PHILIPS TECHNICAL LIBRARY

PRECISION ELECTRONICS

PRECISION ELECTRONICS

G. KLEIN

and

J. J. ZAALBERG VAN ZELST

PHILIPS TECHNICAL LIBRARY

Translated from the Dutch by K.S. Ankersmit, Purley, England

This book contains X + 466 pages and 467 illustrations.

U.D.C. nr. 621.38: 621.317

Printed in The Netherlands

Preface

At a time when the stream of information on electronics is threatening to reach flood proportions, one might ask why we have written a book on what seems to be an already familiar subject. We feel hesitant in giving an explanation, since anything we say is bound to sound a little presumptuous, but nevertheless we have tried in this preface to set down the facts which motivated us to produce it.

An ever increasing number of people are using or designing electronic equipment in factories, laboratories or other institutions. For these people most of the existing books do not quite reach the target for one or more of the following reasons:

1. An inordinately large part of the contents is often devoted to the technology and physics of components, particularly of valves and transistors, without utilizing this knowledge in a practical way throughout the remainder of the book.
2. A disproportion between the elementary knowledge assumed in the reader – particularly in mathematics – and the complexity of the electronic phenomena discussed. This leads to verbose over-simplified discussions, excessive calculations and a loss of general clarity.
3. A tendency towards completeness by including as many as possible of the circuits which have been developed during the history of electronics. Conversely, fundamental aspects of the subject are often treated rather superficially, thus presenting the non-specialist reader with the achievements of the main exponents of electronics without adequate explanation of how these results were derived.
4. Too little attention is paid to practical difficulties – especially to the secondary phenomena which can upset the performance of the practical design. The theoretically trained engineer who attempts to design measurement apparatus and who meets with constant failure because of the unwanted presence of a mains frequency interference signal in the materialization of his "brain-child" is seen all too often in laboratories.
5. Too much attention is paid to the application of electronics to radio and television, usually illustrated with a profuse quantity of circuits. This clouds any insight into the much wider scope of electronics and sacrifices accuracy in the process. For the mentioned groups of people the problems related to accuracy are, however, of prime importance.

We hope we have been able to overcome the aforementioned disadvantages in our approach to the subject in this book. We first discuss the most common components, methods of calculation and basic circuits in electronics, and then turn to principles and methods with special attention to the limits in the design of electronic measurement equipment. The first part of this book is therefore general in character, whilst the remainder devotes special attention to problems met specifically in measurement electronics.

We have tried to avoid the dryness of an over-systematic approach by giving the impression of a stroll through the field of electronics, sometimes making a slight detour to have a closer look at an interesting point, but always keeping the main objective in sight, which is to give the reader the correct approach so that he may be successful in his attempts at designing electronic equipment.

When explaining circuits, we have tried to keep to basic principles and have intentionally avoided giving complete lists of possible circuits. Once the reader is acquainted with these principles, he will certainly be in a position to discover the solution needed for his specific purpose and he can consult the circuits given in the available literature when he feels it necessary. We have, however, taken care to mention a great number of practical and useful points, mostly in the form of simple examples. A disadvantage of our method of presentation may be that it is sometimes difficult to locate some specific point. We have tried to overcome this by making the index as extensive as possible.

One danger of the rapid progress made in electronics is that any book will soon be outdated. However, a critical analysis of this development reveals that progress is chiefly technological and progress in electronics as a science (principles, basic effects and concepts) is extremely slow. We could even say, without wanting to belittle their importance, that the invention of transistors has had a restricting effect on the progress of basic electronics. By placing the main emphasis on the basic aspects, we hope that this book can be usefully consulted for many years to come.

Any success which this book will have depends not only on the contents and presentation as first conceived, but also on the way in which our ideas have been criticised and further elaborated. In this respect we are most grateful to many colleagues who have been so helpful in carefully studying our first drafts.

In particular we would like to thank Ir. J. te Winkel for his especially valuable contributions.

Finally we extend our thanks to all those who have contributed in any other way to make the publication of this book possible.

G. Klein
J. J. Zaalberg van Zelst

Preface to the English edition

Translating a technical work is always difficult and we are very grateful to Mr. Ankersmit for the care and skill he has devoted to the English version of our book. We believe Mr. Ankersmit produced a lively work in English and at the same time retained much of the original style.

Our thanks are also due to Mr. G. F. Steele who has checked the English manuscripts at all stages during their preparation for printing.

In electronic literature the use of international standardized symbols is fortunately under way. For English readers this makes little difference since the standardized symbols agree very largely with those already in use in the English literature. In this book the symbols used show only one notable departure from English practice and this is in the symbol for a transconductance. Here the reader will not find the English g_m but the symbol S. We expect that this substitution will not cause any inconvenience.

CONTENTS

1. Introduction

It is difficult to give a good definition of electronics because of the rapid development of this science and the variety of subjects it comprises. Several definitions have been attempted over recent years, ranging from the brief and inaccurate to the wordy and almost philosophical.

A common definition which is easily understood and which offers a reasonably comprehensive explanation is that "Electronics is the science concerned with the study and application of external electrical phenomena occurring when electrons pass through low-pressure gases and semiconductors".

One might assert that a great deal of what falls under this definition is outside the scope of "electronics proper", and that the applications are certainly not restricted to electronics alone. However, by combining this rather wide definition with the more current one, that "Electronics is the manipulation of currents, voltages and impedances by means of valves and transistors", we achieve a sufficiently accurate definition.

In any case the concept "electronics" covers such a wide field that in practice further subdivisions are usually required. These are often made in a rather haphazard manner – radio, television, medical electronics, microwave technique, digital circuitry etc. which results in a certain amount of overlapping. After discussing the more general aspects of electronics, this book will devote a large section to those techniques where a greater accuracy is demanded than normally required in common applications, especially radio and television. Measurement of a value usually infers its determination with a certain minimum accuracy (about 1 per cent). The techniques and their associated circuitry used to perform this measurement may be called "measurement electronics" and we can consider this concept as a separate field which is rapidly developing.

When a value is measured by means of electronics, it is always possible to recognize a few or all of the following signal operations:

1. Transformation of the signal to be measured into an electrical signal.
2. Amplification of the electrical signal.
3. Application of this amplified signal for control or other purposes.
4. Presentation of the measurement.

The above operations are illustrated by the schematic representation of the measurement of the temperature in an oven (Fig. 1-1). The thermo-couple produces an electrical signal, that is, a voltage which corresponds to the temperature of the oven. Such a voltage has a magnitude of millivolts, which is too small for most applications. However, the amplifier now produces such a large signal that it can control the heat passing to the oven after further conversion (e.g. to an h.f. current). A pen recorder ensures that the electrical signal, which corresponds to the oven temperature, is recorded.

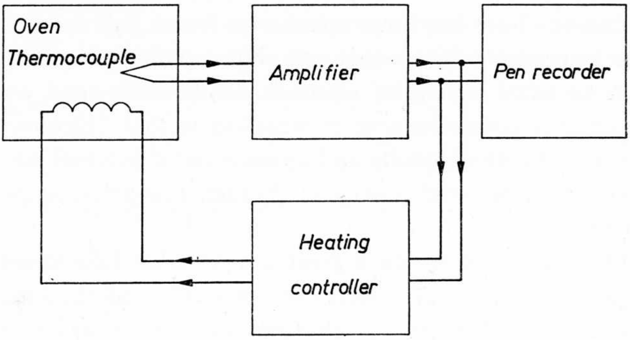

Fig. 1-1

Without discussing this subject in detail, it can be seen that in principle there are unlimited practical solutions for the operations mentioned. In practice, however, one usually attempts to achieve a solution that is as simple as possible, that is, a linear relation between the amplified electrical signal and the original measurement signal. To obtain a clear understanding of what takes place during the signal treatment, especially as far as complications and possibilities are concerned we are well justified in restricting ourselves to a study of this linear relation. We shall not in fact treat the transformation of the signal to be measured into an electrical signal in this book. It is assumed that the electrical signal present is such an accurate conversion of the original signal, that it justifies great accuracy in amplication and further treatment. This is certainly not so in many measurements, where the lack of accuracy in the conversion limits the accuracy of measurement. This explains why so much research in the technique of measurements is being undertaken on the improvement of transducers.

Regarding further treatment and recording, we should mention that the normally rather low accuracy of recording apparatus does not necessarily impose a limitation on the accuracy of measurement. By using compensation methods or additional amplification it is nearly always possible to achieve a scale enlargement which correspondingly reduces the inaccuracies in the recording. We shall therefore not dwell on the various recording methods.

For many of the operations mentioned under 3., the same auxiliary apparatus are often used, such as oscillators, multivibrators and other relaxation circuits and it is therefore necessary to discuss these in detail. It will be obvious that only a limited number of examples of actual operational circuits can be given, and in view of the nature of this book, we have attempted to indicate clearly which practical solutions are available for achieving great accuracy.

Before discussing amplification, we shall first have a close look at the most important components, as well as some properties of linear circuits and the usual calculation methods.

2. Components

We can divide the many kinds of components used in electronics into passive and active components.

Passive components are not capable of supplying energy to the circuit of which they form a part. This group includes the resistor, the capacitor and the inductor as two-terminal components, and the transformer, the gyrator and the cable as four-terminal components. Here, the gyrator is mainly of theoretical importance. Transformers and inductors appear to have so many disadvantages (low versatility, sensitivity to interference, etc.), especially at low frequencies and low energy levels, that it is nearly always worthwhile considering if their use can be avoided. Apart from making connections between circuits, cables are only used in special cases, so that primarily one will meet only resistors and capacitors as passive components in electronic measurement circuits.

Active components, unlike passive components, are capable of supplying energy to the circuit of which they form a part. By energy, we do not mean any random form of energy, but the type which can in some way be used for the intended purpose, that is energy containing information. These components are not themselves sources of energy, but are only able to supply it by abstracting electrical energy from supplies such as accumulators or batteries. This section of active components comprises transistors and electronic valves, including magnetrons, klystrons and similar valves.

"Normal" valves and transistors are primarily used in measurement circuits. In some aspects they exhibit similar properties, but in others they are totally different. When one has to choose between a transistor and a valve, it is advisable to consider all aspects and then make a selection that best meets the specific requirements. In particular, one should not exclude in principle the possibility of combining valves and transistors.

Until a few years ago, greater accuracy could be obtained with valve circuits than with corresponding transistorized networks. This difference in accuracy has now been largely offset by the development of types of junction transistors that have excellent amplification properties, even at very small currents (high current amplification factor and high cut-off frequency), and of various types of field-effect transistors. The accuracy of transistorized circuits is now even greater in some applications.

Since the operation of a valve is usually easier to visualize than that of a transistor (grid current is normally negligible), most circuits illustrating some principle are shown with valves. As the properties of the field-effect transistor used as an amplifying element fully correspond to those of the triode or pentode, its use in circuits will not be dealt with separately. If the conversion of a valve circuit to one with junction transistors results in unexpected consequences, these will be mentioned. We shall also give examples of cases where the use of transistors offers essential advantages.

In the first part of the book, only networks will be presented which incorporate exclusively the most frequently used components such as resistors, capacitors, valves and transistors. The discussion of the other components such as transformers, cables, gas-discharge valves and zener diodes is deferred until they are applied for the first time in a circuit.

3. Resistor, capacitor and inductor

Ohm's Law applies to the ideal resistor:

$$V(t) = R \cdot I(t) \tag{3.1}$$

where $V(t)$ = voltage across the resistor at time t, $I(t)$ = current flowing through the resistor at time t, R = constant of resistor's value.

The signs of current and voltage of components and combinations of components are always chosen so that the current enters at the terminal with positive voltage, and leaves at the terminal with negative voltage (Fig. 3-1). Constant R is then positive for resistors.

The unit of resistance is the Ohm (Ω).
The normal resistor values lie between approx. 10 and 10^8 ohms; in some cases values up to approx. 10^{-2} and 10^{12} ohms are used.

In practice it is particularly important to take into account the following departures from the ideal case:

a) The practical resistor whose value depends on temperature. The temperature coefficient (relative change in resistance per °C) usually lies between approx. 10^{-3} in the case of carbon resistors and 10^{-5} in the case of high-quality metal resistors. The effects of air pressure and humidity are generally so insignificant that they can be neglected.
b) The relation between V and I is non-linear. This can be allowed for by assuming that R in (3.1) is not a constant, but depends on V and I.
c) A fluctuating voltage in series with the applied voltage is also present. This is caused by the thermal motion of the current carriers in the resistor, the "thermal noise". The magnitude of this voltage depends on the resistance value and the temperature of the resistor.
d) Some resistors, particularly the carbon type, exhibit a fluctuating voltage caused by fluctuations in resistance value. Since the value of this voltage is in the first instance proportional to the current passing through the resistor, this voltage is usually called "current noise".
e) A resistor also possesses a certain capacity and self-inductance, which can produce an important effect at high frequencies.

The following formula applies to the ideal capacitor (Fig. 3-2):

$$Q(t) = C \cdot V(t) \tag{3.2}$$

where Q = charge, C = capacitance of capacitor.

The unit of capacitance, the farad (F) is too large for practical use. Capacitances are therefore usually expressed in micro- and picofarads (10^{-6} and 10^{-12} F respectively; symbols μF and pF). Values used in electronics vary between approx. 1 pF and several thousand μF.

The most important departure from the ideal case is caused by dielectric loss in the capacitor. These losses can be allowed for by assuming a resistor R to be present in parallel to the capacitor, whose

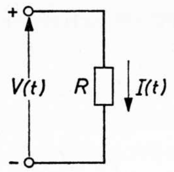

Fig. 3-1

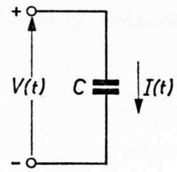

Fig. 3-2

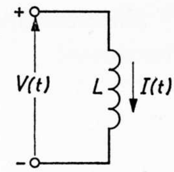

Fig. 3-3

value becomes smaller when the losses increase. As losses increase with the capacitance value and frequency ω, the product ωCR is a good measure of the quality of the capacitor above a given minimum frequency. A good-quality capacitor has a "loss factor" $1/\omega CR$ of about 10^{-3}. In the presence of a d.c. voltage between the terminals, a certain leakage current will occur. This can also be represented by a resistor R in parallel to the capacitor. In this case, product CR is a measure of quality.

A second important deviation from the ideal case is the temperature dependence of the capacitance value. The corresponding temperature coefficient has a value of approx. $10^{-5}-10^{-4}$, dependent on the quality of the capacitors. Moreover, some types of capacitors also have a certain amount of self-inductance which may become important at high frequencies, whilst electrolytic capacitors show a certain series resistance.

For the instant t_0 equation (3.2) becomes

$$Q(t_0) = CV(t_0)$$

and therefore

$$C\{V(t) - V(t_0)\} = Q(t) - Q(t_0)$$

This change in charge must have taken place between times t_0 and t in the form of supply or withdrawal of current, so that

$$Q(t) - Q(t_0) = \int_{t_0}^{t} I(t)\,\mathrm{d}t$$

therefore

$$C\{V(t) - V(t_0)\} = \int_{t_0}^{t} I(t)\,\mathrm{d}t$$

or

$$V(t) = \frac{1}{C}\int_{t_0}^{t} I(t)\,\mathrm{d}t + V(t_0) \tag{3.3}$$

If $V(t)$ is given, $V(t_0)$ will also be known and $I(t)$ will be entirely defined. But if $I(t)$ is given, $V(t)$ will only be fully defined when $V(t_0)$ is also known.

Since the integral equation (3.3) is not very attractive, it is usually differentiated:

$$I(t) = C\frac{\mathrm{d}V(t)}{\mathrm{d}t} \tag{3.4}$$

In this case too, $I(t)$ will follow unambiguously when $V(t)$ is given, but in order to determine $V(t)$ when $I(t)$ is known, further data must be added, e.g. $V(t_0) = V_0$. In other words, an additional condition must be known.

For an ideal inductor the following formula applies (Fig. 3-3).

$$\Phi(t) = L \cdot I(t) \tag{3.5}$$

where Φ = total magnetic flux through inductor, L = inductance coefficient of inductor.

The henry (H) is the unit of inductance. In practice, inductors also have a certain resistance (because of the finite conductance of the inductor) which may be visualized as being in series with the "ideal" inductor. Because of "skin effect", this resistance increases with frequency. The inductance coefficient will also vary slightly with frequency, but this effect is usually of little importance. Inductors with core materials show additional losses, which can be again accounted for by a resistor. Finally, the windings of an inductor are also coupled in a capacitive manner, which may become a major influencing factor, particularly at high frequencies.

The integral relation of the inductor is derived from (3.5) in a similar fashion as used with the capacitor

$$I(t) = \frac{1}{L}\int_{t_0}^{t} V(t)\mathrm{d}t + I(t_0) \tag{3.6}$$

Here the voltage is fully determined when the current is known. However, in order to determine the current from the voltage, $I(t_0)$ should also be known.

