1 Overview

1.1 Introduction

Sources of power in the radiofrequency, microwave, and millimetre wave regions of the electromagnetic spectrum are essential components in a wide range of systems for telecommunications, broadcasting, remote sensing, and processing of materials[1]. Current research is extending the frequency range into the sub-millimetre region. These sources employ either vacuum electronic, or solid state, technologies. At the higher frequencies and power levels, vacuum electronic devices (tubes[2]) are the only sources available (see Section 1.2).

The purpose of this book is to provide a comprehensive introduction to the theory and conceptual design of the types of tubes which are of continuing importance. The design and operation of vacuum tubes requires knowledge and skills drawn both from electrical and electronic engineering and from physics. The treatment here is intended to be accessible to those whose training has been in either discipline. The use of advanced mathematics has been avoided as far as possible with considerable emphasis on the use of simple numerical methods. The book is designed to be a reference text for designers and users of vacuum tubes, and a textbook for people who are new to the field.

This is a mature field in which much has been published since its first beginnings in 1904 [1]. The sources cited here are those which have been used as the basis for the book. They are believed to comprise most of the most important sources in the field and the reader is invited to consult them for further information. References have been included to sources that provide additional information on many of the topics, but no attempt has been made to provide a comprehensive bibliography of the subject and longer lists of references are to be found elsewhere [2]. A further aim of this book is to provide the reader with the background necessary to read with understanding other papers in the field.

In any book it is necessary to make choices about what should be included and what excluded. The subjects covered are those the author believes to be important for the practical business of designing and using vacuum tubes. Because the focus

[1] For convenience, the term RF is used throughout this book to refer to all frequencies in the range 30 kHz to 300 GHz.

[2] We shall call these devices (vacuum) tubes because that terminology is familiar and concise.

is on theory and conceptual design very little has been said about the technology of tube construction, which is well treated elsewhere [3, 4]. Similarly, little has been said specifically about vacuum tubes for use at sub-millimetre wavelengths since they employ the same principles as those at lower frequencies, and most of the challenges are in the technology of their construction.

This chapter provides an overview of the subject of the book. The next section compares vacuum tubes with solid-state devices to show how the technologies are complementary. Section 1.3 provides an overview of the physical principles on which vacuum tubes are based and definitions of the key terms used to describe their performance. A tube converts the DC power in the initial electron stream into RF output power by interaction with electromagnetic structures. These structures support standing (resonant), or travelling, electromagnetic waves. Coupled-mode theory is introduced as a valuable conceptual tool for understanding the interactions between electron streams and travelling electromagnetic waves. The section concludes with a classification of the principal types of vacuum tube based on the preceding discussion. The principal applications of vacuum tubes are reviewed in Section 1.4 together with some of the factors which govern the availability of tubes of different types. That leads, in Section 1.5 to consideration of the communication between the designers and users of tubes in the form of a Statement of Requirements. This statement specifies both the electrical performance required, and the factors which constrain the design. Many tubes are required to amplify modulated carrier signals whose properties are not normally familiar to people whose primary discipline is Physics. An introduction to analogue and digital modulation, noise and multiplexing is provided in Section 1.6. Finally, Section 1.7 considers some of the principles of the engineering design of tubes including dimensional analysis and scaling, and the use of computer modelling.

The remainder of the book comprises four sections:

i) Chapters 2–4 deal with the properties of the passive electromagnetic components employed in vacuum tubes.
ii) Chapters 5–11 are concerned with aspects of electron dynamics in vacuum that are employed in tubes in a variety of ways. Chapter 10 also includes a discussion of methods of cooling.
iii) Chapters 12–17 show how the fundamental principles introduced earlier in the book are applied to specific types of tube, and their conceptual design.
iv) Chapters 18–20 provide an introduction to some technological issues which are common to most types of tube and their successful use in systems.

