reception. This was achieved largely because of the use of frequency modulation (F.M.). In this process the audio signal is carried by variations in the frequency, not the amplitude, of the radiated signal. It is thus relatively unaffected by variations in propagation; it is also less susceptible to interference. Because it requires a much wider bandwidth in the transmitted signal it is not a suitable method of modulation for transmission at lower frequencies, that is, on longer wavelengths.

An important aspect of broadcasting technology is the design of studios and the technical control of programmes. To obtain suitable acoustic performance of studios for different types of programme and to arrange the adjustment and mixing of the signals from arrays of microphones is a highly specialized matter which has been the subject of much research and experiment.

Developments in receivers. In general appearance radio receivers changed considerably in the decades after 1920, but in general principle there were few significant changes. The two most important were probably the introduction of the superheterodyne system of reception, which provided greatly increased selectivity, and the use of automatic volume control.

We have already described the principle of the superheterodyne system (p. 1237). In broadcast-radio receivers it came into use in the early 1930s. The intermediate frequency (I.F.) is usually set at 465 kHz, so that high selectivity can be obtained by fixed and precisely adjusted tuning at this frequency. To tune in to a signal at, say, 1000 kHz, the local oscillator is set at 1465 kHz, so that the image frequency—that is, the frequency which can also enter the I.F. stages along with the wanted signal—is 1930 kHz, which is within the ordinary radio band. It is therefore usual to have some variable tuning, which need not be very sharp, in the amplifier stages which precede the frequency-changer; this is conveniently 'ganged' to the oscillator-frequency control, thus permitting single-knob tuning. Superheterodyne receivers were, of course, considerably more complex than the older 'straight' receivers, and used more valves, typically six or seven, although in Britain and a few other countries several valve units were often combined in the one envelope structure with a common heater.

Automatic volume control also came into use in the early 1930s; it was a system whereby a direct voltage proportional to the amplitude of the carrier signal at the output of the I.F. stage was used to control the gain of the amplifying valves, thus greatly reducing the variations in the output of the receiver when the input signal fluctuated owing to fading, etc. Its operation depended on the 'variable-mu' valve, introduced in the late 1920s; this is a thermionic valve in which the grid is designed with unequal spacing of the mesh-wires so that the negative-grid characteristic of the valve is greatly prolonged along the negative voltage axis. By this means the amplification of the valve is gradually reduced as the grid bias is made more negative. When the direct control voltage, mentioned above, is fed negatively to the grids of the several valves in the receiver, their overall amplification is reduced by a factor many times greater than that by which the output has increased. Thus a hundredfold fluctuation of amplitude in the input leads to a fluctuation of audio output of perhaps only two or three times. The superheterodyne and automatic volume control methods survived into the 1970s, merely with the replacement of thermionic valves by transistors.

VIII. TELEVISION [23]

Origins. The desire to transmit pictures rapidly to a distant place by some electric-telegraphic method was evinced almost from the beginning of electric telegraphy. As early as 1850, F. C. Bakewell (Britain) constructed a system which could transmit hand-written messages (and therefore presumably also line drawings) written in non-conducting ink on a metal plate which was then scanned by a series of styluses each connected in an electric circuit. At the receiving end the currents in each stylus-circuit (directly, or through relays and a local circuit) caused marks on a rotating drum covered with chemically prepared paper; the original message or picture was thus reproduced by white lines. There were several other developments of this idea, and a public picture-transmission service existed in France as early as 1863.

Methods of this kind could not deal with pictures other than line drawings. For transmitting pictures with tones, something having an electrical response dependent on light intensity was needed. When Willoughby Smith (Britain) discovered in 1873 the light-sensitive property of selenium, by which its conductivity was increased when light fell on it, it seemed that the necessary device was at hand, and many proposals for picture transmission were put forward, notably by M. Senlacq (France) in 1878; W. E. Ayrton and J. Perry (Britain) in 1880; G. Carey (U.S.A.) in 1880; and Shelford Bidwell (Britain) in 1881. In some schemes the picture was projected on to a mosaic of selenium