Differentiation gives here:

$$V(t) = L\frac{\mathrm{d}I(t)}{\mathrm{d}t} \tag{3.7}$$

which must be supplemented by an additional condition, such as $I(t_0) = I_0$.

Assuming ideal components from now on, we find that although linear relations do exist between voltage and current for the capacitor and the inductor, the corresponding integral or differential equations present a considerable difference to the simplicity of Ohm's Law that is valid for resistors. One wonders, therefore, whether it is possible to reduce these equations to a simpler form. This is not possible without introducing some restriction in one way or another. A sensible restriction appears to be not to work with arbitrary waveforms of current and voltage, but with those waveforms where the above-mentioned differential equations change into equations of the same form as Ohm's Law; that is:

$$V(t) = Z \cdot I(t)$$

where Z now represents the "resistance value" of the capacitor or inductor.

We thus obtain for the capacitor:

$$I(t) = C\frac{\mathrm{d}V(t)}{\mathrm{d}t} = C\frac{\mathrm{d}\{ZI(t)\}}{\mathrm{d}t} = CZ\frac{\mathrm{d}I(t)}{\mathrm{d}t}$$

or

$$\frac{\mathrm{d}I(t)}{I(t)} = \frac{\mathrm{d}t}{CZ}$$

therefore:

$$\ln I(t) = \frac{t}{CZ} + \ln A$$

or

$$I(t) = A\mathrm{e}^{t/CZ} = A\mathrm{e}^{pt}$$

where, because no restrictions have been made for Z in this respect, $p = 1/CZ$ may obviously be an arbitrary constant. The corresponding voltage then becomes:

$$V(t) = ZI(t) = \frac{1}{pC}A\mathrm{e}^{pt}.$$

For current and voltage waveforms where the time-dependent part of the waveform is e^{pt}, we can thus say that Ohm's Law is valid for the capacitor:

$$V(t) = ZI(t)$$

with $Z = 1/pC$.
Similarly it can be shown that the same applies to inductors, but in this case $Z = pL$.

The widest range of these current and voltage waveforms is naturally obtained by taking the widest choice of p, that is by selecting for p a complex number $\alpha + j\omega$. At first it may appear strange that complex currents and voltages will occur in this case. We shall, however, demonstrate that calculations can be carried out successfully and to advantage with complex currents and voltages.

This gives us a third manner of describing a capacitor and inductor, namely:

$$V(t) = \frac{1}{pC} I(t) \tag{3.8}$$

and

$$V(t) = pL\, I(t) \tag{3.9}$$

respectively.

These equations thus only apply to a certain class of current and voltage waveforms, but fortunately this class happens to be of very great practical importance. Without conditions, the undefined terms are here also the voltage for capacitors and the current for inductors.

Quantities such as pL and $1/pC$, with the dimension of a resistance, are called impedances, their reciprocals $1/pL$ and pC admittances.

4. Current and voltage sources

Apart from components, as discussed above, we must also consider sources of current and voltages in electronics.

An ideal voltage source supplies across its terminals a voltage which is independent of the current taken. Similarly, an ideal current source produces a current which is independent of the voltage across the terminals. The voltage and current supplied may be dependent on time. In practice, however, no ideal source exists. What we call a voltage source produces a voltage which decreases with increasing current taken, while in the case of a current source the current supplied decreases with increasing voltage across the terminals. In normal circumstances, these dependences can usually be allowed for with sufficient accuracy by considering the voltage source as the series-circuit of an ideal voltage source and a relatively small "internal resistance" r_i, and the current source as an ideal current source with a large parallel internal resistance R_i.

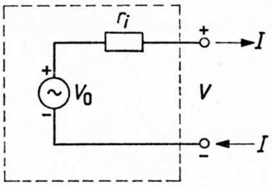

Fig. 4-1

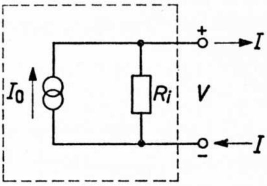

Fig. 4-2

One should note that in practice this makes it impossible to differentiate between the behaviour of non-ideal voltage and current sources. The only characteristic which can be determined external to a source is the relation between voltage and current. For the voltage source of Fig. 4-1 this is:

$$V = V_0 - Ir_i$$

and for the current source of Fig. 4-2:

$$V = (I_0 - I)R_i = I_0R_i - IR_i.$$

and these expressions are identical when $R_i = r_i$ and $V_0 = I_0R_i$.

In the following we shall often refer to the elimination of current and voltage sources. This should be understood to mean that the current or voltage of the source is made zero. With a non-ideal current source, the part representing the ideal current source will therefore be open circuited, and with a non-ideal voltage source, the part representing the ideal voltage source will be replaced by a short-circuit. The non-ideal sources should thus be replaced by their internal resistances.

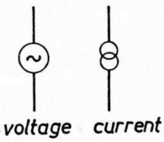

Fig. 4-3

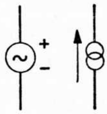

Fig. 4-4

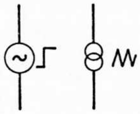

Fig. 4-5

We should also mention that the symbols of Fig. 4-3 will be used in the rest of this book for indicating an applied signal voltage or signal current. The polarity or positive current direction assumed for calculations is sometimes indicated in the circuit (Fig. 4-4), whilst the reading of circuits is often facilitated by a brief indication of the time dependence of the signals (Fig. 4-5).

5. Superposition

We shall now consider a circuit consisting of linear passive components and current and voltage sources. Kirchhoff's Laws ($\Sigma I = 0$ for each point, $\Sigma V = 0$ for each loop) and the relations between current and voltage for all components, will enable us to determine the equations for the currents which flow in the various branches, and the potential differences which exist between several points of the circuit. The known terms will consist of linear combinations of the given (applied) source currents and source voltages and, as appropriate, their integrals or derivatives. As these equations are linear in all voltages and currents, it follows that the value found for each of the voltages and currents equals the sum of the values which would be found if each of the given source voltages and source currents were operative individually. This property of linear electrical systems is known as superposition. We should add that the above reasoning is only valid when the coefficients of currents and voltages do not change. Since these coefficients may include the values of the various internal resistances of the current and voltage sources, the convention mentioned in the previous section for the elimination and re-insertion of these sources, should be taken into consideration.

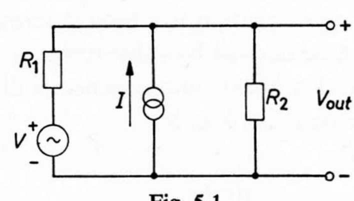

Fig. 5-1

Example (Fig. 5-1)

$$V_{\text{out}} = V_{\text{out}}(I = 0) + V_{\text{out}}(V = 0).$$

$$V_{\text{out}}(I = 0) = V \frac{R_2}{R_1 + R_2} \quad \text{and} \quad V_{\text{out}}(V = 0) = I \frac{R_1 R_2}{R_1 + R_2}$$

therefore $$V_{\text{out}} = V \frac{R_2}{R_1 + R_2} + I \frac{R_1 R_2}{R_1 + R_2}.$$

6. Equations

If, in deriving the equations for a system, we make use of the integral relations for inductors and capacitors, the set of equations thus obtained will fully describe the circuit. The solving of a set of integral equations is usually done indirectly by first determining the general solution of the set of differential equations which is produced by differentiation of the integral equations. Values for the integration constants are chosen so that the integral equations are satisfied; in other words, we adapt the solution of the differential equations to the additional conditions. In our case it is not necessary to derive this set of differential equations by first producing the integral equations. These differential equations can be obtained directly by using the differential relation for each inductor and capacitor.

The general solution of a set of linear differential equations is found by eliminating all unknown voltages and currents with the exception of a single one. Differentiation will be necessary for this elimination, and a higher-order linear differential equation is produced for the remaining unknown, where the constant coefficients consist of a combination of resistance values, capacitances and inductances. The known righthand side of this equation is a linear combination of the given source voltages and source currents, and possibly their derivatives.

Once the solution of this equation has been determined, the solutions of the other unknowns can be derived by substituting in the original set.

The general solution of a linear inhomogeneous differential equation of the n-th order with constant coefficients

$$B_n \frac{\mathrm{d}^n x}{\mathrm{d}t^n} + B_{n-1} \frac{\mathrm{d}^{n-1} x}{\mathrm{d}t^{n-1}} + \ldots + B_0 x = f(t) \tag{6.1}$$

consists of the sum of a single "particular solution" of this equation and the general solution of the corresponding homogeneous differential equation:

$$B_n \frac{\mathrm{d}^n x}{\mathrm{d}t^n} + B_{n-1} \frac{\mathrm{d}^{n-1} x}{\mathrm{d}t^{n-1}} + \ldots + B_0 x = 0 \tag{6.2}$$

The solutions of this last equation, that is the solutions which are possible for the system if all sources are eliminated, are called the "natural modes or oscillations" of the system. These natural modes are usually of the form $x = x_0 e^{p_i t}$, where x_0 is a constant and p_i a root of the equation

$$B_n p^n + B_{n-1} p^{n-1} + \ldots + B_0 = 0 \tag{6.3}$$

There are thus n natural "frequencies" p_i.

Only when (6.3) has multiple roots, can solutions of the form $t^k e^{p_i t}$ occur. If p_i is a k-fold root, we have:

$$x = \{x_0 + c_1 t + c_2 t^2 + \ldots c_{k-1} t^{k-1}\} e^{p_i t}$$

with k constants: $x_0, c_1, c_2 \ldots c_{k-1}$.

The word "frequencies" is used in this context because complex roots $p_i = a_i + j\omega_i$ occur in many cases, so that we can write

$$e^{p_i t} = e^{a_i t} (\cos \omega_i t + j \sin \omega_i t)$$

which will produce damped, constant, or increasing sinusoidal oscillations according to whether a_i is smaller, equal, or larger than 0. With all passive systems occurring in practice, the natural modes will be damped oscillations because the energy will always be decreased by dissipation. If active components are also used, oscillations with constant or increasing amplitude may occur. We then speak of oscillating systems. We shall later refer to the necessary calculations for these systems, so that we shall limit ourselves here to systems whose natural modes are damped.

The extent to which natural modes will occur is determined by the additional conditions. In the general solution of (6.2), the n constants which occur as coefficients of the natural modes, must be chosen so that the general solution of (6.1) satisfies the additional conditions. It often happens that the latter refer to the initial situation of a system; they are then called "initial conditions". By choosing in this case the particular solution of (6.2) in which no natural modes occur, this solution will describe the situation in the system if the initial conditions have been eliminated i.e. the "steady-state" situation has been reached. "Switching-on phenomena" which belong to the initial conditions are then completely allowed for in the coefficients of the natural modes.

In many cases, especially when dealing with sinusoidal signals, we are chiefly interested in the steady-state situation. It is easy to determine the corresponding particular solution when the righthand side of equation (6.1) can be written as the sum of a number of terms Ae^{pt}, where p is not a root of equation (6.3), that is the terms do not correspond to the natural modes of the system.

The particular solution of a linear differential equation whose righthand side is a sum $f_1(t) + f_2(t) + \ldots$, is equal to the sum of the particular solutions obtained by considering each term $f_1(t)$, $f_2(t)$ separately. This means that we can limit our attention to the particular solution of equations of the form

$$B_n \frac{\mathrm{d}^n x}{\mathrm{d}t^n} + B_{n-1} \frac{\mathrm{d}^{n-1} x}{\mathrm{d}t^{n-1}} + \ldots + B_0 x = A\, \mathrm{e}^{pt}$$

Following the theory of linear differential equations with constant coefficients, we can put

$$x = D\mathrm{e}^{pt},$$

where

$$D\left\{B_n p^n + B_{n-1} p^{n-1} + \ldots + B_0\right\} = A \tag{6.4}$$

In this case all currents and voltages in the circuit thus possess exactly that time-dependence for which the simple algebraic relations $V = I/pC$ and $V = LpI$ can be used for the capacitor and the inductor. Equation (6.4) could have been obtained directly from these relations which means that only algebraic equations have to be solved.

It is normal practice to depart from the algebraic relations for capacitor and inductor. We thus obtain equations in the "operator" p which corresponds with the operation "differentiation with respect to time". When the source voltages and currents do not have the required form $A\mathrm{e}^{pt}$, this method can be regarded as a quick way to arrive at differential equation (6.1).

As stated, the righthand side will usually contain all the applied voltages and currents, as well as a number of derivatives. However, superposition allows us to restrict the righthand side to a single applied value $f(t)$ and its derivatives. The general equation which must be solved for linear circuits is therefore:

$$B_n \frac{\mathrm{d}^n x(t)}{\mathrm{d}t^n} + B_{n-1} \frac{\mathrm{d}^{n-1} x(t)}{\mathrm{d}t^{n-1}} + \ldots + B_0 x(t) =$$

$$A_m \frac{\mathrm{d}^m f(t)}{\mathrm{d}t^m} + A_{m-1} \frac{\mathrm{d}^{m-1} f(t)}{\mathrm{d}t^{m-1}} + \ldots + A_0 f(t)$$

or, with

$$\begin{cases} B(p) = B_n p^n + B_{n-1}\, p^{n-1} + \ldots + B_0 \\ A(p) = A_m p^m + A_{m-1}\, p^{m-1} + \ldots + A_0 \end{cases}$$

$$\left\{B(p)\right\} \cdot x(t) = \left\{A(p)\right\} \cdot f(t) \tag{6.5}$$

where brackets are used to avoid misunderstanding and to indicate the symbolical way of writing a differential equation with operator $p = \mathrm{d}/\mathrm{d}t$. Functions $A(p)$ and $B(p)$ have no common roots because otherwise simplification would be possible.

In the next section we shall consider the calculation methods which greatly facilitate the solution of (6.5) for a few of the most important signal forms.

7. Calculation methods

Amongst the most important of the signal forms which occur in practice are those which have a sinusoidal form when expressed against time.

It appears from $\cos \omega t = \frac{1}{2}(e^{j\omega t} + e^{-j\omega t})$ and $\sin \omega t = \frac{1}{2j}(e^{j\omega t} - e^{-j\omega t})$ that these signals belong to the class where the time-dependent part has the form e^{pt}. Since we are normally only interested in the steady-state situation of these sinusoidal signals, it is useful to determine the corresponding solution for $f(t) = V_0 e^{pt}$ from differential equation (6.5), which can also be written in the symbolical form:

$$x(t) = \left\{ \frac{A(p)}{B(p)} \right\} \cdot f(t) \tag{7.1}$$

We have seen in the previous section that in the presence of a single function e^{pt} in the righthand side of (6.5), the function $\{B(p)\}$ could be replaced by the algebraic form $B(p)$, provided that e^{pt} does not correspond to a natural mode of the system. However, from $d^k/dt^k \cdot e^{pt} = p^k \cdot e^{pt}$ follows that we can also replace $\{A(p)\}$ in the righthand side by $A(p)$, which gives:

$$x(t) = \frac{A(p)}{B(p)} V_0 e^{pt} \tag{7.2}$$

This means that with the time-dependence e^{pt}, the steady-state solution is found by replacing the operation $\{A(p)/B(p)\}$ by the algebraic form $A(p)/B(p)$. This obviates any necessity for solving differential equations in this case.

If e^{pt} does correspond to a natural mode of the system, that is $B(p) = 0$, the solution will be:

$$x(t) = \frac{tA(p)}{B'(p)} V_0 e^{pt}$$

where $B'(p)$ is the first derivative of $B(p)$. If p is a k-fold root of $B(p) = 0$, we have likewise:

$$x(t) = \frac{t^k A(p)}{B^{(k)}(p)} V_0 e^{pt}$$

where $B^{(k)}(p)$ is the k-th derivative of $B(p)$.

Let us now consider the most important practical case, when p is purely imaginary, $p = j\omega$, and signal $V_0 e^{j\omega t}$ is the sum of a real voltage $V_0 \cos \omega t$

and an imaginary voltage $jV_0 \sin \omega t$. Since the natural oscillations of the system are damped, these signals cannot coincide, so that equation (7.2) is always valid. In this case we find:

$$x(t) = \frac{A(j\omega)}{B(j\omega)} V_0 \, e^{j\omega t}$$

We can write (see Fig. 7-1):

$$\frac{A(j\omega)}{B(j\omega)} = \left| \frac{A(j\omega)}{B(j\omega)} \right| \cdot e^{j\varphi}$$

where $|A(j\omega)/B(j\omega)|$ is the modulus and φ the argument of $A(j\omega)/B(j\omega)$, so that

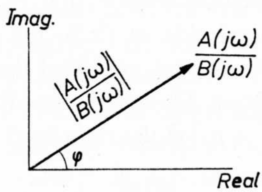

Fig. 7-1

$$x(t) = \left| \frac{A(j\omega)}{B(j\omega)} \right| V_0 \, e^{j(\omega t + \varphi)} =$$

$$\left| \frac{A(j\omega)}{B(j\omega)} \right| V_0 \left\{ \cos(\omega t + \varphi) + j \sin(\omega t + \varphi) \right\}$$

In this case the real part will correspond to the real part of $V_0 \, e^{pt}$, i.e. to $V_0 \cos \omega t$, and the imaginary part to $jV_0 \sin \omega t$, which brings us to the following result:

If $$x(t) = \left\{ \frac{A(p)}{B(p)} \right\} \cdot f(t)$$

is valid for the relation between the "response" $x(t)$ and the "signal" $f(t)$ in a system with a "transfer function" $A(p)/B(p)$ and if $f(t)$ is a sinusoidal signal with frequency ω, we find:

$$\text{ampl}\ x(t) = \left| \frac{A(j\omega)}{B(j\omega)} \right| \text{ampl}\ f(t)$$

and

$$\arg x(t) = \arg \frac{A(j\omega)}{B(j\omega)} + \arg f(t).$$

Although we shall make repeated use of this property in the following pages, it seems still useful to illustrate the above by means of a simple example.

