From HISTORY OF TECHNOLOGY, ed. T. I. Williams, Vol. 7,
Oxford, 1978, pp. 1091–1125

46

ELECTRONIC ENGINEERING

J. R. TILLMAN AND D. G. TUCKER

ELECTRONIC engineering as such is a product of the twentieth century, for the electron itself was not identified and described until almost the very end of the nineteenth century. Nevertheless, the subject grew naturally out of electrical engineering and, as will be seen from this chapter, many aspects of the subject as it is now broadly recognized by engineers have their roots in the nineteenth century. It was the development of the thermionic valve that created electronic engineering; in this device electrons are emitted from a heated surface in a rarefied gas or in a vacuum and are subjected to electric forces in such a way as to provide a variety of effects which can be exploited to generate, amplify, and control electrical signals. It is the processing of signals by electronic devices which essentially constitutes electronic engineering. The thermionic valve, having created electronic engineering, has now been largely displaced by solid-state devices such as the transistor; but this development had barely commenced by 1950.

The early applications of electronic engineering were almost entirely in electrical communications, and these have been described elsewhere (Ch. 50). Subsequent applications to industry, computers, radar, etc. are also described elsewhere. This chapter, therefore, concentrates on the history of the fundamental devices, circuits, and basic systems of electronic engineering. In preparing it, we quickly found that we could not write for the wholly non-scientific layman without making the chapter unacceptable to the physicist or engineer. We have therefore assumed a little knowledge of physics. For example, it is not possible to evaluate the contributions of such pioneers as Lee de Forest and E. H. Armstrong without some explanation of the vitally important feedback principle. Readers daunted, even by this, should nevertheless be able to pick out the historically significant events in the history of electronic engineering with which we are here concerned. In the few circuit diagrams that are included we have used British Standard symbols.

I. THE THERMIONIC VALVE

When once established in triode form as a reproducible device, the thermionic valve was steadily developed to exploit more fully its ability to amplify, rectify, and modulate electrical signals, and to act in other capacities; for example as a switch, a generator of various kinds of waveform, an information store, and a logic unit in computers. But its earlier history was piecemeal, offering little prospect of wide-ranging usage. It began with observations of effects which were later shown to be manifestations of electronic emission.

Electronic emission. In 1873, F. Guthrie [1] noted effects with heated metals brought near to an electroscope, for which we now know electrons emitted from the metals must have been responsible. Ten years later, T. A. Edison [2] reported that a current could flow in one direction only between a second electrode and the heated filament in a vacuum incandescent lamp. The effect defied correct explanation until 1897 when J. J. Thomson [3] established the existence and elementary properties of free electrons. Thus he measured the ratio of the charge to the mass of the electron, and showed evidence that all free electrons, however generated, were indistinguishable from one another. In 1898, by making an approximate measurement of the charge, he deduced that the mass of the electron must be about one-thousandth that of the hydrogen atom.

Other electronic effects were discovered during the next decade, laying the foundations for electronic engineering. It became recognized that when moving electrons collided with matter, radiation was produced; X-rays, discovered by W. K. Röntgen in 1897, were a notable example. A converse effect had been investigated by J. Elster and H. Geitel (1889 onwards), following H. Hertz's observation that the discharge of a spark gap was influenced by ultraviolet illumination. They demonstrated that such illumination caused metals to emit electrons in numbers proportional to, but with emission velocities independent of, its intensity; moreover, above a limiting wavelength there was no emission. These findings, and M. Planck's quantum theory (1900) put forward to explain the failure of the classical wave theory to account for a key feature of radiation from hot bodies, led Einstein, in 1905, to propose his simple photoelectric emission equation

$$h\nu = q\phi,$$

where ν is the frequency of the longest wavelength capable of exciting emission, h is Planck's constant, q is the electronic charge, and ϕ is the work function of the metal ($q\phi$ is defined as the energy needed to remove an electron from the metal to a great distance).

Of other studies, closely related to the above results, that of O. W. Richardson [4] (1901) was of most direct relevance to the thermionic valve; using hot platinum sources, he concluded that the temperature dependence of thermionic emission of electrons was dominated by a term exp $(-q\phi/kT)$ where k is Boltzmann's constant and T is absolute temperature.

The cathodes of the first valves used refractory metals, principally tungsten, in the form of filaments. Such metals have rather higher values of work function (e.g. 4·5 electron volts (eV)) than do other metals obtainable in wire form, necessitating a high operating temperature (e.g. 2500 K) but ensuring a sufficient margin between this temperature and their melting-points. Thoriated tungsten, first introduced by Irving Langmuir (U.S.A.) in 1914, had a work function of only 2·7 eV; running at about 1650 K, it showed a clear advantage and found use for a while. But in turn it was largely superseded by the oxide cathode, due originally to A. Wehnelt [5] in 1904 and comprizing a thin layer (about 0·0025 cm) of mixed oxides of the alkaline-earth metals, particularly strontium and barium, suitably activated by a partial reduction. The work function was sufficiently low (about 1·2 eV) to enable adequate emission to be obtained at temperatures of only 1000–1100 K, with considerable saving of power and increase of life.

The thermionic diode. Although, as we have seen, the unidirectional conductivity between a plate and the filament inside an electric filament lamp had been well explored by the turn of the century, no practical application seems to have been put forward until J. A. Fleming proposed its use as a detecting device in radio receivers in 1904 [6]. The conventional receiver of that time used a coherer or an electromagnetic detector (see Ch. 50), either of which could quite strictly be called a detector because its response to the reception of a radio signal was to bring in an auxiliary circuit of greater power. It had been known since 1898 that another way of making evident the receipt of a radio signal was to rectify it, so that either the envelope component (which was usually at an audible frequency) or the direct current (d.c.) component could be 'detected' by earphones or galvanometer respectively. It occurred to Fleming that the thermionic diode was an ideal device to perform this rectification, and he used it in a radio receiver in the way exemplified in Fig. 46.1(a). The way it worked in principle is shown in Fig. 46.1(d). The radio signal, which at that time would probably be in Morse code, consisted

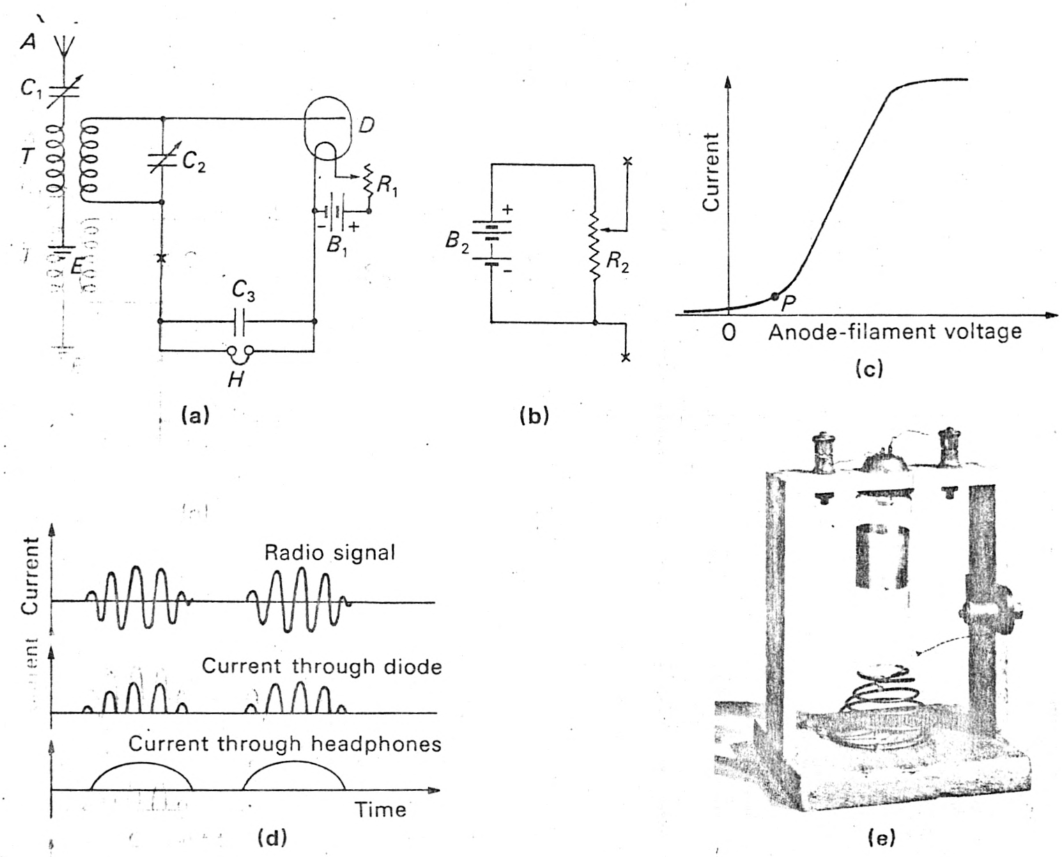


Fig. 46.1(a). A radio detector circuit using a Fleming diode, *D*.
(b). A bias circuit to improve sensitivity, inserted at *X* in (a).
(c). The current/voltage curve for a diode. *P* is the operating point under biased conditions, as in (b).
(d). The principle of operation.
(e). The Fleming thermionic diode (1905), in a wooden stand.

A = aerial; B_1 = battery for filament; C_1, C_2, = variable capacitors (condensers); C_3 = capacitor to bypass headphones, H, for radio-frequency currents; T = aerial tuning coils or transformer; E = earth; R_1 = resistance for adjusting filament temperature; R_2 = resistance for adjusting bias voltage; B_2 = bias battery.

of short transmissions, controlled by a telegraph key, each made up of a series of bursts of oscillations from a spark transmitter as shown in the first line. The diode would pass only the positive parts of the current waveform, so that the current in the diode was as shown in the second line. The radio-frequency components of this passed through the capacitor C_3, leaving the current through the headphones as the enveolpe component shown in the third line.

After a time, Fleming realized that significantly greater sensitivity could be achieved if the diode were used a little differently. Instead of relying on the idealized concept of a device that passed current equally easily on all positive parts of the wave, but no current on negative parts, which we see from Fig. 46.1(c) is a long way from the truth, he added a biasing battery in the diode circuit, as shown in Fig. 46.1(b), so that there was a current even when no signal was being received, as indicated by the operating point P on the curve in Fig. 46.1(c). The difference in current between positive and negative swings of the received wave is now greater than the 'rectified' wave previously obtained. This method was patented in 1908.

Thermionic diodes have been used for a wide variety of purposes over the years; in small sizes as signal rectifiers, in somewhat larger sizes as rectifiers for electric mains supplies in domestic radio receivers, in large sizes as power rectifiers. High vacuum has been needed for some applications; in others, residual gas has been tolerable or even desirable.