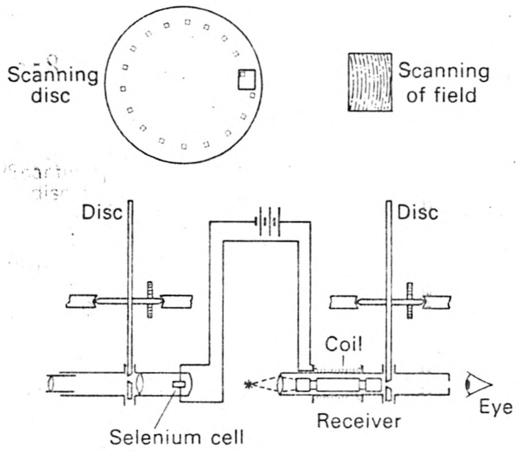

FIG. 50.12. Nipkow's patent, 1884; a redrawing of the diagram published in the original patent specification showing his scanning disc with apertures on a spiral track. A selenium cell was proposed as the light-sensitive element at the transmitter, while rotation of the plane of polarization in a magnetic field was proposed as the light-control at the receiver.

cells, each connected by a separate wire to the receiver (which might be of the electrochemical type); others used the scanning of the picture by a single selenium cell, the transmission over one telegraph line of the sequence of signals corresponding to each scan, and a receiver presumably synchronized. P. Nipkow's scheme of 1884 used a method of scanning that remained important for a long time: viz, a large rotating disc with a spiral series of small holes in it (see Fig. 50.12). Mirror-drum scanning was also proposed at about this time.

It became apparent during these years that selenium had a fundamental limitation to its suitability for picture transmission; its response to variations of light intensity was too slow for there to be any possibility of its serving for the transmission of moving pictures. While it remained a suitable material for the slow transmission of still pictures over telegraph and telephone lines, and was indeed widely used for this purpose as late as 1930–40, it clearly had no future for television. The development of television lapsed for a couple of decades, until the possibilities of the cathode-ray tube became evident in about 1907–11.

The cathode-ray tube had been introduced in 1897. The idea of using it as the means of reproducing the picture at the receiving end was due to B. Rosing (Russia) in 1907, but he still used mechanical devices for scanning at the transmitter, with a photocell of some kind. A. A. Campbell Swinton (Britain) in 1908 specified the nature of a successful television system in terms of cathode-ray tubes at transmitter and receiver:

> . . . two beams of kathode rays (one at the transmitting and one at the receiving station) synchronously deflected by the varying fields of two electromagnets placed at right angles to one another and energized by two alternating electric currents of widely different

> frequencies, so that the moving extremities of the two beams are caused to sweep synchronously over the whole of the required surfaces within one-tenth of a second necessary to take advantage of visual persistence . . .
>
> The real difficulties lie in devising an efficient transmitter which, under the influence of light and shade, shall sufficiently vary the transmitted electric current . . . and further in making this transmitter sufficiently rapid in its action to respond to the 160 000 variations per second that are necessary as a minimum.

The transmitter, he proposed three years later, should have a special cathode-ray tube with a mosaic screen of photoelectric elements on which an image of the scene to be transmitted was projected, and the consequent charges stored on the elements were scanned by, and modulated the current in, the cathode-ray beam. This quite accurately describes the modern television system in its basic principles, but could not be implemented with the technology of 1911, as Campbell Swinton took care to explain.

It was during this period at the beginning of the twentieth century that photoelectric (as opposed to photoconductive) cells were developed; instead of varying their conductivity under the influence of light, they generated a voltage or current, and proved to be much faster in operation, thus permitting fast scanning.

Practical systems: the low-definition interlude. In the mid-1920s determined efforts to launch television were made by a number of workers, especially D. von Mihály (Hungary); C. F. Jenkins (U.S.A.); H. Ives (U.S.A.: Bell Telephone Laboratories); and J. L. Baird (Britain). Of these, the last-mentioned attracted most attention. Starting work in 1923, he gave frequent public demonstrations of his sytem from 1925 onwards. He used a Nipkow disc, with lenses in the holes, for scanning the scene, which had to be very brightly lit. The picture had 30-line scanning at five frames a second. At the receiver, another Nipkow disc, driven in synchronism with the transmitter, rotated between the observer's eye and a neon lamp modulated in brightness in proportion to the current from the photocell at the transmitter. This was, of course, now the era of electronics, and Baird used carrier waves and valve amplifiers as necessary. Although the picture was small (about 5 cm by 4 cm) and dull, and flickered badly, it was just possible to recognize faces.

Improvements came gradually, including the flying-spot scanning system, in which the scene was illuminated by an intense spot of light caused, by a revolving mirror-drum, to scan the scene in 30 vertical lines; and in 1929 the British Broadcasting Corporation allowed Baird's company to start public television broadcasts, still with 30-line scanning but now at 12·5 frames a second; in 1932 the B.B.C. took over the broadcasts themselves. Suitable receivers were marketed from 1930. The service continued until 1935.