1.2 Vacuum Electronic and Solid-State Technologies

The characteristic size of any active RF device is determined by the distance travelled by the charge carriers in one RF cycle. Thus the size of a device decreases as the frequency increases and as the velocity of the charge carriers decreases. The application of any electronic technology is limited, ultimately, by temperature since the operation of a device generates heat. Hence, the maximum continuous, or average, RF power which can be generated by a single device is determined by the power dissipation within it. At low power levels semiconductor devices have the advantages of small size, and low voltage operation. But these become disadvantages at high power levels because the current passing through the device is high. Thus there are large conduction losses, generating heat, within a small volume. They can be reduced, to some extent, by operating the device as a switch rather than in its active mode.

Transistors can currently deliver around 100 W of continuous power or 1 kW of pulsed power at frequencies in the region of 1 GHz [5–7]. Further developments, including transistors using diamond as a semiconducting material, may increase the power to several kW [8, 9]. High power amplifiers can be made by operating many transistors in parallel but the penalties of increased complexity set limits to this. The power combining can take place in space, as in active phased-array radar, where average powers up to tens of kilowatts can be achieved [10]. Alternatively, a power combining network may be used as in the 190 kW, 352 MHz, amplifier at Synchrotron SOLEIL [11]. This amplifier combines the power from four 50 kW towers each containing twenty 2.5 kW units. A unit comprises eight 315 W modules each having a pair of transistors operated in push-pull. This arrangement means that the loss of power from the failure of an individual transistor is small, and the degradation of the amplifier from this cause is gradual. However, the power output is still almost an order of magnitude less than that achieved by vacuum tubes at the same frequency, and the frequency is at the lower end of the range for which vacuum tubes have been developed. The use of low DC voltages reduces problems of reliability caused by voltage breakdown. But the consequent need for high currents leads to DC losses in the connecting bus-bars. A review of high power semiconductor RF power technology is given in [12].

Vacuum tubes, in contrast, operate at high voltages and low currents. The charge carriers have a much higher velocity than in a semiconducting material, and they are not subject to energy loss through collisions as they pass through it. Thus the active volume can be large with only RF losses within it. In many cases, the greater part of the heat generated is dissipated on electrodes which are separate from the active region, and whose size can be increased to reduce the power density. There is a common misconception that vacuum tubes are fragile, short-lived, unreliable, and inefficient. In fact, modern vacuum devices are mechanically robust and able to survive short-term electrical overloads without damage. They have demonstrated outstanding reliability and lifetimes in the very demanding environment of space. Vacuum tube amplifiers can have maximum conversion efficiencies of up to 70%, and 90% has been achieved in oscillators. However, the most reliable performance of any technology is achieved by allowing generous design margins and operating well within their limits. Vacuum tubes can be operated in parallel, in the same way as semiconductor devices, but the number of parallel devices is usually quite small [13, 14]. An exception to this is in phased-array radar where the output from a larger number of microwave power modules (see Section 20.1) is combined in space [15].