The thermionic triode. It was an American, Lee de Forest, who first discovered the advantages of adding a third electrode to the thermionic device [7]. There was much bitter litigation over the invention; Fleming claimed, not without considerable justification, that the triode was a modification of his diode and depended on it; de Forest claimed it was a new invention. Uncertainty as to who had the patent rights delayed the development of applications throughout the patent period. Nevertheless, the triode led to a number of vitally important new achievements.

De Forest had been working on detector systems for radio signals during 1905–6, and, allegedly without knowing of Fleming's work, had tried using a filament lamp with a platinum plate added. He went on to add a third electrode, and tried several different forms of it: a piece of foil outside the tube; a second plate inside the tube, opposite to the first plate; and, finally, a wire grid between the filament and the original plate. It was this final arrangement which became the standard triode valve or tube (Fig. 46.2). (The name 'valve', derived from the valve action of Fleming's diode, carried on in Britain as the accepted name; but in the U.S.A. the device soon became called the 'vacuum tube'.) De Forest called his device the 'audion', but the name lasted barely a decade.

For some years the audion was used only as a more sensitive detector in radio-telegraph reception. The additional electrode, which soon became called the 'grid', was connected to the signal input via a capacitor, C_4 in the typical circuit of Fig. 46.3(a), the resistor R_2 shown there being absent in these early days. With the capacitor in the grid connection, the audion could operate as a detector of telegraph signals because the grid current which flowed

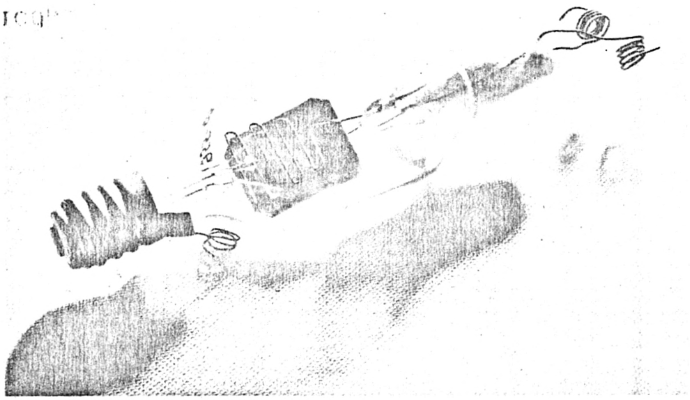


FIG. 46.2. The de Forest audion valve.

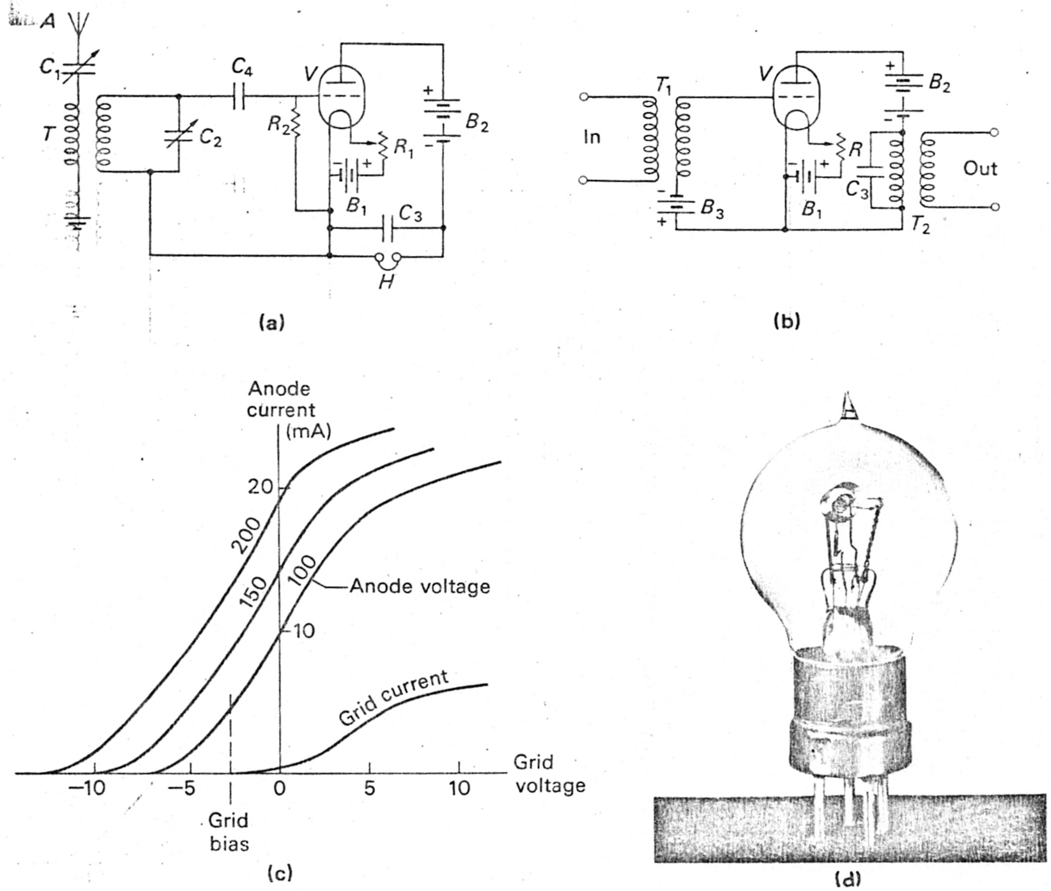


FIG. 46.3(a). A radio receiver using a triode valve, V.
(b). An amplifier using a triode valve, V.
(c). Typical characteristic curves of a small triode.
(d). The French or Army 'R' triode valve, *c.* 1915.

B_1 = filament battery; B_2 = anode (or plate) battery; B_3 = grid bias battery; C_4 = grid condenser; R_2 = grid leak; T_1, T_2 = input and output transformers. Other symbols as in Fig. 46.1.

on each positive excursion of the incoming bursts of damped waves (for example, from a spark transmitter) drained charge from the capacitor and made the grid voltage negative with respect to the filament, thus reducing the anode current. At the end of each burst, the capacitor would recover by the reverse flow of current due to the ionization of the gas in the tube. Headphones in the anode circuit would give an audible signal from these variations of current at the repetition frequency of the bursts of waves. When 'hard' valves came into use, a grid leak (R_2 in Fig. 46.3(a)) was necessary to provide for the recovery of the capacitor charge, and this was invented by H. J. Round (Britain) in 1914.

The real importance of the triode, however, lay not in its detection capability but in its power of providing amplification of a signal. With the capacitance-coupled grid this could not be achieved because the modulation of the anode current did not even roughly follow the signal waveform. It was therefore a big step forward when F. Lowenstein (U.S.A.), in 1912, discovered the favourable effect of a negative bias of moderate voltage, applied by a battery in series with the grid circuit, as shown in Fig. 46.3(b). Then, for the first time, the valve could be worked under its best operating conditions. The signal voltage swing could be kept within the range where at the positive extremity the grid current became excessive, and at the negative extremity the valve cut off. In this way the signal power absorbed could be made very small indeed, and the power amplification comparatively large. It was later found that Round's grid-leak and capacitor arrangement (Fig. 46.3(a)) could, by suitable choice of values, give a suitable steady grid bias voltage when the signal was of the continuous-carrier kind, as in radio telephony, thus dispensing with the need for a battery in many applications.

Triode amplifiers with two or more stages, to give an amplification of perhaps 20 dB, came into use, and in 1915 were applied to the trans-continental telephone line between New York and San Francisco.

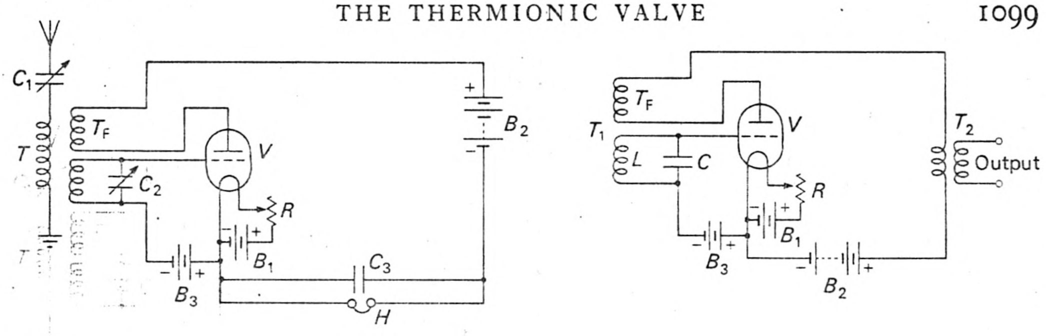
Early valves contained residual gas, and became known as 'soft'. Their performance was irregular and they could not stand more than a few tens of volts on the anode. 'Hard', or high-vacuum valves came into use in the early 1910s. Better mechanical design enabled electrode spacings to be reduced, and greater control of the anode current by the grid voltage was obtained, thus making higher amplification possible.

Although de Forest can justly be credited with the invention of the triode, it is clear from his writings that he never understood its operation or how properly to use it. The clarification of the physical principles involved was largely due to I. Langmuir (U.S.A.). The formulation of the principles of use in circuits was largely due to E. H. Armstrong (U.S.A.), whose paper of 1915 was masterly [8]. Using measured characteristic curves of the type shown in Fig. 46.3(c), he showed how the triode could be used as a rectifier, amplifier, or both together; how feedback could be applied to increase the amplification or to make a self-oscillating detector; how grid-to-anode capacitance could be exploited, and so on.

The feedback principle, just mentioned, was a vitally important one. There was much dispute and litigation as to who invented it. In the U.S.A. the main contenders were de Forest and Armstrong, but there were also several British and European inventors, all producing feedback circuits in 1913 [9]. There were two main uses of feedback, which was conceived at that time as positive feedback: first, to provide a very high amplification in a triode circuit, accompanied by increased selectivity, and secondly to provide a generator of high-frequency oscillations using a triode. These two uses are illustrated in Fig. 46.4(a) and (b). Fig. 46.4(a) shows a radio receiver circuit in which the triode has a negative grid bias and operates as a combined detector and amplifier through the lack of symmetry of the anode current–grid voltage curve about the bias voltage point. The output part of the circuit includes the winding T_F on the input transformer T, and this feeds some output signal back to the grid in phase with the input signal. Consequently a given output requires a much smaller input; that is, the receiver is much more sensitive. The feedback also makes the tuning sharper. Fig. 46.4(b) shows a triode amplifier circuit in which the output is coupled back to the input, with a coupling strong enough to sustain self-oscillation approximately at the resonance frequency of the tuned circuit LC. Such an oscillation generator was very much more satisfactory for radio transmitters than the spark, arc, and rotating-machine types hitherto used. It also lent itself much better to modulation by speech, and thus helped the advance of radio-telephony.