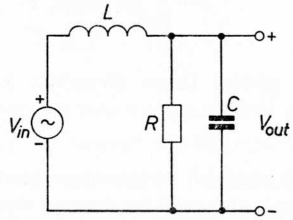

Fig. 7-2

For the circuit in Fig. 7-2:

$$V_{out} = \frac{\dfrac{R}{1+pRC}}{pL + \dfrac{R}{1+pRC}} V_{in} = \frac{R}{R + pL + p^2LRC} V_{in}$$

If $V_{in} = V_0 \sin \omega t$, we must calculate the transfer function for $p = j\omega$:

$$\frac{R}{R - \omega^2 LRC + j\omega L} = \frac{R^2(1 - \omega^2 LC) - j\omega LR}{R^2(1 - \omega^2 LC)^2 + \omega^2 L^2}$$

The argument of the fraction is angle φ:

$$\cos \varphi = \frac{R(1 - \omega^2 LC)}{\sqrt{R^2(1 - \omega^2 LC)^2 + \omega^2 L^2}}, \quad \sin \varphi = \frac{-\omega L}{\sqrt{R^2(1 - \omega^2 LC)^2 + \omega^2 L^2}}$$

It follows that φ is also defined by:

$$\tan \varphi = \frac{-\omega L}{R(1 - \omega^2 LC)}$$

provided φ is chosen in the correct quadrant.

The modulus becomes:

$$\frac{R}{\sqrt{R^2(1 - \omega^2 LC)^2 + \omega^2 L^2}}$$

therefore:

$$V_{out} = \frac{R}{\sqrt{R^2(1 - \omega^2 LC)^2 + \omega^2 L^2}} V_0 \sin(\omega t + \varphi).$$

If one has only to deal with sinusoidal signals, it is of course possible to apply the substitution $p = \mathrm{j}\omega$ when formulating the equations, which means that instead of the relations

$$V = \frac{I}{pC} \quad \text{and} \quad V = pLI$$

for capacitor and inductor respectively, the relations

$$V = \frac{I}{\mathrm{j}\omega C} \quad \text{and} \quad V = \mathrm{j}\omega LI$$

are used, thereby giving these elements an impedance $1/\mathrm{j}\omega C$ and $\mathrm{j}\omega L$ respectively.

With regard to the other signal forms belonging to the class e^{pt}, the above is also valid for damped sinusoidal signals provided that they do not correspond to a natural oscillation of the system, as well as for sinusoidal signals of increasing amplitude, if $A(\mathrm{j}\omega)/B(\mathrm{j}\omega)$ is replaced by $A(\alpha + \mathrm{j}\omega)/B(\alpha + \mathrm{j}\omega)$ when $\mathrm{e}^{\alpha t}$ is the corresponding damping term.

These cases are of no great practical interest, neither are increasing exponential functions (e^{pt} with p real and positive). For decreasing exponential functions the interest lies mainly in switch-on phenomena rather than in the steady-state solution.

Apart from sinusoidal signals, electronics has a great deal to do with transient phenomena, namely those occurring when signal sources are turned on or off, and when introducing open and short circuits in a system.

It is obvious that in this instance it is not the easily determined steady-state conditions which are the most important but rather the phenomena occurring during switching on and off.

Since these phenomena are completely allowed for by the system's natural oscillations, and the time dependence of these is known (see Section 6), only the coefficients of the resonant frequencies need to be determined in these cases. However, since we deal in the first part of this book exclusively with sinusoidal signals, we shall defer discussing the calculation method promulgated by Heaviside to Section 22.

There is little need in electronics, as discussed in this book, for solving methods for more general signal forms.

Laplace transforms can be used where necessary, and are capable of giving the solution of a large number of signal forms in a direct manner. Because of the very comprehensive literature on this method, we shall not discuss it in this book.

8. Thévenin's theorem

Another theorem in circuit theory which is of importance to the electronic engineer is known as Thévenin's theorem which states: If an impedance Z is connected between two points of a circuit containing linear elements and current and voltage sources, a current I_z will flow through Z which is determined by the equation:

$$I_z = \frac{V_0}{Z + Z_0} \tag{8.1}$$

where V_0 is the voltage between the two points before connecting Z (open-circuit voltage, Fig.8-1) and Z_0 the impedance of original circuit, measured between the two points when all current and voltage sources are negated. This can be easily understood when visualizing an ideal voltage source with voltage V_0 connected to one of the two points, with the polarity as indicated in Fig. 8-2. The voltage between the remaining terminals is then zero. If impedance Z is now connected between these remaining terminals, no current will flow through Z (Fig. 8-3). According to the superposition theorem, this current can be considered as the sum of the two currents. The first one of these currents is that which would pass when only the original source was operative, while the second would occur if only the additional

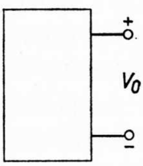

Fig. 8-1

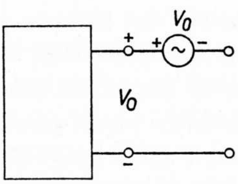

Fig. 8-2

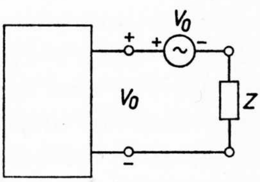

Fig. 8-3

source was present. The first one is the current taken, and therefore has the same value as the current which would be produced by the additional source alone. The latter's value is given by equation (8.1).

Taking the special case $Z = 0$, that is when the two points are short-circuited, the "short-circuit current" passing through this short circuit, is defined by

$$I_0 = \frac{V_0}{Z_0}$$

Conversely, it follows that the "internal impedance" Z_0 can be determined from the open-circuit voltage and the short-circuit current:

$$Z_0 = \frac{V_0}{I_0}$$

Therefore Thévenin's theorem can also be interpreted as follows: any two points of a linear circuit can be considered, observed externally, as the terminals of a non-ideal voltage or current source having a source voltage or current equal to the open-circuit voltage or short-circuit current between these points, while the internal impedance of the source equals the internal impedance of the circuit between these two points.

We should note that although it is permissible to consider the circuit simultaneously for more than one pair of points as a voltage or current source, it is not necessarily possible to combine these derived results.

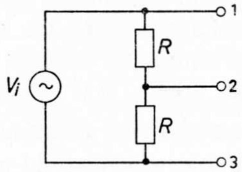

Fig. 8-4

Example: The circuit shown in Fig. 8-4 between points 1 and 2 can be replaced by the voltage source with the open-circuit voltage $\frac{1}{2}V_i$ and internal resistance $\frac{1}{2}R$. The same applies to the circuit between points 2 and 3. Connecting these two circuits in series would result in the possibility of replacing the circuit between points 1 and 3 by a voltage source with open-circuit voltage V_i and internal resistance $\frac{1}{2}R + \frac{1}{2}R = R$, while in reality this resistance is zero.

9. Electronic valves

In any discussion on active components, we should first deal with electronic valves; not only for historical reasons but especially because the mechanism of current conductance is much simpler with valves than with transistors. The action of electronic valves is based on the properties of metals and metal oxides which emit electrons, particularly at high temperatures (Edison effect). If these electrons are emitted in sufficiently rarefied gases, they are able to cover great distances. This flow of electrons can be "controlled" by electrical fields both in direction and in size.

The electronic valve with the simplest design is the diode, which only contains two electrodes *in vacuo*, the cathode and the anode. The design of the diode and other valves with more electrodes is usually concentric (Fig. 9-1). The cathodes of most valves suitable for our purpose are heated indirectly, that is by means of a filament which is situated in the cathode, a hollow metal cylinder.

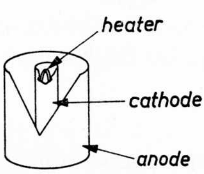

Fig. 9-1

The primary electrons emitted by the cathode are able to move about freely. If they arrive at places that are isolated from the cathode, they will give these places a negative charge and make it more difficult, and gradually even impossible, for electrons following them to reach these same places. When the anode is isolated from the cathode, it will acquire a negative charge and eventually become unattainable to electrons. Further emitted electrons have no alternative but to return to the cathode. However, one nearly always provides an external conducting path between anode and cathode, in which supply sources may or may not be incorporated. The electrons can then return from the anode to the cathode and no permanent electrostatic charging of the anode occurs, so that the anode will continue to attract new electrons from the cathode. The electron flow travels internally from the cathode to the anode and externally in the opposite direction. This external flow can be measured and is defined by the number of electrons which are able to reach the anode. This number in turn depends on the potential difference between anode and cathode. If the anode is much more negative than the cathode, almost no electrons will flow to the anode and the current will therefore be almost zero. If the anode is sufficiently positive with respect to the cathode, all emitted electrons will arrive at the anode. There is a point at which a maximum current occurs and this will not increase when the anode voltage is raised. The diode is then saturated. The saturation current also depends on the cathode heating. Emission, and hence the saturation current, greatly increases at higher cathode temperatures. Apart from a few exceptions, diodes are always used with currents that are much smaller than the saturation current.

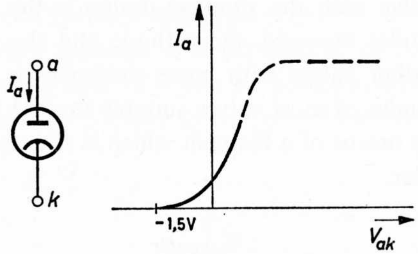

Fig. 9-2

Hence at a fixed operational temperature, we have a fixed relation between current and voltage. This is illustrated in Fig. 9-2. The voltage at which almost no current passes is about —1.5 volt. At a given point on the current-voltage characteristic curve of a diode, the latter will behave as a resistor in series with a d.c. voltage if the voltage and current changes are sufficiently small for the curvature of this curve to be neglected. Such linear behaviour is however of little interest. Almost all applications of the diode are therefore based on the non-linear behaviour obtained with large changes in current and voltage; these will be discussed at a later stage.

10. Triode

The triode offers far more possibilities than the diode. It differs from the latter by the presence of a third electrode in the shape of a thin wire spiral between anode and cathode, Fig. 10-1. This electrode is called a "grid". By giving the grid a negative potential with respect to the cathode, it is possible to influence the electric field between anode and cathode, and hence the anode current. The "controlling" function of such a grid is based on this fact. The grid is usually indicated in diagrams by the letter *g*.

Fig.10-1 shows the three voltages and three currents which occur in a triode. The following relations exist between the three voltages:

$$V_{ag} + V_{gk} = V_{ak}$$

and between the three currents:

$$I_g + I_a = I_k$$

so that the properties of a triode can be described by the relations existing between two currents and two voltages. For the voltages, the grid-cathode voltage V_{gk} and the anode-cathode voltage V_{ak} are usually chosen, and for the currents the grid current I_g and the anode current I_a. It is usual to

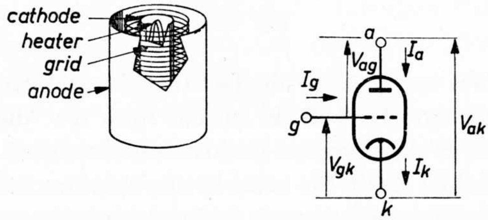

Fig. 10-1

give the currents' dependence on the voltages. Fig. 10-2 shows I_a and I_g against V_{gk} at various voltages V_{ak} for a conventional triode at voltage and current values which are usual in practice. The grid-cathode voltage has a small negative value of a few volts, whilst that of the anode-cathode voltage may amount to several hundred volts positive. The magnitude of the anode current may have values between a few and several tens of milliamps. A much larger variation occurs in the value of grid current, at least in the interesting region of $V_{gk} < -1$ volt. For example the industrial high-quality valve E80CC (=6085) has a grid current between 10^{-9} A and 10^{-11} A whereas in the cheaper "entertainment" valves, this current varies between 10^{-6} A

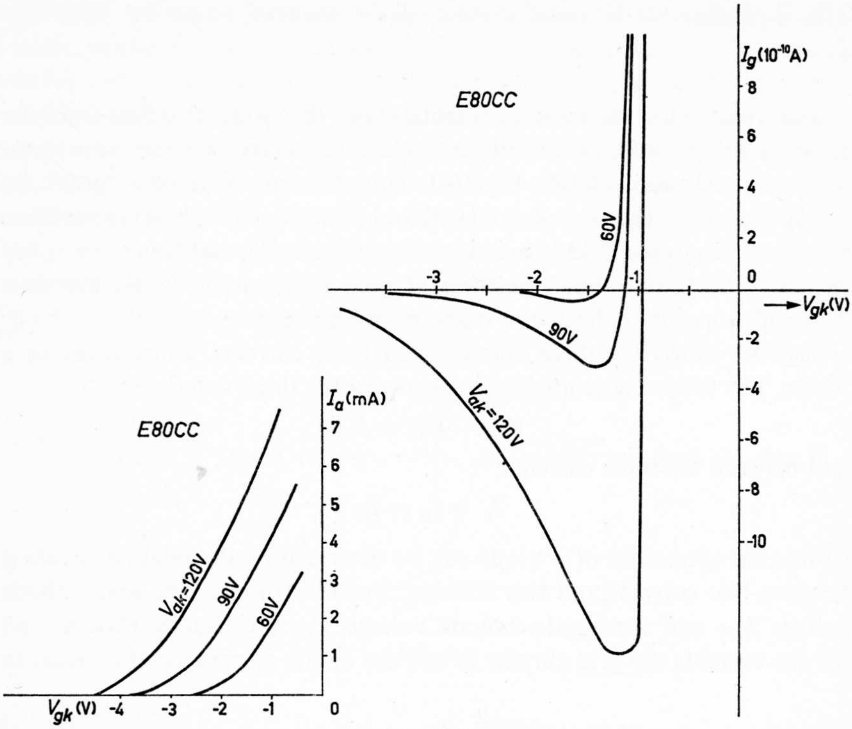

Fig. 10-2

and 10^{-8} A. This is explained by the fact that the grid current is made up of several components: 1) A diode current from the "diode" formed by grid and cathode; 2) A current of positive ions produced during the ionization of the residual gas in the valve by the electrons forming the anode current; 3) A leakage current through the insulator between anode and grid. The "diode current" has the usual characteristic, and the leakage current is mainly determined by the insulation of the valve base. On the other hand, the magnitude of the ion current is closely connected with the quality of the vacuum, which may vary greatly for different valve types. It is important in many applications to operate the valve so that the grid current is small, which means that the negative grid-cathode voltage must be more than about -1.5 volt.

Although there is a point at which the grid current is zero, the influence of changes in the grid voltage on the grid current is there very great.

A closer look at the $I_a - V_{gk}$ curves of triodes shows (at least within certain limits) that it is possible to make the curves for various anode-cathode voltages almost coincide by a displacement parallel to the V_{gk} axis; this displacement is in the first instance proportional to the change in anode-cathode voltage. In other words: we regain the same anode current if we combine a change in anode-cathode voltage, ΔV_{ak}, with a change in grid-cathode voltage $\Delta V_{gk} = -\Delta V_{ak}/\mu$, where μ is almost constant over a large part of the operating region which interests us. This constant is called the "amplification factor" of the valve, for reasons which will be given later.

The amplification factor μ is mainly determined by the geometry of the valve design, namely the mesh width of the grid and the distances of the grid to anode and cathode. The value of μ for most triodes lies between 15 and 60. Extreme values found in some existing types are about 2 and 100.

There will be no change in the anode current if:

$$\Delta V_{gk} + \frac{\Delta V_{ak}}{\mu} = 0$$

or

$$V_{gk} + \frac{V_{ak}}{\mu} = \text{constant.}$$

I_a can thus be written as a function of $V_s = V_{gk} + V_{ak}/\mu$, the "control voltage" of the valve:

$$I_a = f(V_s) \tag{10.1}$$

The relation between grid current I_g and voltages V_{gk} and V_{ak} is more complicated and cannot be expressed as a function of one simple variable. However, I_g, and especially the dependence of I_g on the voltages, is of little importance for most applications. Nevertheless, we shall meet cases where the grid current has to be taken into account.

11. Amplification

An important property of the triode is that it can amplify voltages, that is it can produce a new voltage showing the same proportional variations as the original signal, but having greater absolute values. Fig. 11-1 gives an example of an amplification circuit.