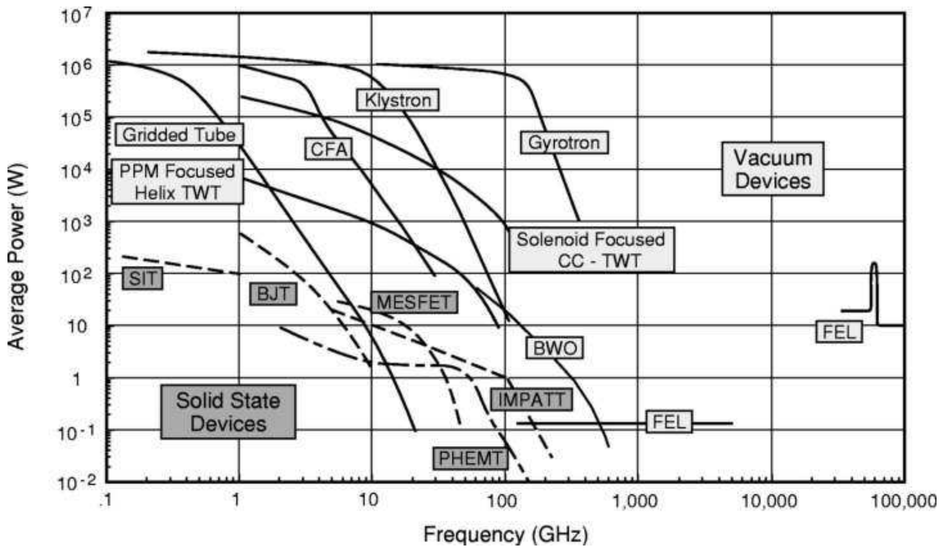

Figure 1.1: Comparison of the performance of single vacuum electronic and solid state RF power devices
(copyright 1999, IEEE, reproduced, with permission, from [16]).

Figure 1.1 shows a comparison between the performances of individual vacuum and semiconductor RF power devices [16]. There have been advances in both technologies since that figure was drawn, but the general picture remains valid. A more recent, but less detailed, figure is to be found in [17]. For further discussion of the relative merits of vacuum tube and solid-state RF power amplifiers see [18, 19]

1.3 Principles of Operation

The tubes which are the subject of this book are all power amplifiers and oscillators. For our present purpose it is sufficient to regard oscillators as amplifiers in which the RF input is provided by internal feedback. All vacuum tube amplifiers can be understood in terms of the block diagram shown in Figure 1.2. A uniform stream of electrons is emitted into the vacuum from the electron source and modulated by the RF input voltage[3]. Radiofrequency energy is extracted from the modulated stream of electrons, and their remaining energy is dissipated as heat on a collecting electrode. The arrows show the direction of motion of the electrons; the conventional current is, of course, in the opposite direction. The functions of the blocks may be combined in various ways in different devices but the overall process is essentially the same. The basic RF performance of an amplifier is defined in terms of its gain, output power, efficiency, and instantaneous or tuneable bandwidth.

[3] The word 'stream' is used here as being more general than the term electron beam which is used for specific types of tube.

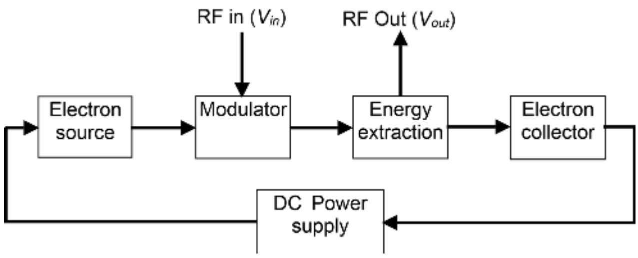

Figure 1.2: Block diagram of a vacuum tube amplifier.

1.3.1 Geometry

The majority of practical vacuum tubes have geometries which are cylindrically symmetrical, or close approximations to it. The flow of the DC current is radial or axial, and driven by a static electric field in the same direction (see Chapters 5, 8, and 9).

1.3.2 Electron Dynamics

The voltages employed in vacuum tubes are high enough, in many cases, for the velocities of the electrons to be at least mildly relativistic. The kinetic energy of a relativistic electron is [20]:

$$T = mc^2 - m_0c^2, \tag{1.1}$$

where m_0c^2 is the rest energy of the electron and the relativistic mass is

$$m = \frac{m_0}{\sqrt{1-u^2/c^2}} = \gamma m_0, \tag{1.2}$$

where u is the velocity of the electron and c the velocity of light. If an electron starts from rest at the cathode then its velocity at a point where the potential is V, relative to the cathode,[4] is found by using the principle of conservation of energy:

$$eV = m_0c^2\left(\frac{1}{\sqrt{1-u^2/c^2}} - 1\right). \tag{1.3}$$

This equation can be rearranged as

$$u = c\left[1 - \frac{1}{\left[1+\left(V/V_R\right)\right]^2}\right]^{\frac{1}{2}}, \tag{1.4}$$