As valve circuits developed and became increasingly versatile, it became essential to integrate the triode into circuit theory. From the use of characteristic curves, as started by Armstrong in 1915, there evolved over the next decade a concept of valve parameters, which soon led into the use of equivalent-circuit representation for the triode, and then, later, for other types of valve. These parameters were essentially partial differentials expressing the interdependence of two variables while all others were kept constant. The three most important parameters were:

(i) The amplification factor, $\partial V_a/\partial V_g$ for constant anode current.
(ii) The mutual conductance, $\partial I_a/\partial V_g$ for constant anode voltage.
(iii) The internal resistance, $\partial V_a/\partial I_a$ for constant grid voltage.



(a) (b)

FIG. 46.4(a). A radio receiver with reaction or retroaction, i.e. with positive feedback.
(b). An oscillation generator. If the coupling in T_1 is fairly loose, the frequency of oscillation is approximately $f = (1/2\pi)(1/LC)^{\frac{1}{2}}$ Hz.

T_F = feedback winding or coil on aerial tuning transformer T; T_1 = the transformer that couples the feedback from the anode (winding T_F) into the grid circuit winding, which has inductance L henries; C = the tuning capacitance in farads. Other symbols as in Figs. 46.1 and 46.3.

For most purposes of circuit design and calculation, these three were sufficient, since grid current was often negligible, and inter-electrode capacitances could be considered as external added elements. (This discussion is, of course, highly simplified, being intended only to introduce the basic ideas.)

Equally it became important for valve designers to understand the characteristic curves in terms of their design parameters. Now Richardson's equation expressed the total (saturation or temperature-limited) current that could be drawn from an emitter—by applying high enough voltage to a collecting electrode. This saturation current is, however, 10 to 100 times the anode current with the recommended electrode voltages. The difference is accounted for by the charge on the electron which produces in the dense cloud of slowly moving electrons, emitted from and very close to the cathode surface, an intense negatively charged (space-charge) region. The region imposes a retarding field which decelerates these electrons so much that most of them return to the cathode. The retarding field, and hence the fraction of electrons not so returned, is, however, influenced by the voltage of the adjacent control grid and to a lesser extent by that of the anode. This leads to the basic equation relating I_a to V_g and V_a, namely

$$I_a = k\,(V_g + V_a/\mu)^n$$

where $\mu = \partial V_a/\partial V_g$ and is dependent on the inter-electrode spacings; k is also a geometry-dependent constant; and n is about 1·5 for indirectly heated and about 2–2·5 for directly heated cathodes. This equation and its constants

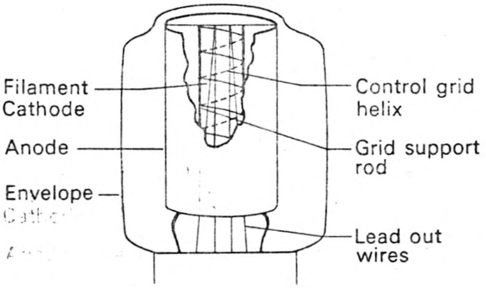


FIG. 46.5. A directly heated triode.

are little affected by any loss of saturation emission so long as that emission remains at least a few times I_a.

The design and construction of valves improved over the years, without however any revolutionary change in the basic structure of the essential triode (see Fig. 46.5), other than the addition of further electrodes (mostly other grids) and the substitution of indirectly heated for directly heated cathodes. In the earlier directly heated valves the cathode took the shape of one or more hairpins, heated by the passage of direct current. The control grid, closely surrounding the cathode, comprised a helix of fine wire wound on two support rods. The anode took the form of a pair of plates (or, commonly, a partially flattened cylinder) at a much greater distance, positioned to receive only those electrons which could reach it through the grid. Where necessary, small electrically insulating spacers, usually of mica, were used to maintain the electrodes at the specified spacings. The electrode structure was mounted on a pinch, integral with the well-evacuated envelope, containing lead-out wires connected to a cemented-on base with positioned contact pins. The introduction of the indirectly heated cathode in about 1930 brought two advantages. Firstly, raw alternating current (a.c.) from the mains, merely transformed to a low voltage, could be used to power it, thus eliminating the need for a battery or a rectifier and smoothing circuit. Secondly the cathode voltage (*vis-à-vis* ground) was more at the circuit designer's choice. The heater was a twisted metal hairpin coated with a refractory insulator, inserted into a cylindrical cathode; the outer surface of this cathode was sprayed with a mixture of the appropriate carbonates and binder, subsequently heated to convert the carbonates to oxides, and activated by a short period of running at a high temperature.

For long-range radio communication and for broadcasting, very high signal power was required for feeding to the transmitting aerials, and special triodes were developed for this use, calling for special materials and water-cooling of the anode or the whole valve.

The thermionic tetrode. As the construction of triode valves improved, giving higher amplification, and the use of multi-stage amplifiers became common, trouble was experienced from the capacitance between anode and grid. This capacitance limited the frequency range over which the valve could operate, and, perhaps more seriously, it caused unwanted feedback leading to self-oscillation in what was supposed to be a signal-amplifier. To overcome this latter trouble, the neutrodyne principle was introduced by L. A. Hazeltine (U.S.A.) in 1923, by which some signal was fed back by a separate path to cancel the undesired feedback due to anode-to-grid capacitance [10]. However, a better solution was a valve which had a greatly reduced capacitance. This was the tetrode valve. It had a most important influence on radio-receiver design in the 1930s.

The idea of introducing an extra grid (later called the screen-grid, or just 'screen') between the normal grid and the anode, to reduce back-coupling between the anode and the grid, was due to W. Schottky (Germany) in 1916–19 [11]. It was made effective, however, by Round in 1926. The screen-grid was maintained at a constant high voltage and so accelerated electrons from the cathode; but being of open mesh allowed them to pass through to the anode. The anode current was thus almost independent of the anode voltage; that is, the internal resistance was nearly infinite. It was a good valve for both voltage and power amplification.

There was one feature of the screen-grid valve which, while being capable of exploitation for some special purposes, limited its use in general. This was the effect of secondary electrons; that is, those emitted by the anode as a result of the bombardment by the normal, or primary, electron stream. These interfered with the amplifying function of the valve when the relative voltages of the screen and anode were such that the secondary electrons were attracted to the screen. Over a limited range of anode voltage, below the screen voltage, the anode current was actually reduced as the anode voltage was raised, giving a negative-resistance effect which could be exploited in a circuit such as the 'dynatron' oscillation-generator. When a tuned circuit was connected across the negative resistance, its resistance losses were neutralized and it oscillated spontaneously at approximately its resonance frequency.

The thermionic pentode. Although the tetrode was, in general, a good valve, and remained in production and use for some decades, the difficulty with secondary emission from the anode could be overcome by the use of yet another electrode, the suppressor grid, inserted between the screen-grid and the anode. This grid would usually be connected directly to the cathode, itself at or near earth voltage, and so repelled the secondary electrons back to the anode. Being of open construction, it did not prevent the primary electrons, accelerated by the screen, from reaching the anode. This five-electrode valve, the pentode, was the invention of G. Holst and B. D. H. Tellegen (Holland) in 1926–7 [12]. It was a most versatile and successful type of valve, and formed the active element of most electronic developments until displaced by the transistor in the 1950s and 1960s.

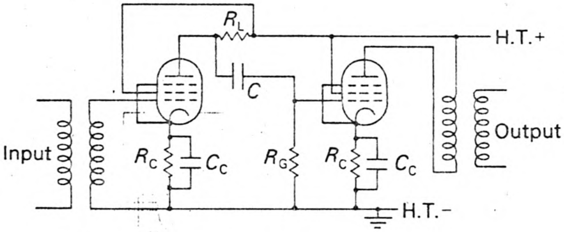
It is impossible to give any account here of the vast variety of circuits which were developed during the 1930s and 1940s to exploit the pentode. All that can be attempted is an explanation of how the pentode was used in some basic amplifier applications.

In Fig. 46.6(a) is shown a two-stage pentode amplifier using input and output transformers. In both valves the suppressor-grid is connected directly to the cathode, and the screen-grid directly to the high-tension positive voltage. Other arrangements were often used, but these are the simplest. The first stage develops its amplified signal voltage across the anode load resistance R_L, and this voltage is transferred to the grid of the second valve by the capacitance-resistance coupling C, R_G which prevents any high-tension direct voltage from reaching the grid. The second valve develops its amplified signal power into the load via the output transformer. Negative bias for the first, or control, grid of each valve is obtained by a method that became almost universal; namely, by the voltage drop in the resistance R_C connected in the cathode circuit. The capacitance C_C is to smooth out any signal component. This system was much more convenient than the use of grid-bias batteries. The valves are shown as having a simple cathode, indirectly heated. Small filament-type pentodes were made for use with dry batteries.

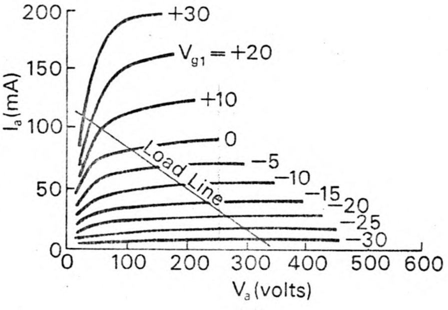
Figure 46.6(b) shows a typical set of characteristic curves for a pentode. The 'load line' has a slope of $1/R_L$ mhos[1] and terminates at the right-hand end at the voltage of the high-tension supply, since this is the voltage of the anode when the anode current is zero. The grid should normally not be swung positive by the applied signal because the consequent flow of grid current would lead to distortion. Equally, it should not be swung negative too far, for the cramping-up of the lines at the larger negative voltages also indicates distortion. So R_C is chosen to give a bias voltage of about -10 in the numerical example shown.

---

[1] The mho is the unit of electrical conductance, and is the reciprocal of the ohm, the unit of resistance.



(a)



(b)

FIG. 46.6(a). An amplifier circuit with two pentodes. (For symbols see text)

(b). Curves for a typical power pentode. V_a = anode voltage; I_a = anode current; V_{g1} = control-grid voltage.

In the output stage, the operation is slightly different. Here, assuming an ideal transformer, the mean, or 'quiescent', anode voltage is equal to the high-tension voltage. The signal voltage from the first stage then causes the anode voltage to swing above and below this high-tension voltage. For a power stage, this is evidently a much more efficient way of working the valve.

Other types of valve circuit. In the preceding discussion of thermionic valve circuits the range of operations involved has been confined to the three very basic ones of rectification, amplification, and oscillation-generation, the last being intended to provide a substantially sinusoidal waveform. These operations provided for most of the electronic requirements of communications, and the very few industrial applications, up to the early 1930s. Other early needs had arisen, however, as exemplified by the time-base circuits of cathode-ray oscilloscopes. During the 1930s the growing development of television and military electronics (for example, radar) put a new emphasis on the need for valve circuits to perform functions such as time-bases with a strictly linear sweep; generators of very short pulses; triggers which could change the state of a circuit from one condition to another on receipt of a suitable pulse signal; circuits which could multiply or divide the frequency of a wave or pulse train by given factors; circuits which could radically change the waveform of a signal in a specified manner; etc. A whole range of new kinds of circuits emerged which is far too complex to be discussed in any breadth or depth here.