Baird's work, more than that of anyone else, showed that television had a future worth working for; but it also showed that the future did not lie with his system. Much higher definition was essential for success, and this could not be obtained satisfactorily, if at all, by mechanical scanning. Campbell Swinton's proposals were now technologically realizable and it was in this direction that development took place.

Practical systems: the high-definition era. From about 1931 the development of high-definition television was aided by a number of important technological advances; it is also arguable that these advances were stimulated by the desire for good television. Some developments proved to be mere interludes, others led directly to post-1950 television practice. The more significant of the advances were four in number.

(i) The improvement of cathode-ray tubes from the older 'gas-focused' tubes where the focus could not be preserved over a range of intensity, to the high-vacuum type with electrostatic focusing.

(ii) The development of electronic circuits as 'time-bases', that is, circuits which generated an output voltage which in each cycle varied linearly from one value to another and then returned rapidly—the 'fly-back'—to the first. Applied to the new cathode-ray tubes these circuits provided the voltage waveform required to scan the beam over the face of the tube: one time-base for the horizontal scan and one for the vertical scan. The time-bases at the receiver could be synchronized with those at the transmitter by the transmission of synchronizing pulses at the start of each cycle.

(iii) The development of a transmitting-type cathode-ray tube (as specified by Campbell Swinton in 1911) by V. K. Zworykin (U.S.A.) from 1925 onwards, and first put forward publicly in 1933. Called the 'iconoscope', later developed to the much more sensitive 'image orthicon', it had a mosaic of photoelectric elements which stored charge produced by the incident light until scanned by the cathode-ray beam.

(iv) The development, a few years later, by J. D. McGee and others of the British E.M.I. Company, of the 'emitron', a device with the same functions as the iconoscope, but somewhat more sensitive. Both these types of tube compared well in sensitivity with photographic film used under the same conditions.

With these developments, picture definition improved rapidly; by 1935

E.M.I. was able to offer the B.B.C. a television system with 405 lines per frame, 50 frames a second, with alternate frames interlaced. This was adopted as standard in Britain, but then the Second World War interrupted public television services in most countries. In America and in other countries in Europe progress was similar, although uniform standards and international television were not provided until after 1950.

Some of the interesting interludes to which reference was made earlier deserve further mention.

(i) The use of velocity modulation. Before the new cathode-ray tubes were developed with their suitability for varying the light intensity on the receiver screen by varying (or 'modulating') the intensity of the cathode-ray beam, the idea of obtaining varying illumination on the screen by varying the speed of scan of the beam was developed by L. H. Bedford and O. S. Puckle of the British Cossor Company. Naturally, the same velocity variations had to be applied at the transmitting tube.

(ii) The use of intermediate film. The Baird Company adopted a 240-line standard when they abandoned the original mechanical system, and developed a flying-spot scanner suitable for scanning 35-mm film. When a live programme was to be transmitted, it was first filmed, the film was processed in about 60 seconds, and then immediately scanned for television transmission.

(iii) Large-screen television. A British firm, Scophony, believed that there was a demand for large-screen viewing, especially in cinemas. They developed a technically successful system, which by 1938 could give a picture about 4·5 m by 4 m with the B.B.C. standard 405-line definition. The demand was, however, only short-lived.

One of the effects of the new standards of picture definition was the need for a wide frequency bandwidth in transmission. The very rapid scanning of the picture gave rise to very rapid changes of voltage, which required a frequency range up to 4 MHz. To modulate this on to a medium-wave carrier of around 1 MHz was clearly quite impossible, and so radio frequencies around 40 MHz (wavelengths around 7 m) had to be used—the so-called VHF band. Later, after 1950, much higher frequencies also came into use—the UHF band.

Public television service. The phenomenal growth in demand for television and its world-wide provision is too obvious to need description here. The introduction of colour television and the superb technology of the production of both equipment and programmes are great achievements. But the importance of television lies not in these matters, but in the ambiguity of its effect on society. With nuclear technology it stands at the head of the league table of technological developments of greatest importance to mankind. Like nuclear technology, television can have vast beneficent influences on life—in its case these include education, entertainment, and the raising of moral and material standards—but also like nuclear technology, it can destroy—in its case the soul and spirit of man. It needs just as careful handling as nuclear weapons, but unfortunately its dangers are more subtle, insidious, and sinister.