[4] In this book voltages are referred to the cathode unless otherwise stated. In practice the tube body is normally at earth potential and the cathode potential is negative with respect to it.

where $V_R = m_0c^2/e = 511\,\text{kV}$ is the rest energy of the electron expressed in electron volts. When $V/V_R \ll 1$ the fraction can be expanded by the Binomial Theorem to give the approximate expression:

$$u = c\sqrt{\frac{2V}{V_R}}. \tag{1.5}$$

Figure 1.3 shows a comparison between the velocities calculated using the exact and approximate equations. The error in the approximate velocity is 1% at 7 kV so that the exact formula should be used for voltages higher than this.

The force acting on an electron is equal to the rate of change of momentum

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{u}) = m\frac{d\mathbf{u}}{dt} + \mathbf{u}\frac{dm}{dt}. \tag{1.6}$$

If the force acts in the direction of the motion of the electrons then

$$F = \frac{m_0}{\left(1-u^2/c^2\right)^{3/2}} \cdot a = \gamma^3 m_0 a, \tag{1.7}$$

where a is the acceleration and $\gamma^3 m_0$ is sometimes called the *longitudinal mass*. When the force acts at right angles to the direction of motion then

$$F = \frac{m_0}{\left(1-u^2/c^2\right)^{1/2}} \cdot a = \gamma m_0 a, \tag{1.8}$$

where γm_0 is the *transverse mass*. If the longitudinal velocity is approximately constant and much greater than the transverse velocity then it is possible to use the classical equations of motion with relativistic corrections to the mass of the electron.

The DC electron current in a tube may be collimated by a magnetic field which is either parallel to the direction of the DC current or perpendicular to it (see Chapters 7 and 8). Tubes in which the field is parallel to the current are described as 'Type O' and those in which it is perpendicular as 'Type M'. The current may also be collimated by a static electric field but this is rather rare.

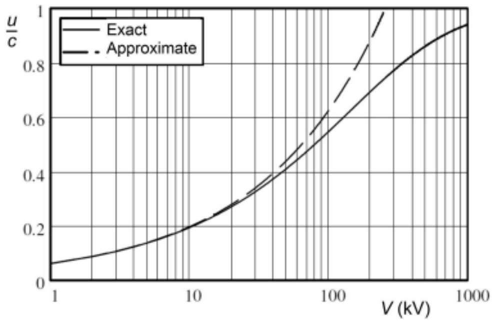

Figure 1.3: Comparison between exact and approximate electron velocities as a function of the accelerating potential.

1.3.3 Modulation of the Electron Current

For the present we will assume that the input RF voltage is purely sinusoidal with amplitude V_1 and frequency ω. Practical input signals are discussed in Section 1.6. It can be shown that the three possible methods of modulation of the current by the RF input voltage are [21]:

- Emission density modulation in which the current emitted from the source is varied.
- Deflection and concentration modulation in which the electrons are deflected sideways.
- Transit-time modulation in which the electron velocity is varied.

The effect of any of these methods of modulation, or combinations of them, is to produce a bunched stream of electrons whose current varies with time at the frequency of the input signal. Because the process is non-linear the time variation of the current is not sinusoidal but can be represented by a Fourier series. The amplitudes and phases of the harmonic components depend upon the amplitude of the input voltage, and on the distance from the source. The modulation may be either in the same direction as the DC current flow, or normal to it. Thus the current at some point z can be written

$$I(z,t) = I_0 + \sum_{n=1}^{\infty} I_n(V_{in}, z)\exp(jn\omega t), \tag{1.9}$$

where I_0 is the current in the unmodulated stream, $I_n(V_{in}, z)$ are complex amplitudes, and modulation in the axial direction has been assumed. The DC and time-varying parts of the current may be in different directions in space. We note that the real current cannot be negative. Its maximum value, relative to I_0, is determined by the process of modulation, or by the maximum current which can be drawn from the source. The ratio I_0/I_0 cannot exceed 2.0 (see sections 11.8.4 and 13.3.4).