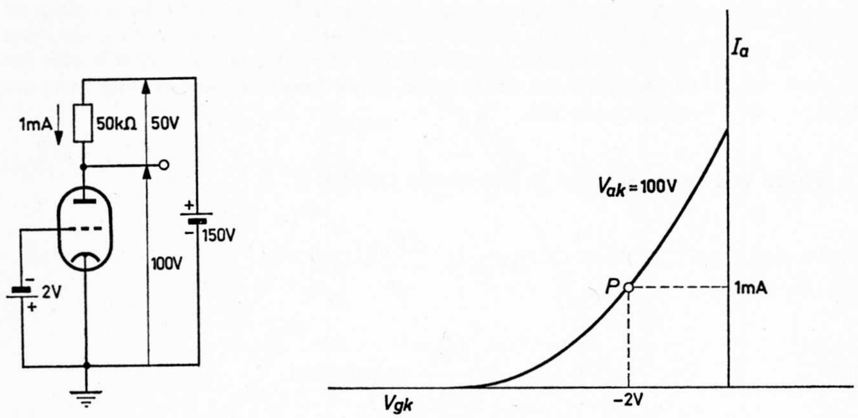

Fig. 11-1

At the selected voltages V_{gk} (−2V) and V_{ak} (100V), a given anode current I_{a0} (1 mA) flows through the valve; this is the "steady-state, standing or quiescent condition" and we say that the valve has been adjusted to a given "working point" (P).

When active elements are being used, one must accept the fact that d.c. voltages and currents are present in the standing or quiescent condition. This follows from the fact that the energy supplied to the system by these elements is derived from these voltages and currents. The signal voltages and currents are deviations from this quiescent condition.

An increase in the grid-cathode voltage will produce an increase in the anode current and hence a decrease in the anode-cathode voltage, because of the presence of a resistor in the anode circuit. This decrease in anode-cathode voltage can be made to be larger than the increase in the grid-cathode voltage. In a given circuit the amount of this "amplification" depends on the relation between the anode current and control voltage as given in (10.1): $I_a = f(V_s)$. The amplification can be calculated by expanding this relation in the vicinity of the working point into a power series convergent to $V_s - V_{s0}$, where V_{s0} is the value of the control voltage at the working point. We thus obtain:

$$I_a = I_{a0} + \left(\frac{\mathrm{d}f}{\mathrm{d}V_s}\right)_{V_{s0}} \cdot (V_s - V_{s0}) + \frac{1}{2!}\left(\frac{\mathrm{d}^2 f}{\mathrm{d}V_s^2}\right)_{V_{s0}} \cdot (V_s - V_{s0})^2 + \\ + \dots \qquad (11.1)$$

Dependent on the required accuracy and the value of $V_s - V_{s0}$, this series may be terminated after a certain number of terms. The need for amplification is, of course, greatest for small signals and the following linear form is therefore adequate for most cases ($V_s - V_{s0} < 0.5$ V):

$$I_a = I_{a0} + \left(\frac{\mathrm{d}f}{\mathrm{d}V_s}\right)_{V_{s0}} \cdot (V_s - V_{s0}) \qquad (11.2)$$

For a given valve and working point the derivative $(\mathrm{d}f/\mathrm{d}V_s)_{V_{s0}}$ is a constant usually denoted by g_m or S. We shall use the latter notation.

$$I_a - I_{a0} = S(V_s - V_{s0})$$

or

$$\Delta I_a = S\Delta V_s = S\left(\Delta V_{gk} + \frac{\Delta V_{ak}}{\mu}\right) \qquad (11.3)$$

where ΔI_a is the anode signal current and ΔV_s the signal control voltage.

In the following pages we shall be concerned mainly with signal voltages and currents. We shall omit the Δ-sign, but write i_a for the anode signal current, v_{gk} for the grid cathode signal voltage, etc. From now on capital letters I and V will be used exclusively for quiescent currents and quiescent voltages.

We can now write for (11.3)

$$i_a = S\left(v_{gk} + \frac{v_{ak}}{\mu}\right) \qquad (11.4)$$

This relation is known as the triode equation. The quantity S is called the slope or trans-conductance of the valve at the working point and has the dimension of a conductance: ohm^{-1} (mho). It is usually expressed in mA/V.

The triode equation can also be derived without using the constant μ by considering I_a as a function of the two variables V_{gk} and V_{ak}, expanding the latter into a power series terminated after the linear term. This method is in any case necessary for deriving the corresponding relation for the grid current i_g.

When $v_{gk} = 0$, that is when there is no signal voltage between grid and cathode, we have $i_a = S/\mu \cdot v_{ak}$, and the valve will therefore behave for signals between anode and cathode like a resistance of value μ/S. This is called the anode or plate resistance r_a of the valve. We thus have:

$$\mu = Sr_a \tag{11.5}$$

which is known as Barkhausen's relation.

The amplification factor is almost constant over the whole working range and it does not change appreciably during the life of the valve. On the other hand the mutual conductance S and the anode resistance r_a can change quite considerably in the course of time. One sometimes finds valves in radio receivers where, at fixed voltages, the mutual conductance has been reduced by a factor of 2. The anode resistance is then increased by the same factor of 2.

For many practical calculations it is convenient to multiply the triode equation by r_a and to write:

$$r_a i_a = \mu v_{gk} + v_{ak} \tag{11.6}$$

Furthermore it is often cumbersome to measure the voltages with respect to the cathode, particularly in circuits with more than one valve, where the various cathodes have different potentials. Here it is almost inevitable that we relate the voltages to a single reference potential, often the earth potential. We shall therefore denominate this reference potential as "earth". Without indication to the contrary, the signal voltage will always be assumed to be positive with respect to ground.

Using v_a, v_g and v_k respectively, for the anode, grid and cathode voltages with respect to earth we obtain for the triode equation:

$$i_a = Sv_g + \frac{S}{\mu}v_a - S\left(1 + \frac{1}{\mu}\right)v_k \tag{11.7}$$

or

$$r_a i_a = \mu v_g + v_a - (\mu + 1)v_k \tag{11.8}$$

An example of a circuit with only one valve to which this last equation can be applied is given in Fig. 11-2. A resistor R_k is inserted in the cathode lead of the valve, across which the quiescent current of the valve gives the desired grid-cathode d.c. voltage. (We shall discuss the adjustment of valves to the correct working point in Section 13.)

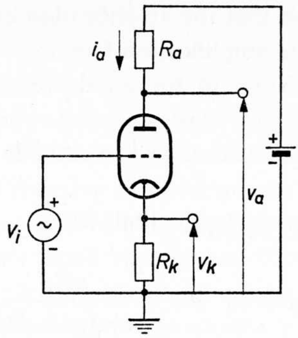

Fig. 11-2

Applying a signal voltage v_i to the grid we have the following equations for this circuit:

$$v_a = -R_a i_a; \quad v_k = R_k i_a$$

When substituted in (11.8) we obtain:

$$r_a i_a = \mu v_i - R_a i_a - (\mu + 1) R_k i_a$$

or

$$[r_a + R_a + (\mu + 1) R_k] i_a = \mu v_i \tag{11.9}$$

Therefore:

$$v_a = \frac{-\mu R_a}{r_a + R_a + (\mu + 1) R_k} v_i \tag{11.10}$$

$$v_k = \frac{\mu R_k}{r_a + R_a + (\mu + 1) R_k} v_i \tag{11.11}$$

It is interesting to see how these voltages v_a and v_k depend on the resistor values while the valve retains the same working point by adjusting the supply voltages accordingly. We then find for the amplification from grid to anode $A_a = v_a/v_i$ that it

1. is negative;
2. increases in absolute value with decreasing R_k, the limit being

$$A_a(R_k = 0) = \frac{-\mu R_a}{r_a + R_a}$$

3. increases in absolute value with increasing R_a, the limit being

$$A_a(R_a = \infty) = -\mu$$

μ is therefore the limit that the amplification can reach in this configuration; hence the name amplification factor;
4. is less sensitive to changes in the anode resistance of the valve with increasing value of $R_a + (\mu + 1)R_k$.

We find for the amplification from grid to cathode $A_k = v_k/v_i$ that it
1. is positive;
2. increases with decreasing R_a, the limit being

$$A_k(R_a = 0) = \frac{\mu R_k}{r_a + (\mu + 1)R_k}$$

3. increases with increasing R_k, the limit being

$$A_k(R_k = \infty) = \frac{\mu}{\mu + 1} \approx 1 - \frac{1}{\mu}$$

The amplification to the cathode is hence always smaller than 1. Nevertheless, the "cathode follower", that is a circuit where the cathode is used as output terminal is often applied. We shall discuss the main reason for this in the next section.

12. Output impedance

When deriving the superposition theorem and Thévenin's theorem, we made use of the linearity of the components, but not of their nature. These theorems thus also apply if active components are included in the circuit, provided they behave linearly. We have already seen that this is the case with triodes for sufficiently small signals, and thus for such signals both theorems can also be applied to circuits containing triodes. It is obvious that when we speak of eliminating voltage sources, we should not assume the supply voltages to be eliminated; the valve must remain at its working point.

Let us now consider Fig. 11-2 where the voltage is taken between anode and earth. These points may therefore be considered as the terminals of a voltage source. The source voltage equals the open-circuit voltage, given by (11.10). The short-circuit current is derived from (11.9) by putting $R_a = 0$ (for signal currents and voltages the battery can be replaced by a short circuit, so that R_a is then in parallel with the output terminals):

$$-i_a(R_a = 0) = \frac{-\mu v_i}{r_a + (\mu + 1)R_k}$$

The internal impedance is found from the quotient:

$$Z_a = \frac{R_a[r_a + (\mu + 1)R_k]}{R_a + r_a + (\mu + 1)R_k}$$

Because of the output nature of the pair of terminals, this internal impedance is also called output impedance; as seen from the equation it consists of the parallel combination of R_a and a resistance $R_p = r_a + (\mu + 1)R_k$.

However, Fig. 11-2 also shows that Z_a corresponds to the parallel combination of R_a and the resistance of the anode through the valve to earth. This latter value is therefore $R_p = r_a + (\mu + 1)R_k$ (see Fig. 12-1). If $R_k = 0$, R_p will equal the anode resistance r_a of the valve analogous to what we found previously.

When writing

$$r_a + (\mu + 1)R_k = (\mu + 1)\left(\frac{r_a}{\mu + 1} + R_k\right) \approx$$

$$\approx (\mu + 1)\left(\frac{r_a}{\mu} + R_k\right) = (\mu + 1)\left(\frac{1}{S} + R_k\right)$$

we see that for normal conditions ($1/S = 100 - 1000\ \Omega$ with an extreme value of approx. 20 Ω in the case of very high-slope valves such as the E810F = 7788), R_k need not be made very large in order to let R_p be mainly determined by $(\mu + 1)R_k$ and acquire very large values.

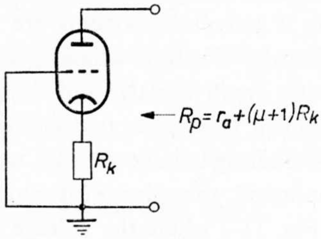

Fig. 12-1

We take advantage of this property when a high resistance is required for the signal currents, but not for the quiescent currents. For example, in the voltage divider in Fig. 12-2 signal v_i appears at the anode with a loss of only three per cent, because the combined resistance of the valve and the cathode resistor amounts to approximately 3 MΩ (28 × 100 kΩ + 30 kΩ). We would need a negative voltage of approximately 3,000 volts to obtain the same result with an ordinary resistance voltage divider. This circuit can therefore be used to bring a signal to a lower quiescent voltage level, if it not possible to apply indirect coupling by means of capacitors.

The ratio of the resistances for the signal and to quiescent currents becomes still greater in the twin valve circuit shown in Fig. 12-3. In this case, the upper valve has a resistance in its cathode circuit which is the same as the above calculated 3 MΩ. The internal resistance of the upper anode has now a value of approx. 80 MΩ (28 × 3 MΩ), but the d.c.resistance is only 300V ÷ 1 mA = 300 kΩ. The loss in signal is now only approximately 0.1 per cent.

Let us now consider the case that the cathode is used as output. The open-circuit voltage is given by (11.11), and the short-circuit voltage follows from (11.9) by putting $R_k = 0$:

$$i_a(R_k = 0) = \mu v_i/(r_a + R_a)$$

which makes the output impedance:

$$Z_k = \frac{R_k(r_a + R_a)}{r_a + R_a + (\mu + 1)R_k} = \frac{R_k \cdot \dfrac{r_a + R_a}{\mu + 1}}{R_k + \dfrac{r_a + R_a}{\mu + 1}}$$

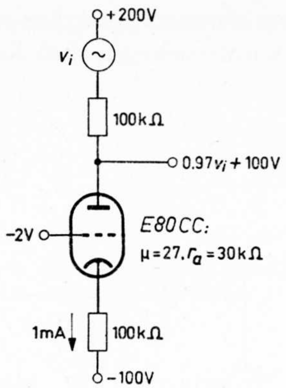

Fig. 12-2

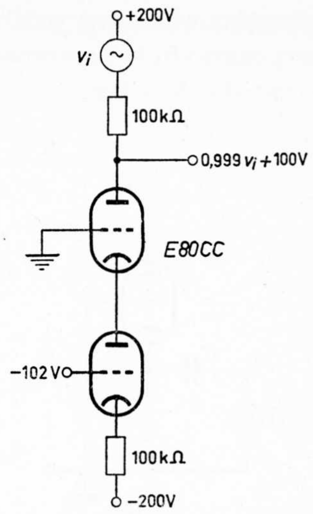

Fig. 12-3

We recognize in this equation the parallel combination of R_k and a resistance $(r_a+R_a)/(\mu+1)$. The latter resistance is therefore the impedance of the valve viewed from the cathode (see Fig. 12-4, where the d.c. supplies have been omitted).

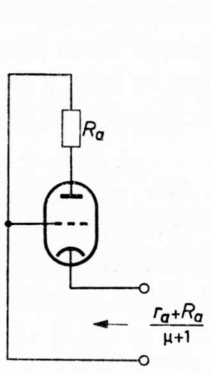

Fig 12-4

As we have seen that the largest amplification occurs for $R_a=0$, and that the output impedance is then at its smallest, this circuit is normally used without a resistor in the anode circuit. The output impedance will then be $r_a/(\mu+1) \approx r_a/\mu = 1/S$, which in normal operation is between 100 and 1,000 Ω. This means that the cathode follower does not produce any amplification of the signal voltage, but offers the facility of transforming a voltage source with a large output resistance into a voltage source with virtually the same open-circuit voltage but with a very small output resistance. It is thus possible to extract more current from the terminals while the voltage only differs slightly from the open-circuit voltage.

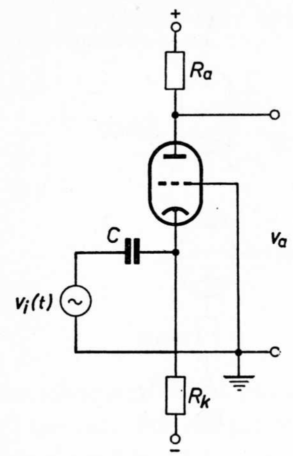

Fig. 12-5

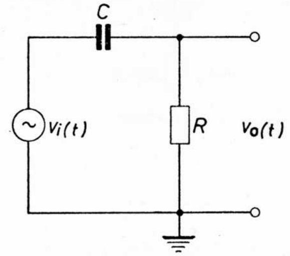

Fig. 12-6

Fig. 12-5 shows an application of the low impedance on the cathode side of the valve where, analogous to that of Fig. 12-2, no control on the grid occurs. If it is required to obtain a voltage which is proportional to the derivative of another voltage, with respect to time, it is theoretically possible to use the circuit shown in Fig. 12-6. The relevant equation is

$$v_o(t) + RC\frac{\mathrm{d}v_o(t)}{\mathrm{d}t} = RC\frac{\mathrm{d}v_i(t)}{\mathrm{d}t}$$

By making the product RC small enough, it is possible to neglect the second term on the left-hand side of the equation with respect to the first term:

$$v_o(t) \approx RC\frac{\mathrm{d}v_i(t)}{\mathrm{d}t}$$

$v_o(t)$ is then actually proportional to the derivative of $v_i(t)$, but the ratio is small. It is of course possible to amplify $v_o(t)$, but often better results can be achieved with the circuit shown in Fig. 12-5. This circuit can be considered as the result of replacing R of Fig. 12-6 by the cathode impedance of the valve, which has the value $(r_a + R_a)/(\mu + 1)$, provided R_k is large with respect to this value. To obtain the same accuracy in differentiation as in the first case, we must put $R=(r_a+R_a)/(\mu+1)$. The current which previously passed through R, will now flow through the valve and resistor R_a, and gives a voltage across R_a which is larger by a factor of

$$\frac{R_a}{R} = \frac{R_a(\mu+1)}{r_a+R_a} = \frac{\mu+1}{\dfrac{r_a}{R_a}+1}$$

With $R_a > r_a$, this method gives a gain of at least $\frac{1}{2}\mu$.