IX. SOUND RECORDING AND REPRODUCTION

Origins. Undoubtedly sound recording and reproduction has for over half a century been a part of electrical communications; in contrast, the first half-century of the art had practically no connection with electricity.

Since reproduction of any sound recorded is an essential feature of any useful system, the beginning of the art may be taken as 1877; however, the idea of recording sound waves on a solid medium covered with lamp-black, by means of lateral undulations of a line, is attributable to L. Scott, who described his 'phonautograph' in 1856. In 1877, C. Cros (France) described a system in which the recording was to be made by Scott's method on a disc with spiral traverse, from which a steel copy was to be made by photoetching; this could be used for reproduction by a stylus coupled to a diaphragm. In the same year, T. A. Edison (U.S.A.) made his 'phonograph' (Fig. 50.13), using metal foil fastened to a rotating cylinder to record sound waves by variation of the depth of impression—the 'hill-and-dale' method; recording and reproduction used the same stylus and diaphragm.

During the subsequent fifty years the purely acoustic (or mechanical) system made some progress. The term 'gramophone' was introduced by E. Berliner (U.S.A.) in 1887; he soon introduced the mass production of disc records from a permanent master with lateral undulations on a spiral groove, the system which became standard thereafter.

Disc recording in the electronic era. The availability of electronic amplifiers in the 1920s meant that voice power was no longer necessary for cutting the master record. A microphone of good quality could be used, generating very little power. Amplifiers provided the power for cutting. This method of 'electrical recording' was used from about 1926. With other acoustical improvements, the frequency range of the recording was improved to 100–5000 Hz. Electromagnetic 'pickups', converting the vibrations of the stylus (or

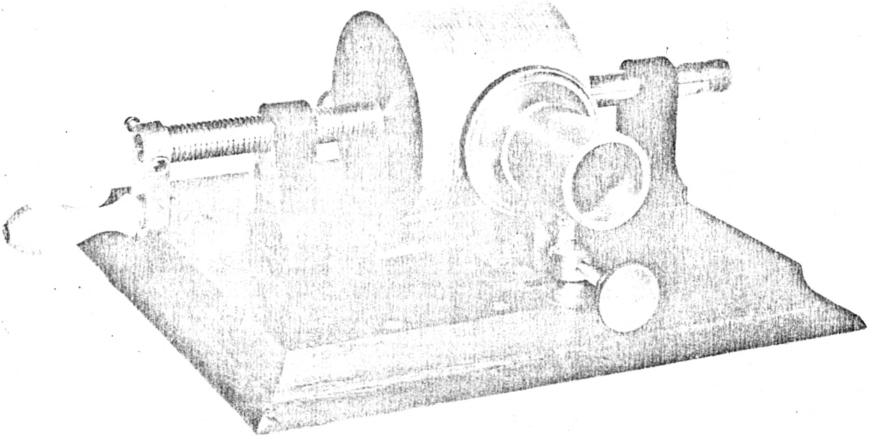

FIG. 50.13. Edison's phonograph, 1877.

'needle') into electrical signals for amplification and use with a loudspeaker, quickly followed. They were rather heavy: about 100 g. In the late 1930s, 'crystal' pickups, of about 25 g weight using piezoelectric elements, became available; by 1950 effective masses of only a few grammes were achieved. This was important for the long playing (LP) type of record that was introduced just before 1950. With a playing speed of $33\frac{1}{3}$ rev/min instead of the 78 rev/min which had been standard for so long, they achieved not only long playing times (20 minutes or more) but also much higher quality of reproduction. The 'hi-fi' age had arrived.

In order to reproduce the spatial effects of music, especially when played by large orchestras spread over a large area of a hall, stereophonic recording and reproduction had been considered from the 1920s onwards. Using two channels (from two suitably placed microphones), techniques largely due to A. D. Blumlein (Britain) for recording and reproduction using a single groove were developed before 1950, but were not generally available until after that date.

Film recording. Early 'talkies', or motion pictures with synchronized sound, introduced in 1927, used disc recording for the sound (Ch. 53). The complications of synchronizing the film and the sound led to the development of direct recording of the sound on a narrow track at the edge of the picture film. For studio work it was found better to record the sound initially on a separate film. Optical systems were used at first. The earliest equipments used a deposit of variable density on the film, produced by modulating the intensity of a narrow light beam. Reproduction was by scanning the film with a narrow light beam, the modulation produced in it by transmission through the film being picked up by a photocell. Later, a variable area system was found preferable to the variable density system. Still later, around 1948, magnetic recording was applied to film work.