Since it is so important, the time-base with linear sweep will be considered as an example of new ways of using valves. It is based on the relaxation oscillator, known in the 1920s. In essence, this simple device is as shown in

Fig. 46.7(a), where N is typically a small neon lamp, and not a thermionic valve. When the battery B is first connected, current flows through the high resistance R to charge the capacitance C, and the rise of voltage is as shown in Fig. 46.7(b). If N were absent, then in a time $t_1 = RC$ the voltage would rise to $V_B[1-\exp(-1)]$, which is approximately $0{\cdot}63\ V_B$. However, the neon lamp has the property of being an open circuit until the voltage across it reaches a certain value V_1 at which the gas ionizes, a glow discharge results, and the lamp provides a very low resistance—a short-circuit in comparison with R. So when V_C reaches V_1, the capacitance begins to discharge very rapidly through the lamp; at the voltage V_2, however, the glow can no longer

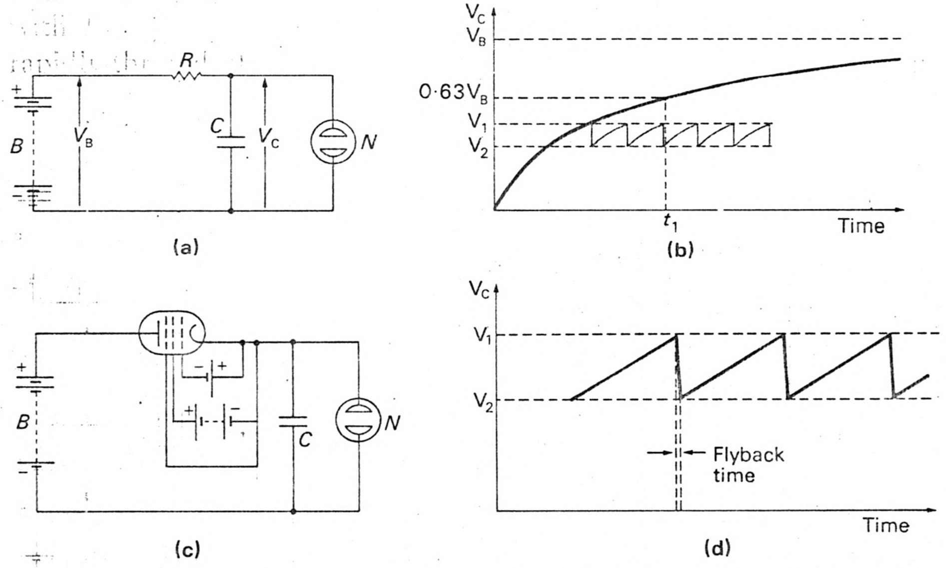


FIG. 46.7(a). A circuit for a simple relaxation oscillator.
(b). The variation of V_C with time.
(c). The use of a pentode valve to give linear sweep of voltage.
(d). An expanded portion of the sweep waveform for a circuit with pentode valve.
(For symbols see text)

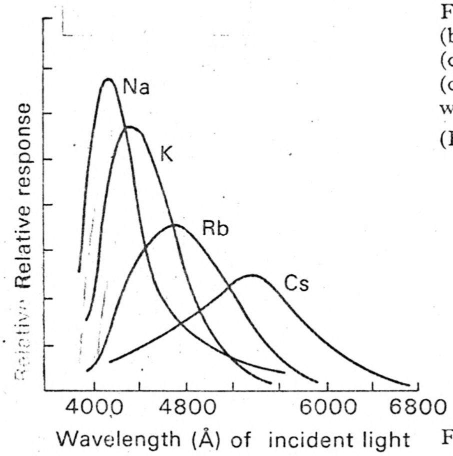


FIG. 46.8. The photoemissive response of the alkali metals.

be maintained and the lamp returns to the open-circuit condition and C begins to charge up once more. The process is repetitive as shown in Fig. 46.7(b), so that the voltage V_C has a succession of gradual sweeps upwards with rapid 'flyback'. These sweeps can be used for the time-base deflection in a cathode-ray oscilloscope, but they have the disadvantage of not being linear. The reason for this is clearly the fact that the rate at which C charges is a function of V_C because the current in R is given by $(V_B - V_C)/R$, and V_C is proportional to the charge acquired.

To obtain a linear sweep, as is so obviously desirable both for instrumental and television purposes, it is necessary to force C to charge at a constant rate; that is, to make the charging current independent of V_C. A method of doing this was invented by L. H. Bedford (Britain) in 1933, using a property of the pentode valve. The typical pentode curves shown in Fig. 46.6(b) indicate very clearly that for a fixed grid voltage, once the anode voltage is above a certain level the anode current is almost independent of it. This is just what is required. If R in Fig. 46.7(a) is replaced by a pentode connected as in Fig. 46.7(c), then C is forced to charge at a rate practically unaffected by the charge acquired. The sweep is then as shown in Fig. 46.7(d).

II. OTHER DEVICES USING ELECTRONIC EMISSION

Photoelectric devices. An early finding of Elster and Geitel, when studying photoemission, had been that the two alkali metals, sodium and potassium, responded well in a vacuum to the shorter wavelengths of the visible spectrum. By 1920 the responses of the other alkali metals had been measured, extending to longer wavelengths for those of rubidium and caesium. These results (Fig. 46.8) stimulated the development of photoemissive cells. The commonest types employed photocathodes of caesium (partially oxidized) on silver, giving a response throughout the visible spectrum and into the near infrared. Straightforward vacuum photodiodes with anode voltages as low as 10 V responded with negligible delay to incident radiation. Their sensitivity could be increased by up to ten times by the admission of an inert gas, at low pressure, and the use of higher voltages (of the order of 100 V), but at the expense of considerable loss of speed of response.

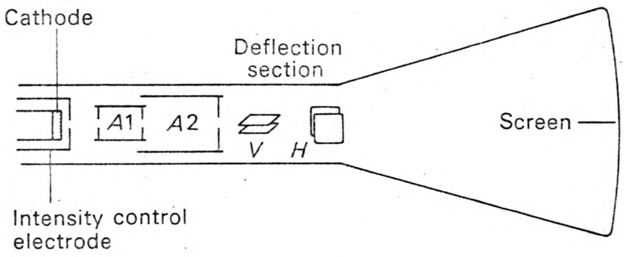
To increase sensitivity, with little loss of speed, the photomultiplier was developed (V. K. Zworykin, G. A. Morton, and L. Walter, 1936) [13]. In it, the primary electrons emitted by the photocathode are accelerated to a first anode (here called a dynode) whose surface is specially prepared to maximize the emission of secondary electrons, which in turn are accelerated towards a second dynode at a higher voltage, where a second multiplication of current takes place. Up to 10 such dynodes might be included before the final anode collects the output current, which could be 10000 times the primary (and even a million times in modern versions).

A more important device (the iconoscope) had earlier been conceived and demonstrated (1929) by Zworykin [14]. It was an image detector, giving more than merely a measure of the total useful optical flux falling upon its photosensitive electrode. An image of an illuminated scene could be focused on to this electrode, which was made up of a mosaic of very small, closely packed globules containing silver and coated with caesium. Light falling on these globules caused them to lose a proportional charge through the emission of electrons. An electron beam, of diameter several times that of a globule, was scanned across the electrode, line by line, in a manner now familiar in television. Each globule restored its charge by extracting electrons from the beam as it passed over it; the sequence of contiguous current pulses thus produced was sensed by a conducting backing plate on the electrode. The iconoscope, and some early developments of it, found increasing use, and were the forerunners of the vidicons, plumbicons, and other television camera tubes so widely used after 1950.

Cathode ray tubes. The development of electronics required new analytical instruments, none more so than one to display the waveforms of electric signals, such as those produced by a microphone when spoken into. Early attempts to meet this need used lightweight pens or galvanometers, but the speed of response, and the bandwidth of the signals satisfactorily handled, were severely limited. F. Braun's earlier work (1897) on deflecting an electron beam [15] led to the recognition that such a beam alone possessed the negligible inertia necessary to allow very high writing speeds, and spurred the efforts to design a suitable cathode ray tube (c.r.t.). The term 'cathode ray' was widely used before 1900 to denote the emission—now known to be a stream of electrons—from the cathode of a gas discharge tube when bombarded by the positive ions of the gas. It survived solely in the context of the c.r.t. and oscilloscopes using it, though no gas discharge is now implied.

The tube has several sections (see Fig. 46.9). The electron beam is produced by the gun, consisting of a thermionic cathode; a sleeve-shaped 'grid' to control the intensity of the beam, with one end closed except for a small central aperture; and a first anode (A1), also sleeve-shaped with diaphragm ends, providing the initial acceleration and the main focusing (on to the final

FIG. 46.9. A cathode ray tube with electrostatic focusing and deflection.



screen) of the beam. Further acceleration is provided by the second anode (A2). The beam then enters the deflection section where the horizontal and vertical deflections are effected separately, the means being either two suitably wound coils producing magnetic forces or two pairs of plates (V and H) producing electrostatic forces. The horizontal deflection is usually a linear time-base repeated regularly, while the vertical deflection represents the amplitude of the signal under observation. The deflected beam finally strikes the luminescent screen, which consists essentially of a thin layer of fine particles of a phosphor exhibiting fluorescence (an instantaneous effect) or phosphorescence (with a delay peculiar to the material used, ranging up to seconds)—or both—under the impact of the beam. Early c.r.t.s used willemite (a zinc silicate containing a manganese impurity) for yellow-green emission, or calcium tungstate giving a blue emission more easily photographed. The needs of radar systems led to improved screen materials, with silicates, sulphides, and fluorides predominating.

Microwave devices. Although conventional thermionic valves met the needs of early radar systems, there soon arose a demand for high peak powers at microwave frequencies, lasting about 1 microsecond (μs) every millisecond (ms) or so, which these valves could not supply. The magnetron, first studied by A. W. Hull in 1921 as a low-power diode using a magnetic field at right angles to the electric field between cathode and anode, was developed, largely by J. T. Randall and H. A. Boot during the early years of the Second World War, to the pitch where it was capable of delivering peak powers of more than 100kW at a frequency of 3 GHz. In this device, the emitted electrons, under the influence of the two fields, pass close to the openings of a series of cavities in the copper block making up the anode (Fig. 46.10). The cavities, together with the remainder of the structure, are of such size as to be resonant at the frequency of the microwave oscillation required. There is complex interaction between the energetic electron stream and the fields

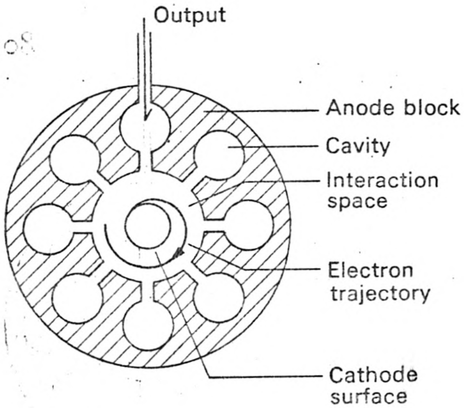


FIG. 46.10. Cross-section of a cavity magnetron (the applied magnetic field is emergent from the paper).

created in the cavities, allowing microwave power to be extracted. The interconnection of alternate segments of the anode by fine wires can improve the performance.