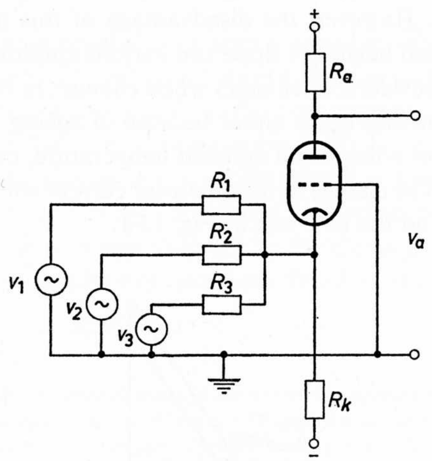

Fig. 12-7

A similar circuit can be used for adding voltages including certain weighting factors. We obtain for the circuit of Fig. 12-7 as a good approximation:

$$v_a = -R_a\left(\frac{v_1}{R_1}+\frac{v_2}{R_2}+\frac{v_3}{R_3}\right)$$

provided that R_1, R_2, R_3 and $R_k \gg (r_a+R_a)/(\mu+1)$.

We shall see later on that the cathode follower possesses other favourable properties. It is therefore understandable that this circuit is very popular, despite the fact that the voltage amplification is smaller than unity. We shall discuss examples of these applications later.

13. Selecting the working point

Before discussing various aspects of amplification, we should first look at the ways in which the valve can be made to work. In order to place the valve at the selected working point, a particular grid-cathode bias voltage is required for a given anode-cathode voltage. For example, a negative grid-cathode bias voltage of 2 volts is required in Fig. 13-1 to obtain a quiescent current of 1 mA at $V_{ak} = 100$ V (point P). This can be achieved as indicated in Fig. 13-2 where the grid-cathode voltage is independent of any other voltages. However, the disadvantage of this method is that the quiescent current, and hence the slope and various amplification characteristics can undergo considerable changes when changes in the emission occur. Changes in emission can occur either because of ageing of the valve or by changes in the heater voltage and ambient temperature, or because the valve has been replaced. The change in the quiescent current will then be from P on the old curve to B on the new one in Fig. 13-1.

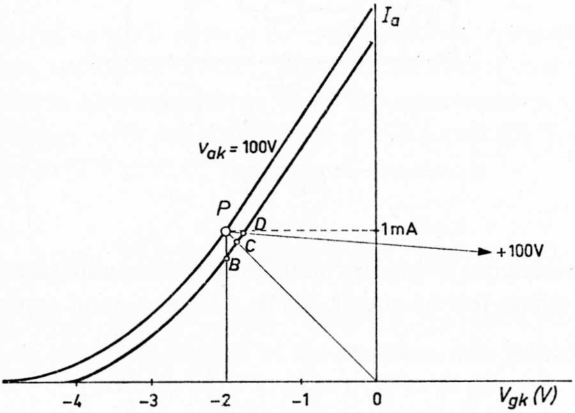

Fig. 13-1

The negative grid-cathode bias voltage can also be obtained by inserting a resistor R_k of the correct value (several hundred to several thousand ohms) in the cathode lead (see Fig. 13-3). This results in a relation $V_{gk} = -I_a R_k$ between this bias voltage and the anode current, represented by line OP. In this case, with the same change of the curve, the working point shifts from P to C and the change in the quiescent current is smaller than in the first case. The improvement becomes considerably greater when connecting

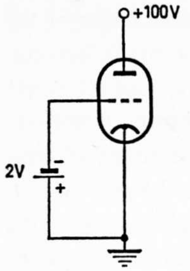

Fig. 13-2

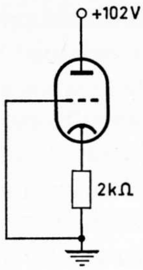

Fig. 13-3

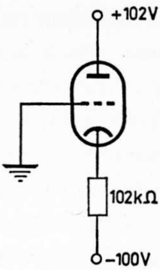

Fig. 13-4

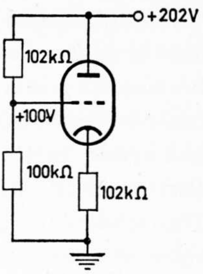

Fig. 13-5

the cathode via a large resistor to a highly negative voltage, e.g. -100 V (see Fig. 13-4 and 13-5). The relation between the bias voltage and the anode current then becomes: $V_{gk} = 100\text{ V} - I_a R_k$, represented in Fig. 13-1 by the line $P \rightarrow +100$ V, so that the working point shifts from P to D when the curve is changed. The change in the anode current is now only very small; the quiescent current is largely determined by the values of the cathode resistor and the negative voltage, and not by the characteristics of the valve. We shall see later how this type of circuit can be adapted to the amplification of signal voltages.

Regarding the changes in the valve characteristics, we should note that in the case of normal heater voltages, the emission decreases only very slowly as the valve ages. Typical values are the following: $10^{-2} - 10^{-3}$ % per hour.

Heat losses of the cathode occur both by radiation and conductance. With the usual types of valves with a power of $1-2$ W for the filament heating, both losses are of the same magnitude. Since the radiation losses vary as the difference between the fourth powers of the cathode and ambient temperature, while the conductance losses are proportional to the difference between these temperatures, a change in ambient temperature will mainly affect the conductance losses. A change of 1° C in the ambient temperature was found to correspond to a change in grid voltage of between 0.3 and 1 mV.

Variations in the heater voltage have of course a great effect on the emission. A change of 10 per cent in the heater voltage corresponds to a change in grid voltage of between 100 and 150 mV.

We shall refer to this behaviour when discussing d.c. amplifiers, where these phenomena are most evident.

The working point should be selected such that the various maximum ratings, indicated by the manufacturer, are not exceeded. In the case of a linear amplifier, these ratings refer, amongst others, to anode voltage, anode dissipation, cathode current and voltage between heater and cathode.

Here we should draw attention to the maximum permissible value of the resistor in the grid circuit, which is usually indicated by the manufacturer. Because the grid is close to the cathode, some emitting material can relatively easily be deposited onto the grid. If, by means of radiation from the cathode and anode, the grid becomes sufficiently hot, it will commence to emit electrons itself, which will also pass through the resistor into the grid circuit. This renders the grid less negative, with the result that the current in the valve increases. This may produce a greater anode dissipation and hence higher anode temperature, thus increasing the grid temperature still further. In this way, thermal runaway can destroy the valve.

The maximum value of the grid resistor indicated by the manufacturer refers to the case where the current through the valve is not stabilized (Fig. 13-2), while the anode dissipation is at maximum. In all cases where stabilization is present and/or the anode dissipation is smaller than the maximum value, this value is therefore certainly not critical.

14. Signal amplitude

The triode equation was derived assuming the condition that the signal voltages were so small that the higher-power terms in the expansion of i_a in terms of v_s: $i_a = Sv_s + \beta v_s^2 + \gamma v_s^3 + \ldots$, could be neglected. How far this condition is satisfied depends on the desired accuracy, quite apart from the value of the coefficients S, β, γ, etc., and of the control voltage v_s. If the working point of the valve is chosen approximately in the middle of the linear part of the $I_a - V_{gk}$ characteristic, the effect of β, γ, etc. will be small in comparison with that of S. The accuracy will then be adequate for most applications, provided the control voltage remains restricted to a few tenths of a volt.

In order to gain an impression of the amplitude of input signals that can be dealt with in a linear fashion by the circuits shown in Figs 13-2--5, we can calculate the control voltage $v_s = v_{gk} + v_{ak}/\mu$ by means of equations (11.10) and (11.11). We thus find:

$$v_s = \frac{r_a}{r_a + R_a + (\mu + 1)R_k} \cdot v_i$$

and we can conclude that the relation between the permissible input voltage $(v_i)_{\max}$ and the permissible control voltage $(v_s)_{\max}$ is given by:

$$(v_i)_{\max} = \frac{r_a + R_a + (\mu + 1)R_k}{r_a} \cdot (v_s)_{\max} \tag{14.1}$$

It will be noted that the effect of R_k on $(v_i)_{\max}$ is very great.

When we now take as an example the valve E80CC ($\mu = 27$, $r_a = 20\ \mathrm{k}\Omega$) in two different circuits (Figs 14-1 and -2) it appears, assuming a value of 0.5 volt for $(v_s)_{\max}$, that equation (14.1) will lead in the first case to a permitted input signal of approx. 5 volts, and in the second case to one of approx. 70 volts.

Since a control voltage v_s produces a current Sv_s through the valve, the permissible output voltages will be

$$(v_a)_{\max} = SR_a(v_s)_{\max} \quad \text{and} \quad (v_k)_{\max} = SR_k(v_s)_{\max}$$

These voltages are therefore the same for the circuits shown in Figs 14-1 and -2.

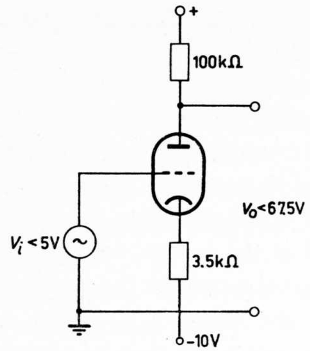

Fig. 14-1

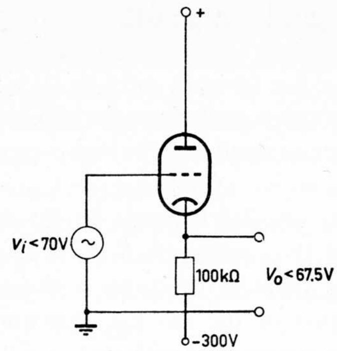

Fig. 14-2

The "paraphase amplifier" (Fig. 14-3) is an example of a circuit in which both the anode and the cathode voltages are used as output signals. Here the anode and cathode resistors, have the same value, so that the two output voltages are also the same but of opposite signs. Provided the resistor values are sufficiently large both will approximately equal the input voltage v_i, and the permissible output voltages will be virtually the same as those produced with the cathode follower circuit of Fig. 14-2.

Of course all this only applies if the supply voltages allow for such large signals. In Fig.14-3 this is not the case: there a signal voltage of 10–20 volts would be permitted.

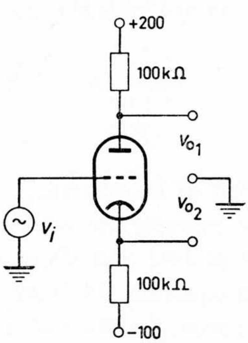

Fig. 14-3

We can therefore conclude that – as far as distortion is concerned – circuits with a large cathode resistor can cope with large signals. We shall see in Section 23 that the actual situation is even more favourable than represented here. This is because a large cathode resistor produces feedback which results in a smaller distortion for the same signal size.

Anticipating the discussion of the frequency-dependence of amplification, we should also mention the necessity of replacing the various resistances in the formulae by the corresponding impedances in the case of a.c. voltages. For example, when cathode resistors are used in a circuit, this means that a possible parasitic cathode capacitance in parallel to R_k will decrease the value of $(v_i)_{max}$ at high frequencies. At very high frequencies the cathode impedance will be solely determined by such a capacitance and may become very small. According to (14.1), $(v_i)_{max}$ will then be of the same order of magnitude as $(v_s)_{max}$ for not too large values of R_a. This obviously means a considerable limitation, particularly so for the cathode follower.

15. A.C. amplifiers

We have seen that d.c. voltages and currents are always present in valves in the quiescent state. This also applies to amplifier circuits using these components. In a broad sense, an amplifier is therefore a device with input and output terminals, in which an alteration in the quiescent state at the input, causes usually a greater alteration expressed in the quiescent state at the output. Especially in the case of small signals, difficulties will then become immediately evident, because even relatively small variations in the quiescent situation, (e.g. as a result of changes in emission) can be of the same magnitude as these signals or even larger, which makes it difficult or impossible to distinguish between them. Fortunately it often happens that one is not interested in deviations from the quiescent situation but solely in the amplitude of changes of an appreciable frequency, and in consequence the level on which they take place is irrelevant. It is possible in this case to make considerable simplifications in the design of the amplifying mechanism. Amplifiers designed in this manner are called a.c. amplifiers, and amplifiers where the absolute level is also amplified are called d.c. amplifiers. The latter therefore not only amplify d.c. voltages, which would not be of much interest, but also a.c. voltages. We shall later refer to the specific problems encountered with d.c. amplification. Let us first discuss the many complications which arise already with a.c. amplifiers.

In the case of a.c. amplifiers we thus deal with the amplification of voltage changes. Dependent on the velocity at which these changes occur, signals must be amplified, whose frequencies lie between certain limits, the "frequency bandwidth". We can make further subdivisions of a.c. amplifiers according to the nature of this bandwidth (loosely: low, middle, high, very high, etc. frequencies) and to its relative or absolute width (selective, broad-band amplifiers). We shall return to these subdivisions later.

When using the triode as the amplifying element in the circuit of Fig. 15-1, where the anode voltage is used as output signal, the amplification $A_a = v_o/v_i$ will be rather small, especially at large values of R_k:

$$A_a = -\frac{\mu R_a}{r_a + R_a + (\mu + 1)R_k} \tag{15.1}$$

The effect of R_k can be virtually eliminated for a.c. voltages in a certain frequency band by inserting a capacitor parallel to R_k. We thus obtain the

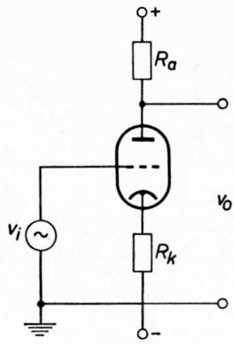

Fig. 15-1

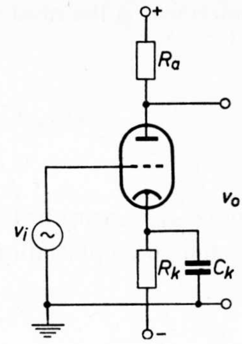

Fig. 15-2

circuit shown in Fig. 15-2. The amplification can be calculated from the above equation by substituting Z_k for R_k, when Z_k is the impedance of the parallel combination of R_k and C_k. We have for a frequency ω:

$$Z_k = R_k \,//\, \frac{1}{j\omega C_k} = \frac{R_k}{1 + j\omega R_k C_k}$$

It is thus possible to reduce Z_k in absolute value as much as required by increasing the value of C_k. The maximum amplification which can be obtained with a given μ, r_a and R_a is, for $Z_k = 0$:

$$-\frac{\mu R_a}{r_a + R_a}$$

If it is necessary to approach this value, equation (15.1) shows that $(\mu+1)Z_k$ should be small compared with $r_a + R_a$ for all frequencies in the band, that is

$$\left| Z_k \right| \ll \frac{r_a + R_a}{\mu + 1}$$

Except for the case that $R_k \ll (r_a+R_a)/(\mu+1)$, when no capacitor is necessary, C_k must therefore satisfy:

$$\frac{1}{\omega C_k} \ll \frac{r_a + R_a}{\mu + 1} \quad \text{or} \quad \omega C_k \gg \frac{\mu + 1}{r_a + R_a}$$

for all frequencies in the band and therefore for the lowest frequency ω_{min} where the requirement is the most stringent:

$$\omega_{min} C_k \gg \frac{\mu + 1}{r_a + R_a}$$

How many times $\omega_{min}C_k$ must be larger than $\varrho^{-1} = (\mu + 1)/(r_a + R_a)$ is determined by the required amplification A_a. If R_k is large, we can write for this amplification:

$$A_a \approx \frac{-R_a}{\varrho + \dfrac{1}{j\omega C_k}}$$

If $\omega_{min}C_k$ is very much larger than ϱ^{-1}, the amplification at ω_{min} will hardly have changed either in magnitude or phase. If we are only interested in the magnitude of the amplification, the form $\sqrt{\varrho^2 + (1/\omega^2_{min}C_k^2)}$ will be the determining factor. At the assumed large values of $\omega_{min}C_k$, this can be approximated by $\varrho(1 + 1/2\varrho^2\omega^2_{min}C_k^2)$. The reduction is then inversely proportional to the square of C_k.

The condition $\omega_{min}C_k \gg 1/R_k$ which is found in some books is incorrect.

Example: Consider an audio amplifier ($\omega_{min} \approx 100$ rad/s) using an E80CC valve ($\mu = 27$, $r_a = 20$ kΩ), an anode resistor of 50 kΩ and a cathode resistor of 100 kΩ (connected to -100 V). $\omega_{min}C_k \gg (\mu + 1)/(r_a + R_a)$ yields: $100\ C_k \gg 28/(20 + 50).\ 10^3$ or $C_k \gg 4\mu$F.

$\omega_{min}C_k \gg 1/R_k$ would yield: $100\ C_k \gg 1/10^5$ or $C_k \gg 0 \cdot 1\mu$F.