Magnetic recording. The recording of sound on magnetized tape or wire came into general use only a few years before 1950, but its history is much longer. The idea was first introduced in a practical form by V. Poulsen (Denmark) in 1897–8. His 'telegraphon' used a steel wire wound spirally on a drum, with a magnetic recording and reading head traversing longitudinally as the drum rotated. The sound to be recorded was converted by a microphone into an electrical current which, flowing through a coil in the magnetic head, varied the magnetization of the wire and was thus recorded until wiped out by demagnetization.

Development of magnetic recording was slow until, in the 1940s, a new magnetic medium was produced: a flexible plastic tape coated with fine magnetic particles. Good dynamic and frequency response became possible and progress became rapid. Since then tape recorders have become commonplace for entertainment, research, and office purposes. Techniques have been developed to the point where even television picture signals, with a frequency range up to 4 MHz, are recorded on videotape.

Microphones. The purpose of a microphone is to convert the sound wave into an electrical signal of corresponding waveform. This it can do in one of two ways: either by generating an electrical signal directly, or by modulating a current or a voltage, or both, already existing so that the amplitude variations of the current or voltage follow the variations of instantaneous amplitude of the sound. The latter method is the more powerful and, as seen under the heading 'Telephony', is the one almost universally used in telephone systems, where carbon-granule microphones are used. In these, the sound pressure varies the contact resistance through a cell of close-packed hard-carbon particles and thus varies the line current. This provides a powerful microphone giving effectively a high amplification, but it is noisy and unsteady. It is difficult to use this type of microphone for high-fidelity work, where electrostatic or capacitance microphones came into use before 1920. In this kind of microphone, the diaphragm forms one plate of a capacitor, the other plate being the solid back-plate, with a spacing of perhaps 2×10^{-5} m. A static voltage, applied between the two plates, is therefore modulated as the diaphragm vibrates under the varying sound pressure.

The class of microphone which generates an electrical signal directly from the sound wave is represented by the electromagnetic microphone which Bell used in 1876, which was thereafter little-used as a microphone but was universally used in reverse as a telephone receiver; by the moving-coil and ribbon microphones, also electromagnetic; and by the piezo-electric microphone which came into use for special purposes in the 1930s.

Loudspeakers. In very general terms, loudspeakers are based on devices for converting electrical signals back into sound waves, which are merely those used in microphones in reverse. The carbon-granule microphone is not reversible in this way, but most other types are. These devices are coupled to a sound-radiator, typically a paper cone, but in the early years of radio broadcasting and electrical gramophones, more usually a metal horn. The acoustical problems of loudspeakers have always given great difficulty, and the contribution of mathematical theory in this field has been large. The performance of early loudspeakers was very poor, but by 1950 'high fidelity' was commercially achievable.

X. SIGNIFICANT TRENDS DISCERNABLE BY 1950

During and after the Second World War progress in electrical communication was rapid, and trends for the future became apparent. The story of their development belongs to the post-1950 period, but some of them, not discussed elsewhere in this work, must be mentioned here [24].

The development of radar stimulated the opening up of the frequency ranges above 300 MHz, generally known as the microwave region, where wavelengths are below 1 m. In this region there are many attractions for communications. Aerials can be highly directive and efficient for point-to-point communication; propagation can be along lines of sight and relatively free from fading and interference over distances of, say, 50 km; long links can be built up by using intermediate relay stations; and very wide bandwidth can be available, for example, 100 MHz when the radio frequency is, say, 6000 MHz. This gives spectrum space not only for the thousands of telephone channels required on major inter-urban links by 1950, but also for television transmission links. The further development of world-wide communication by microwaves reflected or relayed from artificial satellites in space, although not begun until after 1950, was foreseen as early as 1945 [25].

The growing importance of pulse transmission, using pulses shorter by several orders than the conventional telegraph signal elements, heralded the possibilities of time-division-multiplex systems and data-transmission systems, which by 1960 were already being accepted as the communication systems of the future.

Perhaps the most spectacular changes of the post-1950 era have been associated with the replacement of the thermionic valve by the transistor (Ch. 46) as the basis of what are now called 'active' circuits and systems. The trend was hardly noticeable in 1950, for the transistor was then only two years old as a commercial device.

REFERENCES

[1] Hounshell, D. A. *Technology and Culture*, **16**, 133 (1975).