Another promising source of microwave power was the klystron, which owed its origin to studies initiated in 1929 and renewed in the late 1930s. But the development of the magnetron as a high-power source tended to limit the role of the klystron to low-power applications—for example, as a local oscillator in radar receivers—though high-power units were successfully developed after 1950. In the klystron, a broad beam of electrons, already accelerated in a triode section, is passed in turn through an input cavity (called a buncher), a field-free space, and an output cavity (called a catcher) to a collector; the two cavities resonate at the same frequency. The buncher is so called because, for some interval in each cycle of oscillation and relative to electrons entering at the mean time of that interval, its electromagnetic field retards electrons entering earlier and accelerates those entering later, though the actual bunching becomes more pronounced as the beam passes through the field-free space. In the catcher the bunched stream excites an electromagnetic field which can extract more power from the beam than was given it in the buncher, so that suitable coupling together of the two fields leads to oscillation and extractable power at the resonant frequency. This principle of velocity modulation of a stream of electrons offers some advantage over crossed-field operation, as exploited in the magnetron, and many designs of klystron have resulted. A notable example is the reflex klystron, in which one cavity serves as both buncher and catcher, the electron beam being reflected after its first passage through the cavity by a negative-voltage electrode, called a repeller, to return through the cavity.

III. NONLINEARITY AND MODULATORS

It is clear that many, if not most, electronic elements have parameters which depend on the amplitude of the signal applied, because the graphs of response versus stimulus are not straight lines; that is, they are nonlinear. The thermionic valve, for instance, shows this well. From the typical nonlinear characteristic curve relating anode current to grid voltage in a triode (Fig. 46.3(c)) it is easy to see that the mutual conductance $\partial I_a/\partial V_g$ is different for different grid–bias voltages. This was sometimes exploitable and sometimes an undesirable effect. The early use of the triode as an amplifying-detector maximized this effect to give rectification of the signal. In the early telephone and other audio amplifiers, it was realized that distortion in the form of spurious frequency components (that is, harmonics and intermodulation products), was produced, and so the effect was minimized by suitable choice of grid bias; and after about 1930, by compensatory arrangements such as 'push–pull' circuits in which two valves were used to balance out the even-order distortion, as shown in Fig. 46.11.

In radio-telephony it was necessary to modulate the amplitude of a carrier wave by the speech waveform; this was amplitude modulation (A.M.). A good way of doing this was to use the nonlinearity of the triode valve. In

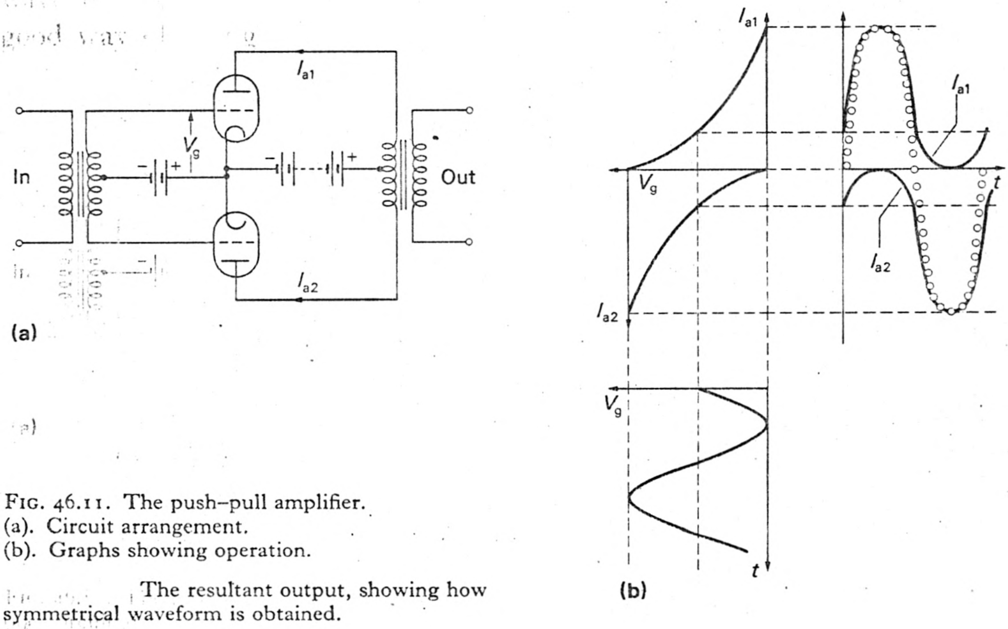


FIG. 46.11. The push–pull amplifier.
(a). Circuit arrangement.
(b). Graphs showing operation.

The resultant output, showing how symmetrical waveform is obtained.

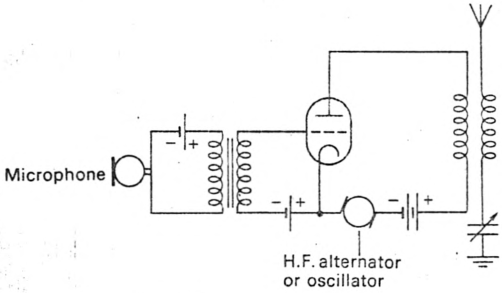


FIG. 46.12. The triode valve used as modulator in radio transmitter.

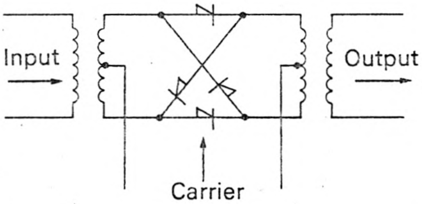


FIG. 46.13. A ring modulator using solid-state rectifiers.

experiments in the early 1900s a carbon microphone had been used directly in the aerial circuit; its varying resistance, which followed the variations of sound pressure, modulated the instantaneous transmitted power by varying the proportion of aerial power absorbed in the resistance of the microphone. Microphones could not, however, absorb large amounts of power. Consequently, when radio-telephony later came into service with transmitters of higher power, a triode valve was used as the modulating resistance; one arrangement is shown in Fig. 46.12. This relies on the fact that the internal resistance $\partial V_a/\partial I_a$ of the triode depends markedly on the grid voltage. Other early methods of modulation included the direct modulation of the amplitude of oscillation in a triode-valve oscillator by applying the speech signal to the grid of the valve. There was much distortion in these early modulators, but improved circuits were rapidly developed. Modulators using saturable iron-cored inductors were also used, but not widely.

In carrier telephone systems (originally on lines, but later also on radio links), and in single-sideband systems generally (Ch. 50), a different kind of modulator came into use in the 1930s. This was based on solid-state rectifiers, at that time of the copper-oxide type. In principle, thermionic diodes could have been used; later, from the 1940s onwards, germanium and silicon rectifiers were used. There were many different circuit arrangements, but one of the most widely used was the 'ring' modulator, shown in Fig. 46.13, so called because the rectifiers pointed round a ring. The operation of the circuit is most usefully described by assuming that each rectifier is a perfect switch, opening when the voltage across it is negative, closing when positive. The carrier-wave voltage is high with respect to the input signal voltage and so controls the 'switches', which thus reverse the polarity of the signal twice in each carrier cycle. The output thus comprises the sidebands of the modu-

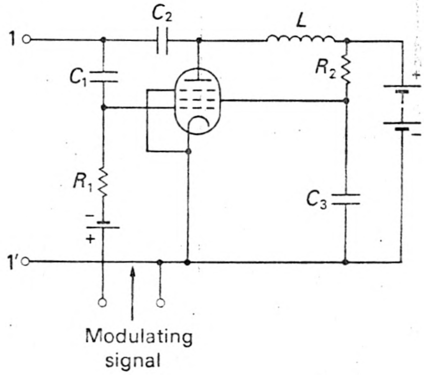


FIG. 46.14. A valve reactor circuit for frequency-modulation.

L = high frequency choke; C_2 = coupling capacitance; C_3 = smoothing capacitance; R_1 = very small compared with reactance of C_1 at carrier frequency; R_2 = fairly high resistance.

lated signal, but not the carrier itself. Rectifier modulators of this family have been used in millions in telephone systems and many other electronic systems.

Frequency modulation (F.M.) became an important process in the late 1930s, although as a concept it was much older [16]. Its principle was to carry the signal information by modulating, not the amplitude, but the frequency of the carrier. To effect this, one method was to connect a circuit called a 'valve reactor' (Fig. 46.14) across the tuned circuit of the carrier oscillator. This circuit produces an effective capacitance between terminals $1,1'$, which is made up of two components: C_1, plus another approximately equal to the product of R_1C_1 multiplied by the mutual conductance of the valve. The modulating signal, applied to the grid circuit with suitable bias, modulates the mutual conductance and hence the effective capacitance across $1,1'$. Since this is connected across the tuned circuit of the carrier oscillator, the frequency generated by the latter is modulated by the signal. At the receiver, the F.M. wave is applied to a circuit which has an amplitude response varying very sharply with frequency, thus converting the F.M. to A.M. which can then be detected in the normal way. Later, more sophisticated ways of forming and detecting F.M. signals were developed.

IV. CIRCUIT AND SYSTEM THEORY

The development of electronic and communication engineering has depended not only on the provision of suitable electronic devices and passive components but also, perhaps *a fortiore*, on the development of circuit (or 'network') theory, which towards the end of the first half of the twentieth century started to extend rapidly into system theory. While much, perhaps most, of the design work in communications up to the 1920s was done on an empirical basis, relying on a knowledge of the physical behaviour of the devices and components, yet there can be little doubt that the rapid progress of the 1930s and 1940s was due to the more general use of mathematical formulation in the conception and design of new circuits and systems.

Circuit and system theory is so abstruse and mathematical that its history can be presented properly only in terms which require a high degree of specialized knowledge. An excellent history in that class was given by V. Belevitch in 1962 [17]. Here we must be content with a history in broad terms with elementary examples. Four stages of development may be identified:

(i) The mathematical statement of the properties of individual circuit elements and of simple combinations of elements.
(ii) The formulation of the properties of complex circuit structures in which each simple combination of elements is represented by a mathematically-defined impedance or admittance function.
(iii) The use of this knowledge and special procedures to synthesize a complex circuit that will have a prescribed performance within stated tolerances.
(iv) The extension of these ideas to systems in which the various parts are complex circuits, each represented only by a mathematical function defining its behaviour as observable externally between one pair of terminals and another.