If the requirement for C_k is satisfied, one might expect that the amplification produced by the circuit of Fig. 15-2 would be almost constant for all frequencies above ω_{min}. However, the actual behaviour is illustrated in Fig. 15-3: the amplification begins to decrease above a certain frequency ω_{max}. At still higher frequencies, this decrease is inversely proportional to the frequency. The reason for this decrease is that the anode resistor is shunted capacitively: the anode and its connections form capacitors with the other electrodes, the surrounding earth, and the supply source leads. The effect of these capacitors can be imagined as being allowed for in the

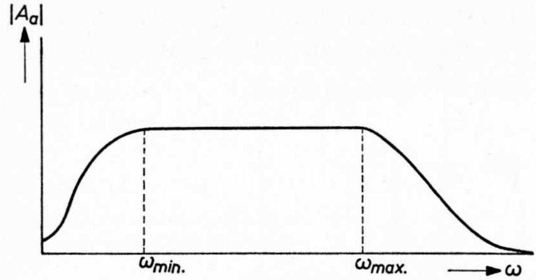

Fig. 15-3

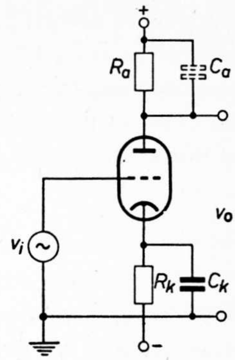

Fig. 15-4

first instance by a "parasitic" capacity C_a to ground (Fig. 15-4). It is then necessary to replace R_a by

$$Z_a = R_a \,//\, \frac{1}{j\omega C_a} = \frac{R_a}{1 + j\omega R_a C_a}$$

and the amplification will then be:

$$A_a = \frac{-\mu Z_a}{r_a + Z_a} = \frac{-\mu R_a}{r_a + R_a + j\omega r_a R_a C_a} =$$

$$\frac{-\mu R_a}{r_a + R_a} \cdot \frac{1}{1 + j\omega \dfrac{r_a R_a}{r_a + R_a} C_a} \qquad (15.2)$$

The decrease becomes noticeable for those frequencies where $\omega C_a r_a R_a / (r_a + R_a)$ approaches unity.

In calculations the product of resistance and capacitance often occurs. It has the dimensions of time and determines the duration of a phenomenon; it is called the time constant and is usually designated by the symbol τ.

In the above, the time constant is $C_a r_a R_a/(r_a + R_a)$, i.e. the product of C_a and the value of r_a and R_a in parallel.

A normal value for C_a is 10–30 pF, dependent on the type of valve and on the method of wiring. Long leads considerably increase C_a, especially when they run close to earthed points. But even with the best arrangement, we still have to contend with the capacitance of the anode itself with respect to the other electrodes and immediate surroundings; this alone usually amounts to several picofarads. It is then no longer possible to achieve an increase in ω_{max} by reducing C_a, but only by reducing R_a. However, this also decreases the amplification in the region between ω_{min} and ω_{max}, where the effect of the capacitance is not noticeable. If we call the absolute value of this amplification A_{a0}, we have according to (15.2):

$$A_{a0} = \frac{\mu R_a}{r_a + R_a}$$

Since ω_{max} is proportional to $(r_a+R_a)/r_aR_aC_a$, we find that the product $A_{a0}\omega_{max}$ does not depend on r_a and R_a, but solely on

$$\frac{\mu R_a}{r_a + R_a} \cdot \frac{r_a + R_a}{r_aR_aC_a} = \frac{S}{C_a}$$

This is the product of amplification and the maximum bandwidth obtainable, which has a constant value for a valve and is proportional to S/C_a. One should therefore select valves with a great mutual conductance and small anode capacitance for the amplification of wide frequency bands.

If it is necessary to amplify over a prescribed frequency band, and a single valve does not supply the required amplification, the process must be repeated by placing several "stages" in series. Fig. 15-5 shows a two-stage amplifier circuit.

Here the input of the second stage is not directly coupled to the output of the first stage. Although this would be possible, the inherent disadvantage of this would be that the d.c. voltage level of the second stage would be higher than that of the first stage. Progressing in this manner would lead to even higher d.c. voltage levels, which would necessitate high supply voltages. The output signal is transmitted through capacitor C, which has the advantage that the input of the second stage can be at a different d.c. potential than the output of the first stage, for example once again at earth potential. This is achieved by inserting resistor R_g. To obtain a good transfer in the frequency band, the impedance of C must be small compared to R_g. This requirement is once again most stringent at the lowest frequency. Therefore:

$$\left|\frac{1}{j\omega_{min}C}\right| \ll R_g \quad \text{or} \quad C \gg \frac{1}{\omega_{min}R_g}$$

If this condition has been satisfied, the impedance of C will be small for all frequencies in the band, and the input of the second stage will then be connected almost directly to the output of the first stage. In other

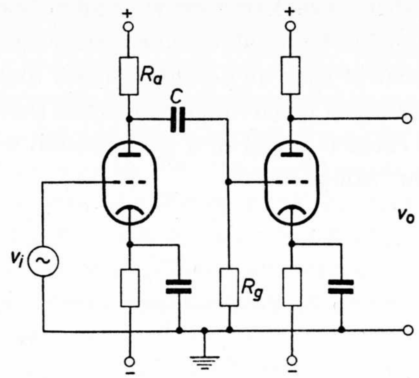

Fig. 15-5

words, R_g will be viıtually parallel to R_a for these frequencies. Because of this, R_g is usually selected large compared to R_a. Since the "coupling" of the two stages occurs by means of a "coupling capacitor" and a resistor, we speak of resistor-capacitor coupling, abbreviated to *RC*-coupling. In order to be independent of a possible d.c. voltage level from the signal source connected to the input of the amplifier, it is normal to use such an *RC*-coupling for the first stage as well (Fig. 15-6). Whilst in the case of "direct" coupling without a capacitor a certain amplification of d.c. voltage signals will occur, this is not so with indirect coupling, so that the effect of slow changes on the point of operation is eliminated.

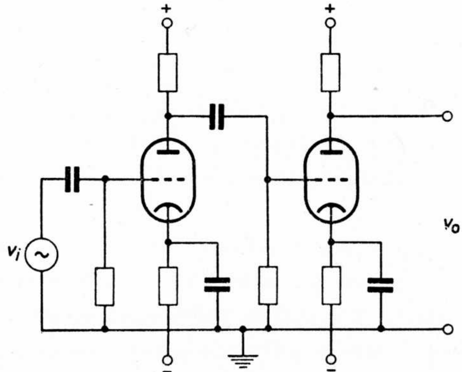

Fig. 15-6

One would expect that for each frequency the amplification of a two stage amplifier would be equal to the product of each stage considered separately. However, measurement of such an amplifier shows that the amplification already begins to decrease at much lower frequencies than is consistent with this expectation. Its cause is found in a phenomenon which has not been considered so far, the "Miller effect".

16. Miller effect

Not only does the anode of a triode show capacitance with respect to the other electrodes, but the grid also possesses this property and has, amongst others, a capacitance C_{ag} to the anode, and a capacitance C_{gk} to the cathode. Once again, both capacitances are of the order of a few picofarads. The result is that the grid will drain current from an a.c. signal voltage source to which it is connected. These capacitances therefore contribute to the input admittance of the circuit containing the valve, but not to the same extent, as we shall see.

In order to calculate the input impedance we should consider Fig. 16-1, where capacitances C_{ag} and C_{gk} are indicated. Current $i_1 + i_2$ is drawn from signal source v_i, where i_1 represents the current flowing through C_{ag}, and i_2 the one through C_{gk}. We obtain for current i_1: $i_1 = j\omega C_{ag}(v_i - v_a) = j\omega C_{ag}(1 - A_a)v_i$. The same current would be drawn from the signal source

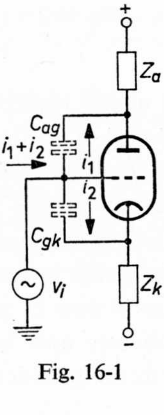

Fig. 16-1

by an impedance $v_i/i_1 = 1/j\omega C_{ag}(1 - A_a)$ to earth. For frequencies in the band to be amplified, A_a is approximately equal to the amplification $-A_{a0}$ calculated in the previous section, so that this impedance corresponds to a capacitance of $(1 + A_{a0})C_{ag}$. This apparent magnification of C_{ag} is called the Miller effect.

For the current i_2 we find: $i_2 = j\omega C_{gk}(v_i - v_k) = j\omega C_{gk}(1 - A_k)v_i$. The value of A_k varies between nearly 1 in the case of a cathode follower and virtually 0 for a circuit with cathode decoupling, so that the corresponding impedance to earth $v_i/i_2 = 1/j\omega C_{gk}(1 - A_k)$ always has a rather high value, and even very high in the case of the cathode follower.

The triode amplification circuit has a large input capacitance $(1+A_{a0})C_{ag}$ which, when successive stages are used, will be parallel to the anode capacitance C_a of the previous stage, thus increasing the effective anode capacitance and correspondingly decreasing the bandwidth.

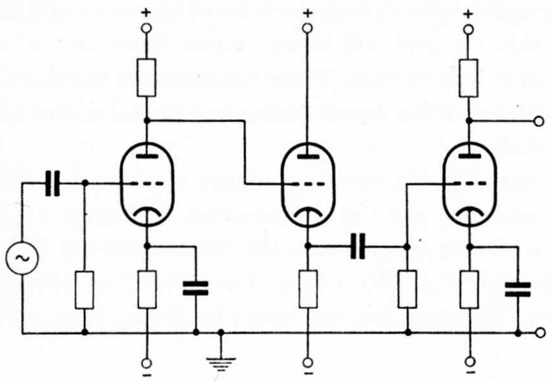

Fig. 16-2

A cathode follower without anode resistor has an input impedance C_{ag} in parallel with a greatly reduced C_{gk}; therefore a much more reduced capacitance value than in the case of the above amplification circuit. If a cathode follower is placed between the two stages of a two-stage amplifier (Fig. 16-2), the Miller effect will no longer be objectionable. The anode capacitance of the first stage is hardly increased at all, and the large input capacitance of the second stage is now in parallel to the very low output impedance of the cathode follower and has therefore very little effect. Indeed, amplification now equals the product of the individual gains of each stage.

Let us finally note that the Miller effect which has here been considered between the first and second stages of an amplifier, can, of course, occur also between the signal source and the first stage when the signal source possesses a sufficiently high internal resistance.

The capacitance C_{ag} of valve ECC81 is approx. 1.5 pF, C_{gk} is approx. 2.5 pF and the capacitance of the grid with respect to points with fixed potentials is approx. 2 pF. When $I_a = 3$ mA in the circuit of Fig. 16-3, S will be 3.75 mA/V, so that, with $\mu = 60$ and $R_a = 33$ kΩ, the amplification at frequencies for which the cathode is sufficiently decoupled will be approx. 40.

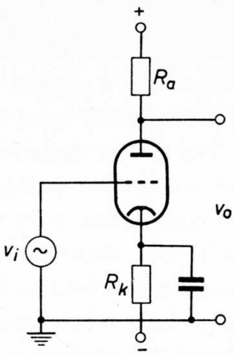

Fig. 16-3

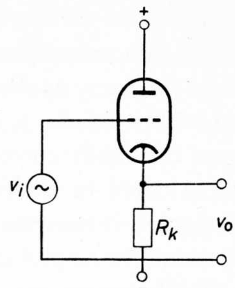

Fig. 16-4

The input capacitance is therefore: $40 \times 1.5 + 2.5 + 2 = 65$ pF, or approximately 70 pF including the wiring capacitance. The output impedance of the anode is $R_a // r_a \approx 11$ kΩ. If two such stages were connected directly, the amplification would already be considerably reduced at a frequency $\omega = (11 \cdot 10^3 \times 70 \cdot 10^{-12})^{-1} = 1.3 \cdot 10^6$ rad/s or $f \approx 200$ kc/s.

In the case of the cathode follower of Fig. 16-4, the input capacitance will amount to approx. 4 pF without the wiring, and to less than 10 pF with the wiring.

We should emphasize that the above-mentioned property of the cathode follower only applies to those frequencies for which A_k is almost entirely real.

At high frequencies, the amplification from the grid to the cathode will have a phase shift due to the cathode capacitance, resulting in an input impedance having both a real and an imaginary part. The real part can even become negative for certain combinations of values, with the inherent danger of parasitic oscillation.

17. Tetrode and pentode

There are two properties of the triode which must be considered as disadvantages for many applications. The first of these is the small gain to be obtained with a triode; μ is rather small, that is the effect of the anode voltage on the anode current is, in comparison with the grid voltage, not as small as could be desired. The second disadvantage is that, although C_{ag} is not great, it nevertheless becomes troublesome because of the Miller effect. It would be helpful if this value were much smaller.

Both disadvantages can be reduced by placing a second grid with a fixed voltage between the grid and the anode. This second grid both reduces the effect of the anode voltage on the valve current and screens the capacitive effect to the grid. It is therefore called the screen grid. In this case, the first grid is called the control grid. The valve with two grids is called a tetrode.

Even if this screen grid is rather wide-meshed and does not have more than 10 per cent "filling", it can drastically reduce the grid-to-anode capacitance, for example, from 5 pF to 0·1 pF. Similarly, the effect of the anode voltage on the electric field in the neighbourhood of the cathode, that is on the space charge, is also reduced. In order to prevent this not only applying to current changes but also to the current itself and therefore to the slope, the loss in the constant part of the anode voltage in this electric field is compensated for by giving the screen grid a fixed, positive voltage with respect to the cathode. The result is that part of the electron flow will be collected by this grid, so that the screen grid will carry a current which, in contrast to the control grid current, forms, an appreciable fraction of the anode current.

The tetrode has an unexpected property which is sometimes useful, but must be considered more often as a serious disadvantage. Measured at fixed control grid voltages and fixed screen grid voltages, the relation between anode voltage and anode current gives curves as shown in Fig. 17-1 (where g_1 = control grid and g_2 = screen grid). This proves that the anode current is reduced over a great part of the normal field of operation when the anode voltage is increased. It may even become negative. This is caused by electrons freed from the anode.

When an electron has just been emitted from the cathode, it has a low velocity and therefore low kinetic energy, but on arrival at the anode, its potential energy is reduced by qV_{ak} (q=charge of the electron) and its kinetic energy has gained the same amount. This means that it arrives at the anode at a higher velocity than it had on leaving the cathode.

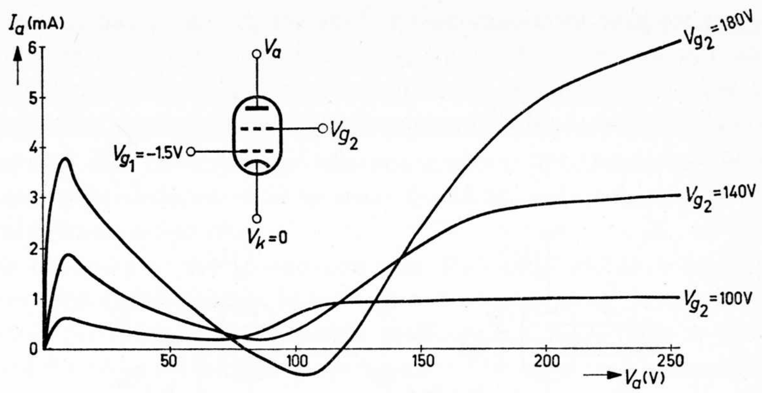

Fig. 17-1

If the anode voltage is low, this kinetic energy at the anode will become available in the form of heat. Above a certain value of the anode voltage, however, the kinetic energy of the electron will become sufficiently large to free further "secondary" electrons from the anode. There may be 3–5 secondary electrons for each primary electron but their average energy is smaller than that of the primary electrons. If the anode voltage is lower than the screen grid voltage, these secondary electrons will be attracted to the screen grid, and since the anode looses more electrons than it gains, the anode current will be negative. The current at the screen grid will be large and positive. If the anode is positive with respect to the screen grid, almost all secondary electrons are re-attracted to the anode and the anode current is then almost independent of the anode voltage, as desired. But, even when the anode material has a smaller secondary emission than assumed above, there is nevertheless always a region where the anode current is very strongly dependent on the anode voltage and also a region where the current will decrease when the anode voltage increases; the anode resistance of the valve is then negative.

Two solutions have been found to eliminate these defects. The first is the beam tetrode where a suitable arrangement of the valve electrodes ensures that the secondary electrons cannot return to the screen grid. However, nowadays the beam tetrode is solely used as an output valve, that is as a valve which is capable of delivering a considerable power, and where the relatively small anode resistance of a few tens of kilo-ohms is an advantage. The second solution is much more important in practice. It consists of the inclusion of a third grid between the screen grid and the anode. This third grid is usually kept at cathode potential, so that the secondary electrons, which as we have mentioned possess little energy, are forced back to the anode. For this reason, the third grid is called the suppressor grid. It can have wide meshes without affecting its operation, and because of its low potential it carries hardly any current. A valve with three grids is called a pentode.