1.3.4 Amplification, Gain, and Linearity

The RF output voltage is obtained by passing the modulated stream through a region at an effective position z_2 where energy is removed from the electron bunches by an RF electric field. This can be represented by an impedance $Z_n(V_{in})$ which depends on frequency and on the magnitude of the input signal. Thus the output voltage is

$$V_{out}(t) = \mathrm{Re}\left\{\sum_{n=1}^{\infty} I_n(V_{in}, z_2) Z_n(V_{in})\exp(jn\omega t)\right\}, \tag{1.10}$$

where, for simplicity we assume that the characteristic impedances of the input and output waveguides are the same. The impedance Z_n is effectively zero above some value of the harmonic number n determined by the nature of the output section. Thus the number of harmonics at which there is appreciable output power is small and, in some cases, limited to the fundamental. If the output power is to be radiated it may be necessary to filter out the harmonics to comply with the regulations determining the bandwidth available to the system, as specified by international agreement and national regulations [22, 23]. The transfer characteristic of the amplifier at the frequency ω is then

$$V_{out}(\omega t) = A(V_{in})\cos(\omega t + \Phi(V_{in})), \tag{1.11}$$

where $A(V_{in})$ is the AM/AM (amplitude modulation) characteristic which is usually plotted on decibel scales and $\Phi(V_{in})$ is the AM/PM (phase modulation) characteristic of the amplifier.

The details of the transfer characteristics depend upon the type of amplifier but many of them show the same features. Figure 1.4 shows, as an example, a typical AM/AM curve of a travelling-wave tube (TWT). The output power is proportional to the input power at low drive levels (slope 1 dB/ dB) but reaches saturation as the input is increased. In the figure the input and output powers have been normalised to their values at saturation. The gain of the amplifier in decibels is given by

$$G_{dB} = 10\log\left|\frac{P_{out}}{P_{in}}\right|, \tag{1.12}$$

where P_{in} and P_{out} are, respectively, the input and output power. Since we have assumed that the input and output waveguides have the same characteristic impedances we can write

$$G_{dB} = 20\log\left(\frac{A(V_{in})}{V_{in}}\right). \tag{1.13}$$

The difference between the linear (small-signal) and the saturated gain is the *gain compression* which may be used as a measure of non-linearity. It is common for a tube to be described in terms of its saturated output power, but the output power available under normal operating conditions may be less than this. For example the tubes used in particle accelerators are normally operated 'backed-off' from saturation to provide a control margin for the operation of the accelerator. In a TWT there is normally some output at second, and higher, harmonic frequencies. This also depends on the input drive level as shown in Figure 1.4. At low drive levels the second harmonic output is proportional to the square of the input power, giving a slope of 2 dB/dB. The maximum second harmonic power does not necessarily occur at the drive level that saturates the fundamental. The intersection between the projections of the linear parts of the fundamental and second harmonic curves, known as the *second-order intercept point*, is a measure of the second-order distortion of the amplifier.

The corresponding AM/PM curve for a TWT is shown in Figure 1.5. The phase is plotted relative to the phase at low drive levels, and the input power is normalised to the input power at saturation. The phase of the output signal relative to the input is constant at low drive levels but changes as the drive level is increased. Both the amplitude and phase of the output depend on frequency, and on the operating

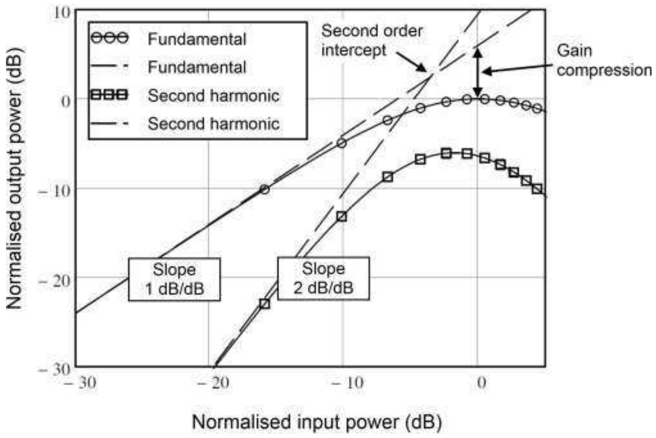

Figure 1.4: Typical curves of fundamental and second harmonic output power of an RF amplifier plotted against the input power, normalised to saturation.