Simple combinations of elements. The basic passive circuit elements are resistance (R (ohms)), inductance (L (henries)), and capacitance (C (farads)), and a simple series combination of all three is shown in Fig. 46.15(a). The term 'passive' is used to distinguish these elements from 'active' elements such as valves, in which voltage and/or current sources operate notionally within the element (see below).

The differential equation for the *RLC* circuit of Fig. 46.15(a) was first set up by William Thomson (later Lord Kelvin) in 1853 [18]. He was concerned with the discharge of a capacitor (for example, a Leyden jar) through a conductor which had resistance and a property which he called 'electro-dynamic capacity', that is, inductance. In more specifically circuit terms, the equation was restated by J. Clerk Maxwell in 1868 [19]; in modern symbols:

$$E \sin \omega t = L\frac{\mathrm{d}i}{\mathrm{d}t} + Ri + \frac{1}{C}\int i\,\mathrm{d}t,$$

where i is the current and $E \sin \omega t$ the applied voltage. Maxwell pointed out that the current would be a maximum when $\omega L = 1/\omega C$. This knowledge was not put to much practical use, however, until work began on radio, with H. R. Hertz in 1887 (he called the effect 'resonance') and Oliver Lodge in 1890 (who called the effect 'syntony'). Electrical power engineers were also conscious of the effects of L and C about this time, although they were concerned with the much lower frequencies of a.c. power supplies.

As early as 1882, F. van Rysselberghe (Belgium) used a capacitance and inductance to separate a telephone and a telegraph channel working on the same wire [20], as shown in Fig. 46.15(b)—the forerunner of the high-pass and low-pass filter pair—and in 1883 he was using the circuit of Fig. 46.15(c) for the telegraph channel: a low-pass filter of the type normally associated with the 1920s. However, he did not examine this circuit mathematically.

The equation given above, and all calculations on more complex circuits, used the basic laws enunciated by G. R. Kirchhoff in 1845–7, that the voltage drops round a closed-loop circuit must sum to zero, and likewise the currents flowing into a junction of several conductors (that is, into a 'node'). Maxwell, in 1873, extended these laws to provide a means of solution for circuits with any number of interconnected nodes; his mesh and nodal equations, as they are now usually called, remain the fundamental method of circuit calculation.

Around 1890, O. Heaviside introduced the idea of an operational calculus for circuits; the use of vector (now called 'phasor') diagrams for the calculation of impedance and the application of the j operator (where mathematically $j = \sqrt{(-1)}$), were introduced in the 1890s. We thus see that the whole basis for circuit calculations was available before 1900.

When the thermionic triode valve came into general use in amplifiers, it

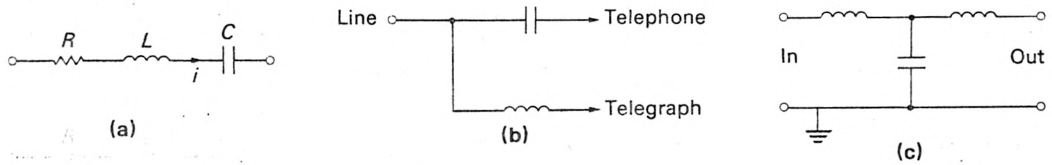
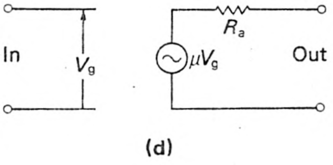


FIG. 46.15(a). A series connection of resistance, inductance, and capacitance.
(b). Van Rysselberghe's separation of high and low frequency channels, 1882.
(c). Van Rysselberghe's low-pass filter of 1883.
(d). Equivalent-circuit representation of a triode valve.
(For symbols see text)



was found useful to have the equivalent circuit shown in Fig. 46.15(d) to represent it in circuit calculations. Here μ is the amplification factor $\partial V_{\mathrm{a}}/\partial V_{\mathrm{g}}$, and R_{a} the internal resistance $\partial V_{\mathrm{a}}/\partial I_{\mathrm{a}}$. Grid current and inter-electrode capacitances could be allowed for by inserting extra elements. Since there is here a built-in source of electromotive force (e.m.f.), the valve is termed an 'active' circuit element.

Complex circuit structures. The beginning of the next stage of circuit theory dates from the work on loaded lines by G. A. Campbell and M. I. Pupin (U.S.A.) in about 1900 [21]. Calculations on cables and lines had, from the days of long telegraph cables, been based on distributed resistance and capacitance and, later, leakance and inductance. However, the introduction of loading coils at regular intervals in a line introduced a 'lumped' element into the system, and the idea of a recurrent circuit or network structure (Fig. 46.16(a)) developed. Here R, L, G, and C are the resistance, inductance, leakance, and capacitance of a line collected into a 'lumped' circuit to represent a given length of line. From the point of view of transmission between terminal pairs (or 'ports') 1,1′ and 2,2′, the 'unbalanced' tee network of Fig. 46.16(b) is equivalent; equally so is the pi network of Fig. 46.16(c) if the length represented is very small. From this developed the generalized networks of Figs. 46.16(d), (e), and (f), where Z_1, Z_2, and Z_3 are impedance functions expressed mathematically; equally well they could be admittance functions Y_1, Y_2, and Y_3.

Networks of these kinds (and of other kinds, too) can be connected in a chain, each two-port section having different responses if desired, so that specified overall transmission properties may be achieved. One very important early use was in 'electric wave filters'.

The idea of an electric wave filter as a formal network structure which will pass signals of some frequencies (in the 'pass-band') while giving a high loss to others (in the 'stop-band'), as indicated in Fig. 46.16(g), appears to have been due to G. A. Campbell in 1915 [22]. A formal system of such network structures giving a very large degree of control over their characteristics, based on the matching of so-called 'image' impedances at the junctions of the tee, pi, or L sections, was invented by O. J. Zobel (U.S.A.) and published in 1923 [23]. The Zobel system of filters was widely used for carrier telephone systems for many years. In practice, abrupt cut-off as shown in Fig. 46.16(g) was not realizable, and methods of optimizing pass-band and stop-band responses were introduced by W. Cauer (Germany) and others in the early

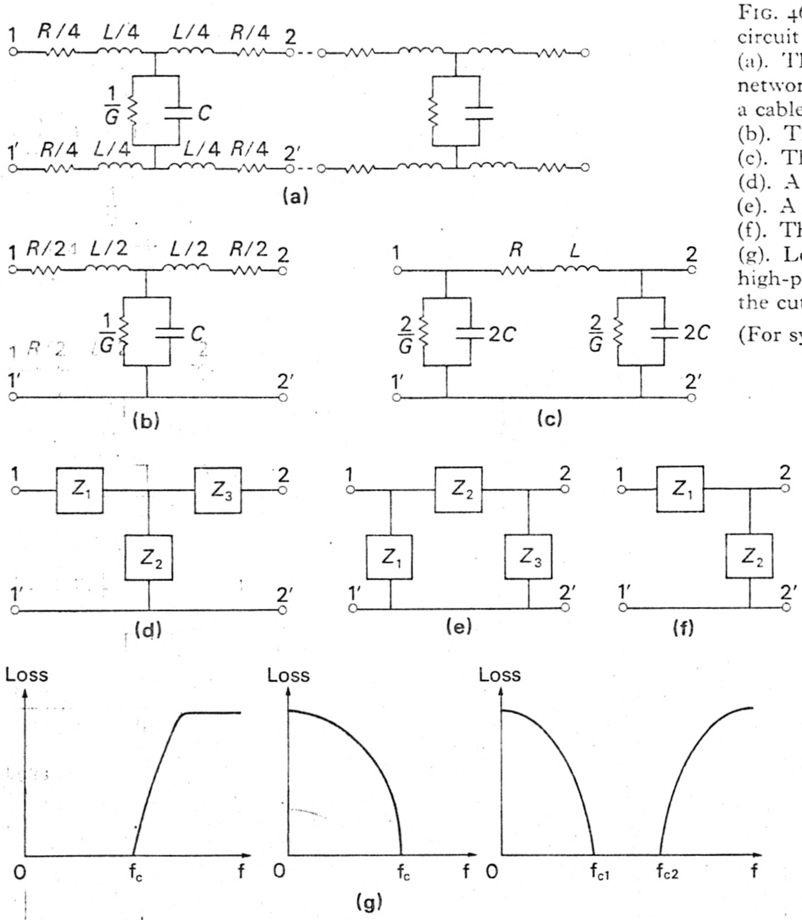


FIG. 46.16. The development of compl
circuit structures and filters.
(a). The recurrent 'ladder' circuit or network, derived from representation o a cable or line.
(b). The Tee or T network.
(c). The Pi or π network.
(d). A generalized Tee network.
(e). A generalized Pi network.
(f). The L network.
(g). Loss/frequency curves for low-pas high-pass, and band-pass filters (f_c is the cut-off frequency).
(For symbols see text)

1930s, leading to formal mathematical synthesis of networks to give a specified performance by S. Darlington (U.S.A.) in 1939 [24].

As electronic systems extended in frequency into the microwave region (centimetric wavelengths, frequencies of thousands of megahertz), the concept of lumped circuits had to be extended to wave concepts: waveguides, cavity resonators, etc.

With the introduction of pulse and waveform-transmission systems, for example, television, the determination and specification of performance in terms of amplitude or loss or gain versus frequency was not always adequate, and performance in terms of time-response had to be considered. The foundation for this had been laid in the work of J. A. Carson, E. A. Guillemin (both U.S.A.), and others during the 1920s and 1930s.

V. SOLID-STATE DEVICES

The thermionic valve, in its triode and later forms, had no serious competitor as an amplifying device until the junction transistor was developed; but it had several as a diode. Two such competitors, distinguished particularly by their geometries, were metal rectifiers and crystal detectors. Both exploited the unidirectional flow of current between a metal and a semiconductor (a class of material with electrical conductivity intermediate between metals and insulators). Metal rectifiers belonged to electrical engineering generally, for example, for battery charging, electroplating, and copper refining. Crystal detectors were initially specific to early radio receivers, a passing application that did not lead to elucidation of the rectifying property that was used. A later application, which valves could not satisfactorily meet, demanded fuller studies, from which sprung the transistor.

Metal rectifiers. L. O. Grondahl's work (1926) on the copper–cuprous oxide interface [25] led to reproducible and stable rectifiers, in contrast to earlier studies of metal to oxide (or sulphide) interfaces. Production, begun in the U.S.A., was simple though empirical. Discs of high-purity copper, about 1 mm thick, were oxidized at about 1000 °C. If this process, or any subsequent annealing, produced a surface film of cupric oxide (CuO), as well as the required layer of cuprous oxide (Cu_2O), it was removed chemically or mechanically. The second (counter) electrode, making good contact to the oxide layer, evolved from lead washers through colloidal graphite to sputtered metals. The direction of easy flow of current was from oxide to copper. Because current blocking in the other direction became unsatisfactory when the applied voltage exceeded about 10 V, many applications required multiple units, series-connected. Units of large area, for heavy current use, also incorporated cooling fins. Applications were as power rectifiers, modulators, and simple switches.