[2] Heaviside, O. *Electrician*, **18**, 302 (1887).

[3] Tucker, D. G. *Proceedings of the Institution of Electrical Engineers*, **123**, 561 (1976).

[4] Susskind, C. *I.E.E.E. Spectrum*, **5**, 90, August (1968); **5**, 57, December (1968); **6**, 69, April (1969); **6**, 66, August (1969).

[5] Ratcliffe, J. A. *Electronics and Power*, **20**, 320 (1974).

[6] Tucker, D. G. *Radio and Electronic Engineer*, **42**, 69 (1972).

[7] Gherardi, B. and Jewett, F. B., *Transactions of the American Institution of Electrical Engineers*, **38**, 1287 (1919).

[8] Tucker, D. G. *Radio and Electronic Engineer*, **41**, 43 (1971).

[9] Tucker, D. G. *Radio and Electronic Engineer*, **40**, 33 (1970).

[10] Black, H. S. *Bell System Technical Journal*, **13**, 1 (1934).

[11] Brittain, J. E. *Technology and Culture*, **11**, 36 (1970).

[12] Blondel, A. *Mémoires du Côngrès de l'Association française à Angers*, 538 (1903).
——. *Compte rendu hebdomadaire des séances de l'Acadamie des sciences, Paris*, **147**, 673 (1908).

[13] Scowen, F. *Transactions of the Newcomen Society*, **47**, 35 (1977).
Flowers, T. H. *Exchange systems*. Wiley, London (1976).

[14] Williams, M. B. and Ireland, J. C. *Electronics and Power*, **22**, 173 (1976).

[15] Shannon, C. E. *Bell System Technical Journal*, **27**, 379 and 623 (1948).

[16] Johnson, J. B. *Physical Review*, **32**, 97 (1928).

[17] Nyquist, H. *Bell System Technical Journal*, **3**, 324 (1924).

[18] Hartley, R. V. L. *Bell System Technical Journal*, **7**, 535 (1928).

[19] Nicholls, H. W. *Journal of the Institution of Electrical Engineers*, **61**, 812 (1922–3).

[20] Oswald, A. A. and Deloraine, E. M. *Electrician*, **96**, 572 and 666 (1926).

[21] Drybrough, D. A. S. *Proceedings of the Institution of Electrical Engineers*, **122**, 953 (1975).

[22] Brinkley, J. R. *Radio and Electronic Engineer*, **45**, 551 (1975).

[23] Garratt, G. R. M. and Mumford, A. H. *Proceedings of the Institution of Electrical Engineers*, **99**, Part 3A 25 (1952).
[24] Halsey, R. J. *Engineering*, **184**, 432 (1957).
[25] Clarke, A. C. *Wireless World*, **51**, 305 (1945).

BIBLIOGRAPHY

Black, H. S. *Modulation theory*. Van Nostrand, New York (1953).

Fahie, J. J. *A history of wireless telegraphy, 1838–1899*. Blackwood, Edinburgh and London (1899).

50th anniversary number of *Proceedings of the Institute of Radio Engineers* (1962).

Heaviside, O. *Electrical papers*. Macmillan, London (1892).

Hunt, F. V. *Electroacoustics*. Wiley, New York (1954).

Jolly, W. P. *Marconi: a biography*. Constable, London (1972).

Marland, E. A. *Early electrical communication*. Abelard-Schuman, London (1964).

Moncel, Count du. *The telephone, the microphone and the phonograph*. Kegan Paul, Trench and Co., London (1884).

Palmer, L. S. *Wireless principles and practice*. Longmans Green, London (1928).

Pawley, E. L. E. *B.B.C. engineering, 1922–1972*. British Broadcasting Corporation Publications, London (1972).

Pocock, R. F. and Garratt, G. R. M. *The origins of maritime radio*. H.M.S.O., London (1972).

Rhodes, F. L. *Beginnings of telephony*. Harper, New York (1929).

Ruhmer, E. *Wireless telephony*. Crosby Lockwood, London (1908). [First published in German, 1907.]

Snel, D. A. *Magnetic sound recording*. Philips, Eindhoven (1958).

Sturmey, S. G. *The economic development of radio*, Duckworth, London (1958).

Symposium on the trans-atlantic telephone cable, special volume of the *Proceedings of the Institution of Electrical Engineers*, **104**, Part B, Supplement 4 (1957).

Transistors—the first 25 years, 1948–1973. Special volume of *Radio and Electronic Engineer* (1973).