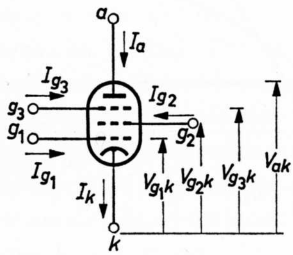

Fig. 17-2

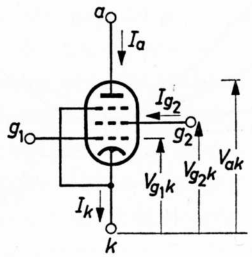

Fig. 17-3

With the pentode we have, in principle, to deal with four independent voltages and four independent currents (Fig. 17-2). However, it is usually possible to simplify matters considerably because I_{g1} and I_{g3} are so small as to be negligible, whilst the suppressor grid is usually connected to the cathode, so that $V_{g3k}=0$. The relation between the remaining voltages and currents (Fig. 17-3) is usually expressed as the two functions $I_a=I_a(V_{g1k}, V_{g2k}, V_{ak})$ and $I_{g2}=I_{g2}(V_{g1k}, V_{g2k}, V_{ak})$. As with the triode, these functions can be reduced to linear relations for small signal voltages. This yields the following equations:

$$\begin{cases} i_a = S_{1a}v_{g1k} + S_{2a}v_{g2k} + S_{aa}v_{ak} \\ i_{g2} = S_{12}v_{g1k} + S_{22}v_{g2k} + S_{a2}v_{ak} \\ i_k = i_a + i_{g2} \end{cases} \tag{17.1}$$

where all coefficients S_{ij} have the dimension of a conductance. Similarly to the triode, where the amplification factor is almost constant for the entire region of operation, the ratios $S_{1a}/S_{2a}=\mu_{g2g1}$ and $S_{1a}/S_{aa}=\mu$ show little tendency to change, so that once more the anode current can be written as a function of a "control voltage" v_s:

$$i_a = S_{1a}v_s = S_{1a}\left(v_{g1k} + \frac{v_{g2k}}{\mu_{g2g1}} + \frac{v_{ak}}{\mu}\right)$$

As a result of the screening caused by the three grids, the effect of the anode voltage on the cathode current is extremely small. The anode voltage only affects the distribution i_a to i_{g2} to a small degree. The internal resistance of the anode $(i_a/v_{ak})^{-1} = 1/S_{aa}$ will therefore be very great; for most pentodes its value lies between 0.5 and 3 MΩ. Since the slope S_{1a} is of the same order of magnitude as in the case of a triode (milliamps per volt), the amplification factor will also be very large (maximum about 10^4). Factor μ_{g2g1} of an ordinary pentode has a value between 20 and 40, and in the case of power valves between 5 and 20.

We can state of the i_{g2}-equation that $S_{12} = aS_{1a}$ and $S_{22} = aS_{2a}$ apply in very good approximation. In the region of normal operation, a is almost a constant determined by the valve's geometry; its usual value is between 0·1 and 0·3. Since the anode voltage has almost no effect on i_k but exclusively on the distribution i_a to i_{g2}, it follows that S_{a2} almost equals $-S_{aa}$ and is therefore also very small. In most applications it is thus permissible to approximate equations (17.1) by:

$$\left\{\begin{aligned} i_a &= S_{1a}\left(v_{g1k} + \frac{v_{g2k}}{\mu_{g2g1}} + \frac{v_{ak}}{\mu}\right) \\ i_{g2} &= a i_a \\ i_k &= (1 + a) i_a \end{aligned}\right. \tag{17.2}$$

Only when high calculation accuracy is essential are the approximations of (17.2) not allowed. Especially the ratio $a = i_{g2}/i_a$ cannot be assumed to be constant in this case. If a is known, I_a can be expressed as a function of V_{g1k} at various values of V_{g2k}, and as a function of V_{ak} at different values of V_{g1k}. This usually gives us data for a sufficiently accurate calculation of the behaviour of the pentode as a component for linear amplification. Fig. 17-4 shows these characteristics for valve EF86 (=6267).

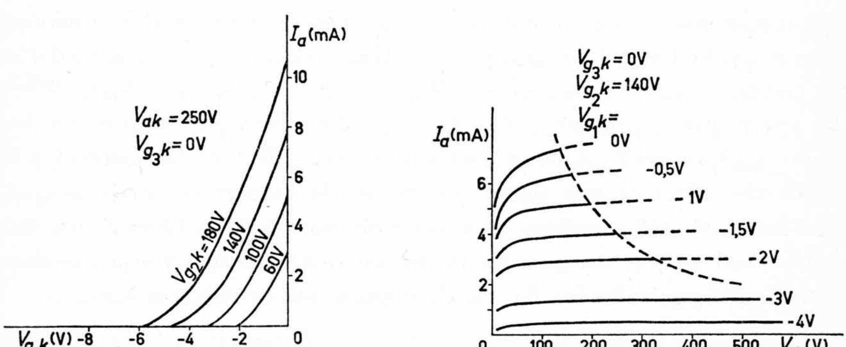

Fig. 17-4

As the cathode current, that is the sum of I_{g2} and I_a, is almost independent of V_{ak}, the right-hand side of the diagram shows that the distribution between I_a and I_{g2} undergoes drastic changes if the anode voltage becomes considerably lower than the screen grid voltage. In our example this would be for $V_{ak}<60$ volts at $V_{g2k}=140$ volts. In this case the above-mentioned approximations are no longer valid.

For the valve EF86 (= 6267) with the quiescent voltages

$$V_{g1k} = -1.5\text{ V},\ V_{g2k} = 120\text{ V and } V_{ak} = 200\text{ V}$$

the following values were found for the quiescent currents and the coefficients of (17.1):

$$I_a = 3\text{ mA};\ I_{g2} = 0.57\text{ mA};$$
$$S_{1a} = 1.9\text{ mA/V};\ S_{2a} = 47\ \mu\text{A/V};\ S_{aa} = 0.5\ \mu\text{A/V};$$
$$S_{12} = 0.38\text{ mA/V};\ S_{22} = 10.5\ \mu\text{A/V};\ S_{a2} = -0.2\ \mu\text{A/V};$$

This gives:

$$a \approx 0.2;\ \mu_{g2g1} \approx 40;$$
$$\mu \approx 3800;\ r_a = 1/S_{aa} \approx 2\text{ M}\Omega.$$

As the positive supply voltage is often too high to serve directly as screen grid voltage, it is usual to connect the screen grid to this supply voltage through a resistor, so that the quiescent current flowing to the screen grid provides the required voltage drop (Fig. 17-5). However, this would cause a signal voltage to appear on the screen grid, which would be opposite in phase to the one on the control grid and would therefore both reduce the amplification of the control grid to the anode, and increase the input capacitance of the valve because of the consequent Miller effect capacitance between g_1 and g_2.

For this reason, with a.c. amplifiers, the screen grid is decoupled by means of a capacitor (C_2) in a similar way as was done for the cathode. The impedance of the capacitor at the signal frequency should be so small that the signal voltage appearing at the screen grid has negligible effect on the screen grid current and anode current. Both these currents are then almost entirely determined by the control grid voltage. This makes it very simple to derive the condition that C_2 must satisfy. We have in this case $i_{g2}=aS_{1a}v_i$. When the screen grid impedance is $C_2//R_2=Z_{g2}$, the screen grid voltage will be: $v_{g2}=-i_{g2}Z_{g2}=-aS_{1a}Z_{g2}v_i$. Compared to the effect of the control grid signal, the effect of this signal on the anode current is: $aS_{1a}Z_{g2}/\mu_{g2g1}$. This factor should therefore be small with respect to 1. Once more, this requirement is most stringent for the lowest signal frequency $\omega_{\min}$, so that, assuming a large value for R_2, the decoupling will be effective when:

$$\frac{aS_{1a}}{\mu_{g2g1}}\,\frac{1}{\omega_{\min}C_2} \ll 1 \quad\text{or}\quad C_2 \gg \frac{aS_{1a}}{\mu_{g2g1}\omega_{\min}}$$

Example: Audio amplifier stage, $\omega_{\min} \approx 100$, using the valve EF 86: $I_a = 3$ mA; $a = 0.2$; $S_{1a} = 2$ mA/V; $\mu_{g2g1} = 40$ give for C_2;

$$C_2 \gg \frac{0.2 \cdot 2 \cdot 10^{-3}}{100 \cdot 40} \text{ F} \approx 0.1\ \mu\text{F}$$

If the screen grid is decoupled to the cathode instead of to earth (Fig. 17-6), and the value of the screen grid resistor is sufficiently large,

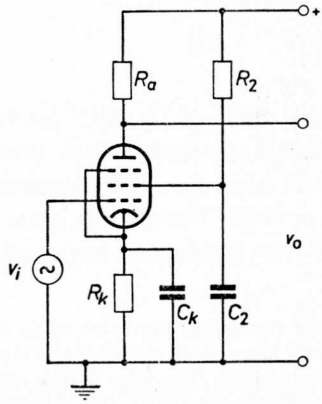

Fig. 17-5

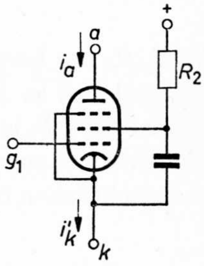

Fig. 17-6

the signal current i_k', which flows through cathode terminal k, will be equal to the anode signal current. The pentode then behaves as a triode with a very high amplification factor. Because, when $v_{g2k} = 0$, we have:

$$i_a = S_{1a}\left(v_{g1k} + \frac{v_{ak}}{\mu}\right) \tag{17.3}$$

and

$$i_k' = i_a - \frac{v_k}{R_2}$$

The property of a triode, that the anode and cathode currents are the same within very narrow limits, is used in a great number of measuring circuits. The above formulae show that the pentode is less suitable for this purpose.

As long as v_{ak} is much smaller than μv_{gk} (which is usually the case because μ is very large), equation (17.3) can be approximated to:

$$i_a = S_{1a} v_{g1k} \tag{17.4}$$

The amplification given by the circuit of Fig. 17-5 then yields:

$$A_a = \frac{v_a}{v_i} = -S_{1a}R_a$$

The d.c. voltage drop through anode resistor R_a is $V_+ - V_a = R_a I_a$, where I_a is the anode standing current.

After elimination of R_a from both equations:

$$A_a = -\frac{S_{1a}}{I_a}(V_+ - V_a)$$

The ratio S_{1a}/I_a can have a theoretical value of 10 volt^{-1} (at very small current densities), but in general practice one finds values smaller than 2 volt^{-1}, so that the gain in this case will be of the same magnitude as the d.c. voltage across the anode resistor in volts. This affords gains of a few hundred; the amplification factor μ (10^3–10^4) is therefore not reached at all.

S_{1a}/I_a can be increased by decreasing the value of I_a (Fig. 17-7). This is achieved by using a very large anode resistor if this is permissible in relation to the bandwidth required. The valve is then adjusted to a very small quiescent current (less than 0.1 mA). This is called a starved amplifier, which can yield a gain of more than 1000 from a single valve at a voltage of, say, 200 V across R_a. However, at normal adjustments the gain in amplification obtained with the pentode, in comparison with the triode, is much smaller than the ratio of the amplification factors would lead us to believe.

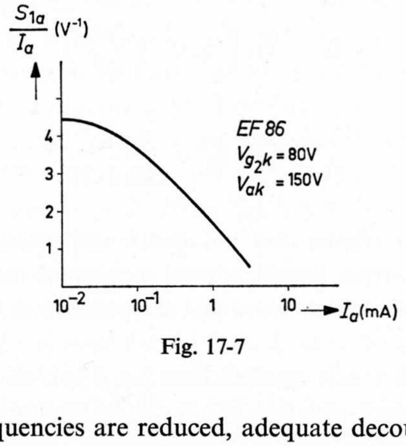

Fig. 17-7

As the signal frequencies are reduced, adequate decoupling of the screen grid can be achieved by using still larger capacitors. This can be avoided by using a cathode follower for decoupling in the manner shown in Fig. 17-8. Here we use the fact that the cathode of the triode follows the grid accurately and has a low internal resistance. This circuit has the additional advantage for many applications that the pentode cathode is hardly loaded by the voltage divider $R_1-(C//R_2)$, for which large resistors can be used.

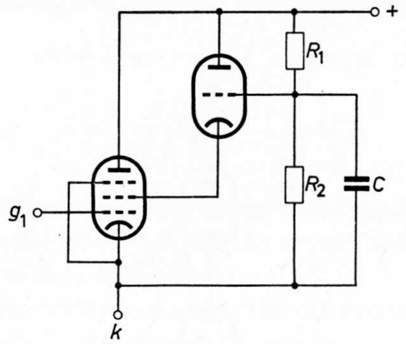

Fig. 17-8

We can thus conclude by saying that, although the pentode has some very real advantages over the triode, on the other hand the pentode often introduces a number of complications in measuring circuitry. The pentode should therefore primarily be considered as an amplifier valve for not too low frequencies.

18. Cascode

Especially in instrumental electronics, a valve combination is often used that combines a number of pentode advantages (high gain and small grid-anode capacitance) with the accuracy given by the triode. This combination consists of two triodes in series and is known as a cascode (Fig. 18-1). Both

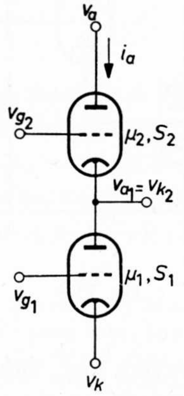

Fig. 18-1

triodes are adjusted so that their grid currents are negligible and therefore carry the same anode current i_a. By applying the triode equation to both valves, we obtain for this current:

$$\begin{cases} r_{a1}i_a = \mu_1 v_g + v_{a1} - (\mu_1 + 1)v_k \\ r_{a2}i_a = \mu_2 v_{g2} + v_a - (\mu_2 + 1)v_{a1} \end{cases}$$

where r_{a1} and r_{a2} are the anode resistances μ_1/S_1 and μ_2/S_2.

After elimination of v_{a1} from both equations, we find:

$$[(\mu_2 + 1)r_{a1} + r_{a2}]i_a = \mu_1(\mu_2 + 1)v_g + \mu_2 v_{g2} + v_a - (\mu_1 + 1)(\mu_2 + 1)\, v_k$$

One now asks if such a linear combination $av_g + bv_k + cv_a$ can be selected for v_{g2} so that this equation for the anode current becomes once more a "triode equation", that is that i_a is exclusively determined by the voltage differences v_{gk} and v_{ak}, so that the sum of the coefficients of v_g, v_a and v_k is zero. This proves to be the case for $a+b+c=1$. Of all possible combinations, only two are of practical interest: $a=1$, $b=c=0$; and especially $a=c=0$, $b=1$.

In the first case, $v_{g2}=v_g$, and the cascode then behaves as a triode with:

$$r_a = (\mu_2 + 1)r_{a1} + r_{a2} \;, \; \mu = \mu_1\mu_2 + \mu_1 + \mu_2$$

and
$$S = \frac{\mu_1\mu_2 + \mu_1 + \mu_2}{(\mu_2 + 1)r_{a1} + r_{a2}} \approx S_1$$

In the second case, $v_{g2}=v_k$, and the cascode behaves as a triode with

$$r_a = (\mu_2 + 1)r_{a1} + r_{a2} \;, \; \mu = \mu_1\mu_2 + \mu_1$$

and
$$S = \frac{\mu_1(\mu_2 + 1)}{(\mu_2 + 1)r_{a1} + r_{a2}} \approx S_1$$

The properties of the combinations do not differ much in either case. The difference is the control of the grid of the upper triode, which is usually effected more easily with v_k than with v_g.

Although the characteristic values of the cascode are therefore of the same magnitude as those of the pentode, the cascode shows a number of advantages.

Because the grid of the upper triode carries neither a direct current nor a signal current, decoupling is much easier than that of the screen grid of a pentode. Moreover, the anode current is exactly the same as the cathode current, which is important for some applications in measurement circuits as we have already stated. We shall see in Section 31 that the noise properties of the cascode are better than those of the pentode. Finally, the Miller effect

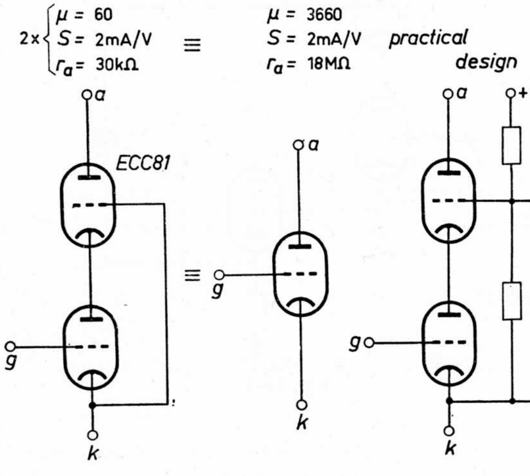

Fig. 18-2

is also rather small in the case of the cascode. One could even think that it had completely disappeared because the upper anode and the lower grid do not necessarily possess a mutual capacitance (two separate valves), but this is not correct. The lower anode also carries a signal, although approximately a factor μ_2 smaller than that on the upper anode. This signal still gives cause to a Miller effect because of C_{ag} of the lower valve.