The copper oxide rectifier had a rival in the selenium type, developed a little later, though first studied much earlier by Fritts (1883). In production, a pure powdered form of selenium was spread thinly on plates of steel or aluminium (preferred for its better thermal conductivity, for larger units), and heat treated just below its melting point (217 °C), converting it to the required crystalline form and making good electrical contact with the plates. The rectifying interface was then prepared by spraying the exposed selenium surface with an alloy of low melting-point, for example, of bismuth and cadmium with lead or tin, serving also as the second electrode. Forming, by the passage of current in the 'non-easy' direction of flow (from the second electrode to the plate), much improved the performance. The maximum usable inverse voltages, though higher than for the copper-oxide type, were still insufficient to avoid the necessity for stacking for most applications.

Crystal detectors. A crystal detector was essentially a fine wire with pointed tip (called a cat's whisker) pressed flexibly, with a force of about 1 gram, on to the surface of a small crystal (or one in a polycrystalline piece) of a semiconductor. Several minerals (for example, galena) were used as suitable semiconductors for this purpose. The whisker acted as one electrode and a metallic clamp for the crystal as the other. Variations in the ratio of forward to reverse conductivity were considerable as the whisker tip was moved over the crystal surface. Units for early radio receivers allowed for such movement, permitting a search for a sensitive spot. Unfortunately, usage or shock could quickly spoil the performance, necessitating a further search.

A later application, in radar receivers of the Second World War and primarily for a 'mixer' (modulator) rather than a detector, demanded a much better controlled and reliable unit. Efforts concentrated on silicon with low levels of added impurity as the semiconductor, and wires of tungsten or molybdenum. Empirical searching gave way to (rather less) empirical tapping which could much improve performance, though lightly welding the wire tip to the silicon was sometimes practised. The maximum withstandable reverse voltage—before breakdown and permanent damage occurred—was low.

Because the element germanium resembles silicon in being tetravalent and crystallizing with the same (diamond) structure, attention was given to germanium point-contact diodes. Units with much higher peak inverse voltages became possible, allowing use in many valve switching and logic circuits such as were used in early electronic digital computers. With its lower melting-point, facilitating the preparation of single crystals, and a higher mobility for conduction electrons, germanium received more study in the late 1940s (and indeed the early 1950s) than silicon. The discovery of transistor action, launching a new era in electronics, was thus brought forward.

The transistor. Scientists had long speculated on the possibilities of adding a third electrode either to a solid-state rectifier to modulate the current–voltage relationship or to a piece of semiconductor to modulate its conductivity, with the object of achieving amplification or controlled switching.

Thus J. Lillienfeld had sought patents in 1925, 1927, and 1928 for three different structures. The most notable was the third, an insulated-gate field-effect device resembling the metal-oxide–silicon transistor (MOST) of the 1960s; copper sulphide was cited as the semiconductor with an aluminium foil as the gate electrode insulated from the semiconductor by having been anodized. But no demonstration was given; the necessary materials technology was lacking. Twenty years later, both the preparation of semiconductors and understanding of their properties, had much advanced with the work on silicon and germanium point-contact diodes. One fact, in particular, had been fully established. Semiconductors could conduct in one of two ways: either by 'free' electrons (electrons surplus to those needed to bind together the atoms in the crystal, and normally present in silicon and germanium if a pentavalent impurity has been added) when it is labelled *n*-type ('*n*' for negative), or by 'holes' (deficiencies in the binding electrons, normally present if a trivalent impurity has been added) when it is labelled *p*-type ('*p*' for positive). The holes do indeed respond to electric fields as if positively charged. Although the properties of the barrier to current flow at the contact of a metal and a semiconductor—which is lowered when a voltage of one polarity is applied and raised for the other polarity—were not wholly predictable, the possibility of efficient rectification at the junction between an *n* region and a *p* region integral in one crystal was emerging.

A series of experiments, backed by relevant theoretical studies, at the Bell Telephone Laboratories revealed and explained new findings. Thus J. Bardeen and W. H. Brattain (1947), using a second whisker to probe the surface of *n*-type germanium close to (within 0·1 mm of) the point of contact of a whisker which was carrying a forward current (I_E), discovered that when the second whisker was reverse biased it passed an abnormally high current (I_C) approximately proportional to I_E [26]. Indeed dI_C/dI_E could exceed unity, but—most important of all—because the first forward-biased contact presented low impedance and the second, reverse-biased, a high impedance, power gain could be obtained. Here, therefore, was a solid-state amplifying device, to be called a point-contact transistor (Fig. 46.17(a)). Its electronic behaviour was explained as follows. The first point contact (the emitter) injected holes into the *n*-type germanium (the base); under the influence of the voltage of the second contact (the collector) the majority of the holes travelled to that electrode undiminished despite the presence of the free electrons in the base. That presence, it had hitherto been assumed, would destroy holes instantaneously; that is, make good any deficiencies (holes)

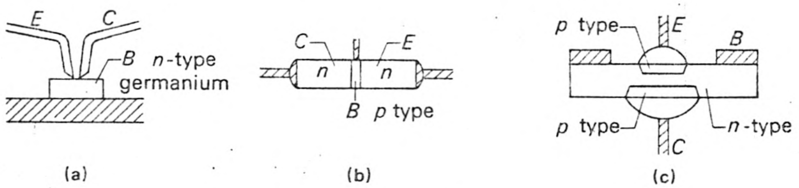


FIG. 46.17. The earliest transistors. (a). Point contact. (b). Germanium n–p–n grown junction. (c). Germanium p–n–p alloyed.

E = emitter; B = base; C = collector; hatched areas are metallic connections.

occurring in the binding (valency) electrons. The current gain at the collector turned out to be of less importance.

W. Shockley, encouraged by his colleagues' findings, then postulated a junction transistor, in which a piece of monocrystalline semiconductor comprised three regions (emitter, base, and collector) respectively of *n*-, *p*-, and *n*-types [27]. He predicted very promising electrical behaviour for it, involving electrons being injected from the *n*-type emitter into the *p*-type base, and diffusing to the collector. Efforts to fabricate the structure followed—a formidable task. Success came in 1950 (W. Shockley, M. Sparks, and G. K. Teal [28]) with a structure grown by adapting the Czochralski technique already in use for pulling monocrystalline ingots from a melt held close to its freezing-point. The conventional seed crystal was dipped into *n*-type molten germanium, but after a few millimetres of solid had been grown on, the melt was over-doped to *p*-type by the addition of a suitable impurity; when an additional length of up to 0·05 mm had been grown the melt was over-doped in the opposite direction, reconverting it to *n*-type, and a further short length grown. The ingot was cut lengthwise into bars, each containing the three regions (*n*-type collector, *p*-type base, and *n*-type emitter), and wire connections made to each region of a bar (Fig. 46.17(b)). The performance of the grown junction transistor confirmed Shockley's predictions completely. The device proved reproducible, and offered a power gain higher than that of the point-contact transistor which quickly lost favour, the more so when R. R. Law [29] and, independently, J. S. Saby (1952) [30] developed a method of making junction transistors well suited to large-scale production. They started with small wafers of *n*-type germanium of thickness about 0·1 mm, and alloyed pairs of small pellets of indium—a trivalent element of low melting-point—to the pairs of faces. Hence *p-n-p* structures resulted (see Fig. 46.17(c)). Contact-making was straightforward.

Although the good performance of early junction transistors was largely confined to frequencies below 100 kHz, their electrical characterization was simple, presenting few problems to circuit designers. They could be used with input signals applied to the emitter and output taken from the collector (with grounded base), or have input to the base with output from collector (with grounded emitter) or from emitter (with grounded collector). Their potential as low-power switches and as logic elements was quickly seized upon for many applications.

Further developments concentrated on reducing the base width—the major factor restricting bandwidth—from 0·025 mm to 0·001 mm and less, and on methods of processing silicon to make similar devices, thereby to benefit from the smaller leakage currents inherent in silicon devices and their ability to perform better than germanium counterparts at higher ambient temperatures. But the resulting predominance of silicon stemmed as much from the excellent electrical, and chemically passive properties of its thermally grown oxide (a vitreous silica), permitting the mass production of wideband devices by a planar technology. All that, however, belongs to the second half of the twentieth century.

VI. PASSIVE COMPONENTS

The most common passive components used in electronic circuits were resistors, capacitors, inductors, and transformers. Electrical engineering had made available examples of all four types by 1900. Electronic engineering began by using the same general designs, but from around 1920 made demands for specific improvements, to which research and development responded, generally successfully.

Resistors. The resistance (R) of a conductor depends on its geometry and the resistivity (ρ) of its material. For a wire of length L and cross-sectional area A, $R = \rho L/A$. Electronic engineering requires resistors having R in the range from about 10 to 10^6 ohms. They must also have the ability to dissipate the heat produced by the current flowing, a small temperature coefficient of resistance, and specified tolerances of resistance.

By 1900 both wire and carbon-composition resistors were available. For the former, pure metals had generally too low a resistivity and too high a temperature coefficient (typically 4000 p.p.m./°C). Alloys were developed with more suitable values. German silver (copper, zinc, and nickel) and Constantin (copper and nickel) had resistivities some twenty times that of copper, and coefficients around one-twentieth. Manganin (copper, manganese, and nickel) had a coefficient less than 20 p.p.m./°C at 20 °C. The wire was wound on a former, but even the best designs introduced enough self-capacitance or self-inductance to upset performance at frequencies above about 1 MHz. Carbon-composition resistors greatly extended the frequency range, enabled high values of resistance (for example, 1 megohm) to be easily produced, and were cheap; but manufacture to close tolerances proved impracticable and their temperature coefficient was large and negative, typically −1000 p.p.m./°C. By 1930 they were in large-scale production, with wire types restricted to special uses.

Carbon-film resistors, developed in Germany in 1928, represented a major advance, and later found much use. The film was deposited on a ceramic rod by the pyrolysis of a hydrocarbon and a thin spiral cut in it to make what was effectively a long narrow strip of film. Tolerances on nominal values could be finer than ±1 per cent; the temperature coefficient was typically −250 p.p.m./°C.

Thermistors (that is, resistors with very pronounced sensitivity to temperature, making them suitable as temperature sensors, regulating elements, etc.) were developed before 1950; they used oxides of manganese and nickel.

Small-sized variable resistors and potentiometers were in continuous demand, for example as volume controls of radio receivers. Wire and carbon-film types, on circular formers with a sliding contact attached to a rotatable central spindle, generally met the demand.