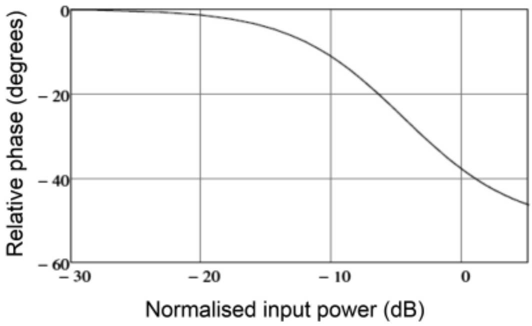

Figure 1.5: Typical curve of the phase of the output voltage of an RF amplifier against input power normalised to saturation.

conditions of the amplifier, including the voltages applied and the external RF matches. Thus any ripple in the voltages applied may result in amplitude and phase modulation of the output signal. In an oscillator, voltage ripple may also produce frequency modulation.

1.3.5 Power Output and Efficiency

An important consideration for many applications of tubes is the efficiency with which the DC input power is converted into useful RF power. The principle of conservation of energy requires that, in the steady state, the total input and output powers must balance, that is

$$P_{RF\ in} + P_{DC\ in} = P_{RF\ out} + P_{Heat}. \tag{1.14}$$

The DC input power includes the power in the electron stream, the cathode heater, and electromagnets. If the tube has a depressed collector (see Section 10.3) the

DC power is reduced by the power recovered by the collector. When a comparison is made between alternative tube types, or between vacuum tube and solid-state amplifiers, then the DC input power should be specified at the input to the power supply to allow for losses in it. Any power required for cooling fans or pumps must also be included. The RF output power comprises power at the fundamental frequency and its harmonics. Heat is generated by the impact of electrons on the collector and the tube body and by RF losses in the tube body, connecting waveguides and windows.

A number of different definitions of efficiency are in use and it is important to distinguish between them. The *overall efficiency* is defined here as the ratio of the fundamental RF output power (P_2) to the total input power to the tube

$$\eta_O = \frac{P_2}{P_{DC} + P_{RF\,in}}. \tag{1.15}$$

If the gain of the amplifier is high, the RF input power is much smaller than the DC input power so that approximately

$$\eta_O \approx \frac{P_2}{P_{DC}}. \tag{1.16}$$

The heater and electromagnet powers are typically much smaller than the stream power in continuous wave (CW) tubes but they can be comparable with the stream power in pulsed tubes. The *power added efficiency*, sometimes used when the gain is small, is

$$\eta_A = \frac{P_2 - P_{RF\,in}}{P_{DC}}. \tag{1.17}$$

This efficiency is effectively identical to that in (1.16) if the gain is 20 dB or more. The efficiency of tubes with depressed collectors is discussed in Section 10.3. The *RF efficiency* is defined here as the ratio of the useful RF output power to the power input to the electron stream (less any recovered by the collector) plus the RF input power.

$$\eta_{rf} = \frac{P_2}{P_{stream} - P_{recovered} + P_{RF\,in}}. \tag{1.18}$$

The *electronic efficiency* is the efficiency with which power is transferred from the electron stream to the RF electric field of the output circuit.

$$\eta_e = \frac{P_{RF\,out} + P_{loss}}{P_{stream}}, \tag{1.19}$$

where P_{loss} is the total RF power loss in the output circuit. If the output circuit is resonant so that only power at the fundamental frequency is included

$$\eta_e = \frac{P_2 + P_{2,loss}}{P_{stream}}, \tag{1.20}$$