Capacitors. Circuit needs for discrete capacitance (called capacity until about 1950) were met by capacitors (previously called condensers), deriving from the Leyden jar (1745). The late nineteenth century saw more compact forms, but all comprised essentially a pair of conducting plates, parallel to one another and separated by an insulating gap filled with gaseous, liquid, or, more usually, solid material possessing suitable dielectric properties. The capacitance (C) is approximately proportional to the area of the plates, inversely proportional to their separation, and proportional to the permittivity (k) of the dielectric (called the specific inductive capacity in the nineteenth century and later the dielectric constant).

The first dielectric of industrial importance was wax-impregnated paper ($k = 1{\cdot}2$ for paper, but 3 for waxed paper). The capacitor would be made up of a strip of tinfoil, two sheets of the paper, another strip of tinfoil, and two more sheets of paper, all rolled up together to make a very compact component which could have capacitances of up to 10 microfarads (μF) and be able to withstand direct voltages of 100 V or more.

Dielectric loss in a capacitor introduces a resistive component in its impedance. Although paper capacitors were acceptable for most uses, lower loss and smaller size could be obtained—at greater cost—with mica as the dielectric ($k \simeq 6$). Other materials also found applications, culminating in the 1940s with rutile (TiO_2), and some titanates, having $k \simeq 100$.

Where large capacitance (for example, 25 μF) was required in a small volume, without the need for precision or low loss (for example, in smoothing power supplies), dry aluminium electrolytic capacitors became important. Evolving from an earlier wet type (demonstrated in 1892), they used a strip of absorbent material saturated with (usually) a borate, sandwiched between an anodically oxidized aluminium foil and a plain foil. The voltage across this capacitor had to be of stated polarity.

Variable capacitors, for example, for tuning radio receivers, almost always comprised a set of fixed vanes with rotatable cam-shaped vanes interleaved, with air as the dielectric (Fig. 46.18).

Inductors and transformers. An inductor is essentially a closely wound coil of copper wire, with inductance (L) proportional both to the area and the square

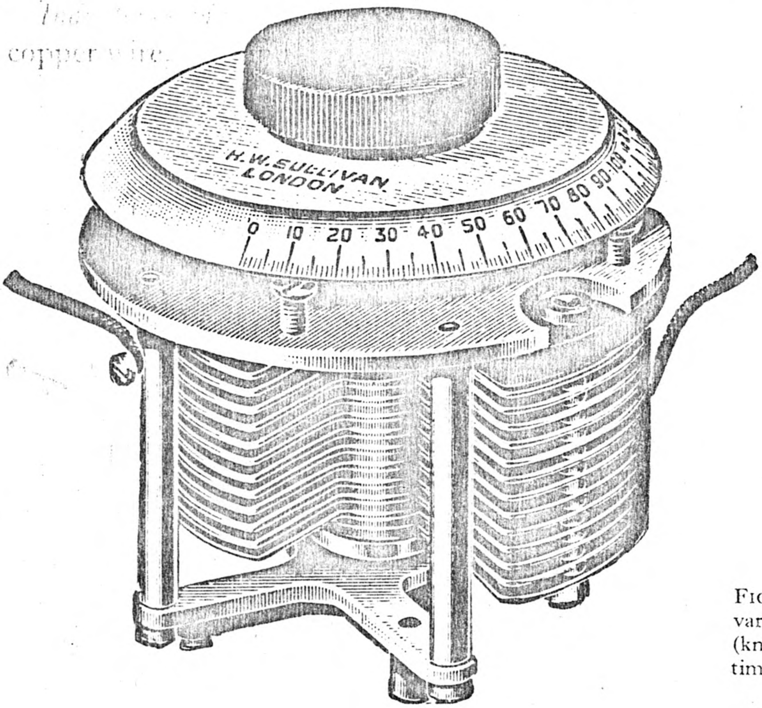


Fig. 46.18. A continuously variable air capacitor, *c.* 1920 (known as a condenser at that time).

of the number of turns. When the coil is immersed in a material of magnetic permeability μ, L increases μ times. In practice a magnetic core is used, filling the cross-section of the coil and forming a loop of its own, achieving much the same increase. Air-cored coils found some favour in early radio receivers, but more latterly only at higher frequencies.

A transformer closely resembles an inductor in construction, except in having more than one coil. It offers transformations of impedance, voltage, and current between the circuits connected to its separate coils.

Coil winding changed little effectively between 1900 and 1950, beyond being mechanized; inductor and transformer development consequently centred on improved core materials.

In 1900 the electricity supply industry was already seeking materials better than soft iron for its transformers. R. A. Hadfield had led the way, with alloys of iron with a few per cent of silicon; production of bulk material began in 1903 followed a few years later by laminations which, stacked with thin insulation between, much reduced the power losses due to the eddy currents induced in the core.

Electronics used this family of alloys ($\mu \simeq 5000$) for its power transformers and smoothing inductors (familiarly called 'chokes'), but sought still higher values of μ for components handling signal frequencies. G. W. Elmen (1917) showed the promise of alloys of nickel (80 per cent) with iron (20 per cent) [31]. Several families were developed, with small additions of other metals; two, mumetal and permalloy, offered values of μ up to 100 000. They found wide use. But, even with lamination thickness reduced to 0·025 mm, their eddy-current losses were unacceptable at the higher radio frequencies. Attention turned to the use of insulated particles (the dimensions of which were to be measured in microns) of iron or nickel-iron mixtures, compressed together and fired, to form 'dust' cores. Despite their considerable sacrifice of permeability, they found use from about 1920. Then, around 1935, the magnetic properties of ferrites—first reported on by S. Hilpert in 1909 [32]—were re-examined with marked success, leading to production and many applications. Ferrites have the generic formula $MO.Fe_2O_3$, where M is a divalent element, as in nickel ferrite, $NiFe_2O_4$, or a mixture of two such elements, as in zinc manganese ferrite, $ZnMnFe_4O_8$. They are insulators, thereby minimizing eddy current loss, and the various compositions offer a range of values of μ up to about 1000. Both dust and ferrite cores were frequently offered in toroidal shape, necessitating coil-winding directly on to them, instead of on to bobbins or formers into which earlier cores fitted.

REFERENCES

[1] Guthrie, F. *Philosophical Magazine* (4th series) **46**, 257 (1873).
[2] anon. *Engineering*, **38**, 553 (1884).
[3] Thomson, J. J. *Philosophical Magazine* (5th series), **44**, 293 (1897).
[4] Richardson, O. W. *Proceedings of the Cambridge Philosophical Society*, **11**, 286 (1901).
[5] Wehnelt, A. *Annalen der Physik*, **14**, 425 (1904).
[6] Fleming, J. A. British Patent 24 850; 1904.
[7] de Forest, L. *Transactions of the American Institute of Electrical Engineers*, **25**, 735 (1906).
[8] Armstrong, E. H. *Proceedings of the Institute of Radio Engineers*, **3**, 215 (1915).
[9] Tucker, D. G. *Radio and Electronic Engineer*, **42**, 69 (1972).
[10] Hazeltine, L. A. *Proceedings of the Radio Club of America*, **2**, No. 8 (1923).
[11] Schottky, W. *Archiv für Elektrotechnik*, **8**, 1 and 299 (1919).
[12] Holst, G. and Tellegen, B. D. H. U.S. Patent 1 945 040; filed 1927.
[13] Zworykin, V. K., Morton, G. A., and Walter, L. *Proceedings of the Institute of Radio Engineers*, **24**, 251 (1936).
[14] Zworykin, V. K. *Proceedings of the Institute of Radio Engineers*, **22**, 16 (1934).
[15] Braun, F. *Annalen der Physik*, **60**, 552 (1897).
[16] Tucker, D. G. *Radio and Electronic Engineer*, **40**, 33 (1970).
[17] Belevitch, V. *Proceedings of the Institute of Radio Engineers*, **50**, 848 (1962).
[18] Thomson, W. *Philosophical Magazine*, **5**, 393 (1853).
[19] Maxwell, J. C. *Philosophical Magazine*, **35**, 360 (1868).
[20] van Rysselberghe, F. British Patents 2466, 1882; and 5503, 1883.
[21] Brittain, J. E. *Technology and Culture*, **11**, 36 (1970).
[22] Campbell, G. A. U.S. Patents 1 227 113 and 1 227 114; filed 1915.
[23] Zobel, O. J. *Bell System Technical Journal*, **2**, 1 (1923).
[24] Darlington, S. *Journal of Mathematics and Physics*, **18**, 257 (1939).
[25] Grondahl, L. O. *Physical Review*, **27**, 813 (1926).
[26] Bardeen, J. and Brattain, W. H. *Physical Review*, **74**, 230 (1948).
[27] Shockley, W. *Bell System Technical Journal*, **28**, 435 (1949).
[28] Shockley, W., Sparks, M., and Teal, G. K. *Physical Review*, **83**, 151 (1951).
[29] Law, R. R. *Proceedings of the Institute of Radio Engineers*, **40**, 1352 (1952).
[30] Saby, J. S. *Proceedings of the Institute of Radio Engineers*, **40**, 1358 (1952).
[31] Elmen, G. W. Canadian Patent 180 539; filed 1916.
[32] Snoek, J. L. *Philips Technical Review*, **8**, 353 (1946).

BIBLIOGRAPHY

Cruft Laboratory, Harvard University. *Electronic circuits and tubes.* McGraw–Hill, New York (1947).

Fleming, J. A. *The thermionic valve.* Wireless Press, London (1919).

Guillemin, E. A. *Communication networks.* Wiley, New York (1935).

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

Latham, R., King, A. H., and Rushforth, L. *The magnetron.* Chapman and Hall, London (1952).

Lubszynski, H. G. Photocells for the visible and ultra-violet, in *Electronics* (ed. B. Lovell), pp. 53–95. Pilot Press, London (1947).

MAXWELL, J. C. *Electricity and magnetism.* Oxford University Press (1873).
REICH, H. J. *Theory and applications of electron tubes.* McGraw-Hill, New York (1939).
ROUND, H. J. *The shielded four-electrode valve.* Cassell and Bernard Jones, London (1927).
SAY, M. G. (ed.). *Cathode-ray tubes.* Newnes, London (1954).
SCOTT-TAGGART, J. *Wireless vacuum tubes.* Radio Press, London (1922).
SHIVE, J. N. *The properties, physics and design of semiconductor devices.* Van Nostrand, New York (1959).
SHOCKLEY, W. *Electrons and holes in semiconductors.* Van Nostrand, Princeton (1950).
TERMAN, F. E. *Radio engineers' handbook.* McGraw-Hill, New York (1943).
TORREY, H. C. and WHITMER, C. A. *Crystal rectifiers.* McGraw-Hill, New York (1948).
TUCKER, D. G. *Modulators and frequency changers.* Macdonald, London (1953).
TURNER, L. B. *Wireless telegraphy and telephony.* Cambridge (1921).
ZWORYKIN, V. K. and RAMBERG, E. G. *Photo-electricity.* Wiley, New York (1930).