A disadvantage of the cascode is that the series connection of the valves necessitates approximately twice the anode d.c. voltage of that used for a single pentode or triode.

In theory there is no reason for stopping at a cascode consisting of two triodes only. A three-element cascode with still more striking properties can be obtained by connecting a third valve in series. When the upper grids are controlled by the cathode, we obtain:

$$r_a = r_{a3} + (\mu_3 + 1)r_{a2} + (\mu_3 + 1)(\mu_2 + 1)r_{a1}$$

and

$$\mu = \mu_1(\mu_2 + 1)(\mu_3 + 1)$$

S has once more approximately the same value as S_1.

Fig. 18-2 gives an example of a cascode circuit.

If, in a practical design, the cathode must be completely free from the load imposed by the voltage divider, a cathode follower can be inserted, (Fig. 18-3). This circuit can, of course, also be used with the pentode.

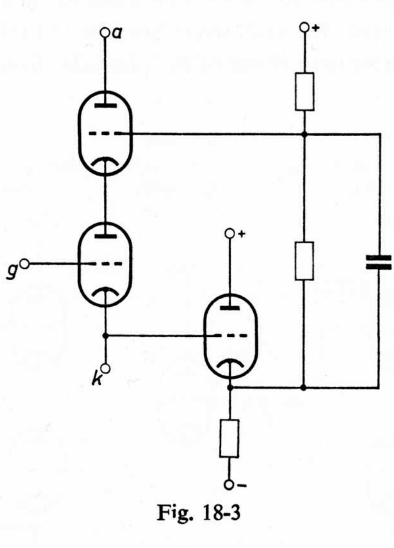

Fig. 18-3

19. Balanced amplifiers

Apart from the cascode, other combinations of valves can also offer attractive properties. One of these is the balanced amplifier, where two valves (triodes or pentodes) are connected, not in series, but in parallel (Fig. 19-1). These valves have a common cathode resistor connected to a voltage supply. The latter is negative (sometimes zero) with respect to the d.c. voltage level of the two control grids. As will appear from the analysis, this circuit possesses a certain degree of symmetry, even with differing valves. This symmetry is greater when the cathode resistor is larger. Therefore, a design with a very large cathode resistor is most frequently used, particularly for measurement circuits. This circuit is known as "long-tailed pair" because of the high value of the cathode resistor.

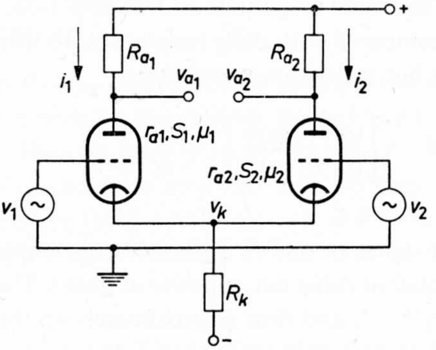

Fig. 19-1

We can visualize a balanced amplifier as originating from the single triode with a cathode resistor, but where the cathode is not decoupled by the low impedance of a capacitor, but by the low internal resistance of the cathode of a second triode.

This decoupling is not only effective for frequencies down to a certain $\omega_{\min}$, but to direct current as well.

The following equations apply to the circuit of Fig. 19-1:

$$\begin{cases} r_{a1}i_1 = \mu_1 v_1 + v_{a1} - (\mu_1 + 1)v_k \\ r_{a2}i_2 = \mu_2 v_2 + v_{a2} - (\mu_2 + 1)v_k \\ v_k \quad = R_k(i_1 + i_2), \quad v_{a1} = -R_{a1}i_1, \quad v_{a2} = -R_{a2}i_2 \end{cases}$$

By substituting the values of v_k, v_{a1} and v_{a2} from the last equations in the first two equations, we obtain after division by (μ_1+1) and (μ_2+1):

$$\begin{cases} (R_k + \varrho_1)i_1 + R_k i_2 = v_1' \\ R_k i_1 + (R_k + \varrho_2)i_2 = v_2' \end{cases} \tag{19.1}$$

where
$$v_1' = \frac{\mu_1}{\mu_1 + 1} v_1 \ , \ v_2' = \frac{\mu_2}{\mu_2 + 1} v_2,$$

$$\varrho_1 = \frac{r_{a1} + R_{a1}}{\mu_1 + 1} \ , \ \varrho_2 = \frac{r_{a2} + R_{a2}}{\mu_2 + 1}$$

We have from (19.1)

$$\begin{cases} [R_k(\varrho_1 + \varrho_2) + \varrho_1\varrho_2]\, i_1 = (R_k + \varrho_2)v_1' - R_k v_2' \\ [R_k(\varrho_1 + \varrho_2) + \varrho_1\varrho_2]\, i_2 = -R_k v_1' + (R_k + \varrho_1)v_2' \end{cases} \tag{19.2}$$

ϱ_1 and ϱ_2 are of the same magnitude as $1/S_1$ and $1/S_2$. It is therefore easy to make R_k large compared with these resistances. In this case we may write for (19.2) as a first but good approximation:

$$\begin{cases} (\varrho_1 + \varrho_2)i_1 = v_1' - v_2' \\ (\varrho_1 + \varrho_2)i_2 = -v_1' + v_2' \end{cases} \tag{19.3}$$

so that $i_1 = -i_2$.

By connecting a signal to one or both grids, we obtain currents in both valves which are equal in value but opposite in phase. They are proportional to the difference $v_1'-v_2'$, and thus approximately to the difference v_1-v_2, because $\mu/(\mu+1)\approx 1$. The proportionality factor $(\varrho_1+\varrho_2)^{-1}$ is for anode resistors of the order of the internal resistance, approximately equal to $\frac{1}{4}S$ and increases to $\frac{1}{2}S$ in the case of smaller anode resistors.

The calculation shows that the degree of symmetry and the extent to which currents i_1 and i_2 are proportional to the difference v_1-v_2, are determined by the ratio ϱ/R_k and the difference between $\mu_1/(\mu_1+1)$ and $\mu_2/(\mu_2+1)$. It is possible to render the factor ϱ/R_k small by giving a large value to R_k. Hence the "long-tailed pair" circuit. The consequent degree of symmetry is adequate for many applications. In Section 28 we shall discuss in detail the measures which can be taken, when necessary, to improve the symmetry so that the valve currents and therefore the output voltages v_{a1}, v_{a2} or $v_{a1}-v_{a2}$ will be very accurately proportional to the difference v_1-v_2. A possible improvement is indicated in the following example.

A "long-tailed pair" circuit using the valve ECC81 (=12AT7) with a

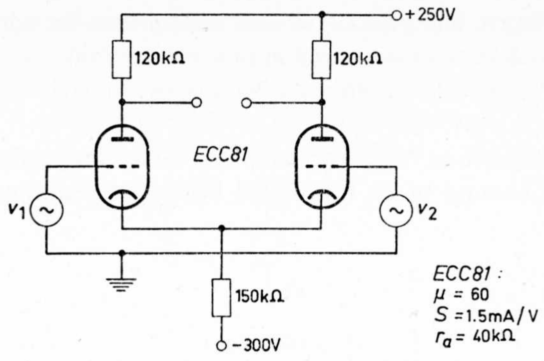

Fig. 19-2

cathode resistor of 150 kΩ and anode resistors of 120 kΩ (Fig. 19-2) gives the following values for the quantities ϱ_1 and ϱ_2 in (19.2):

$$\frac{r_a + R_a}{\mu + 1} = \frac{(40 + 120) \cdot 10^3}{61} \approx 2.6 \text{ k}\Omega$$

so that $R_k + \varrho_1$ and $R_k + \varrho_2$ differ from R_k by approximately 1.7 per cent.

If a valve circuit is used for the cathode resistor (see p. 34), we obtain the circuit of Fig. 19-3. Here we have: $R_k \approx 60 \cdot 75$ k$\Omega = 4.5$ MΩ, and the difference between $(R_k + \varrho)$ and R_k is now only 0.06 per cent.

The difference between the factors $\mu_1/(\mu_1 + 1)$ and $\mu_2/(\mu_2 + 1)$ is determined by the values of μ_1 and μ_2. Presuming the most unfavourable case, that μ_1

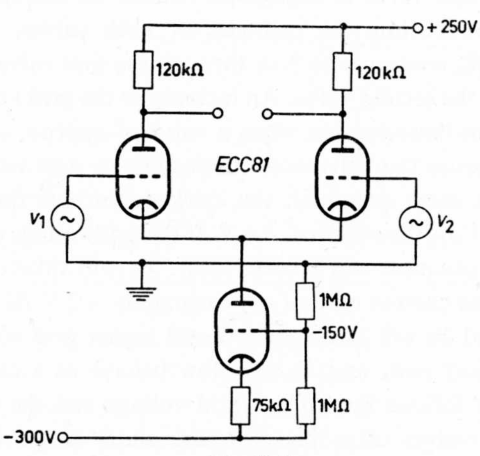

Fig. 19-3

is 10 per cent larger, and μ_2 is 10 per cent smaller than the nominal value 60 (the maximum deviations occurring in practice) we find: $\mu_1/(\mu_1+1)=66/67$ and $\mu_2/(\mu_2+1)=54/55$, the difference being approximately 1/300.

Apart from its use as "difference amplifier", the "long-tailed pair" is also often applied because of its favourable behaviour regarding large signal voltages.

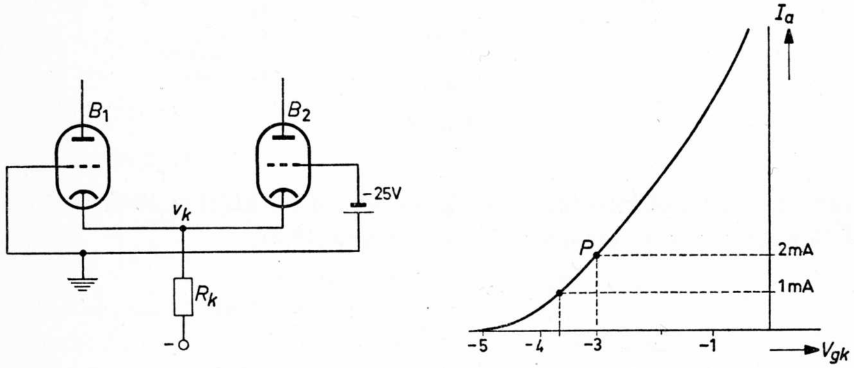

Fig. 19-4

We assume in Fig. 19-4 that valve B_1, having a large cathode resistance R_k, is adjusted to point P of its characteristic. The valve current is 2 mA and the cathode voltage +3 V. If we now adjust the cathode of the corresponding valve B_2 also to +3 V and the grid to −25 V ($V_{gk}=-28$ V), the current flowing through this valve is negligible, so that no change occurs in the situation when connecting the cathodes of both valves. All the current flowing through R_k continues to flow through the first valve and no current will pass through the second valve. An increase in the grid voltage of B_2 only causes a change in this situation when a value of approx. −2 V is reached ($V_{gk}=-5$ V), because this valve now commences to pass current. When the grid has attained earth potential, the currents through the two valves are about equal, and V_k is then approx. 3.6 V. If the grid voltage of B_2 is increased still further, the situation will reverse itself: B_2 will take over an ever increasing part of the current until at approximately +2 V B_2 will pass almost all the current and B_1 will be cut off. At still higher grid voltages of B_2, B_1 no longer takes any part, and B_2 will now behave as a cathode follower. The cathode thus follows the highest grid voltage and the current transfer between the two valves takes place in the short range from $V_1-V_2=-2$ V to $V_1-V_2=+2$ V (Fig. 19-5). The width of this range may be 1−10

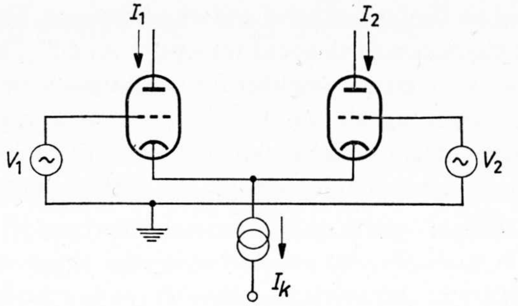

Fig. 19-5a

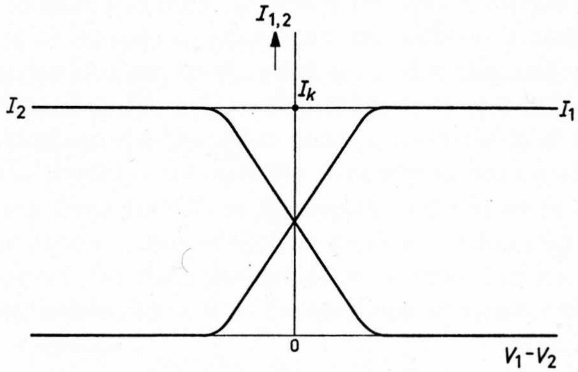

Fig. 19-5b

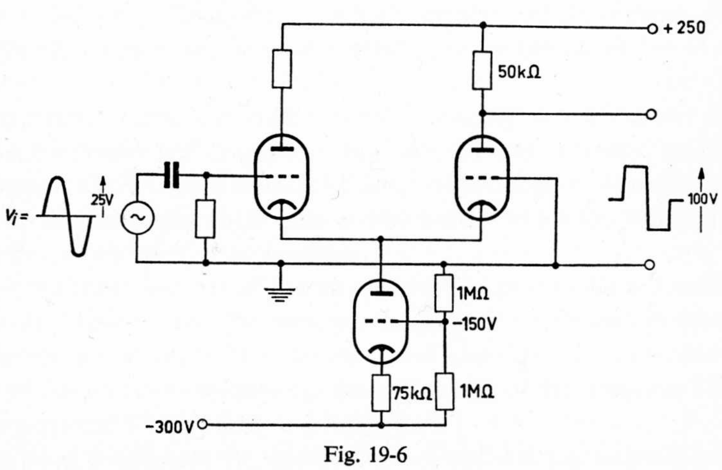

Fig. 19-6

volts, depending on the type of valve and its adjustment. The steepest slope in this current characteristic is about $(\varrho_1+\varrho_2)^{-1} \approx 2/S$. The circuit therefore operates as a difference amplifier for small signals, but it will operate as a limiter for larger signals. In the ideal case of a perfectly constant cathode current we obtain curves I_1 and I_2 of the diagram. Fig. 19-6 shows a possible design of such a limiting circuit. We shall discuss its applications in Section 37.

20. Semiconductor diode and transistor

Apart from the valve types discussed earlier, transistors also play an important part as active components. It is commonly thought that the transistor is basically different from the valve, but this is not true, at least with respect to its external behaviour. As an active amplifier element, there is hardly any principle difference between the behaviour of the transistor and that of the valve, and any differences that do exist, are mainly quantitative. On the other hand, the physics of the transistor are entirely different. The physics of a valve are so simple, apart from some refinements, that it is not difficult to explain its physical action, but the physics of the transistor are much more complex and certainly need much more explanation.

Although it is possible to present the transistor as a component with certain properties that can be measured at its terminals, this would appear rather unsatisfactory, especially in a book which devotes attention to accuracy and limitations. On the other hand, a comprehensive discussion of semiconductor physics is not desirable. This section will therefore compromise by giving those facts from the wealth of information in this field, which help to form a picture of how the idealized semiconductor diode and transistor operate. We shall also devote some space to various transistor manufacturing techniques, and to the operation of a rather different type, the field-effect transistor.

In the next section we shall discuss the same points in relation to transistors as we have done with valves: operation as linear amplifying elements, simple fundamental circuits, properties at high frequencies, and choosing the working point. Because of the high temperature dependence, this last subject will be discussed in greater detail than for the valves.

The current carriers do not move *in vacuo* (or rarefied gases) in semiconductor diodes and transistors, but in solid state crystals, usually germanium or silicon. These materials belong to the category of semiconductors, a group of substances which have a conductivity which lies between that of a conductor and that of an insulator.

Only valence electrons of the atoms can take part in current conduction in solids. The difference between good and poor conductors is determined by the energy necessary to make the atoms lose these electrons, and by the ratio of this energy to the average thermal energy of the electrons. The latter is proportional to the absolute temperature.

For good conductors the situation is that with a single, isolated, atom, the valence electrons are tied rather strongly to this atom, necessitating considerable energy to free the electron from the atom. If this has succeeded, there will be a "hole" left in the atom which has a great affinity to other electrons. On the other hand, the mutual exchange phenomenon between the atoms in the crystal lattice makes the adherence of the valence electrons to the atom very weak, so that hardly any energy is necessary to liberate such an electron. This can be visualized as follows: no real hole is formed in the atom by the absence of an electron, but the absent negative charge has been spread out. The energy of thermal motion is sufficient at room temperature – and even at much lower ones – to liberate all valence electrons; these can move quite freely through the solid. These "free electrons" have a random thermal motion, as well as a field movement directed by an electric field, if present.

The velocity of this directed movement is determined by the "mobility" μ_n of the electrons in the material, which is defined as the mean velocity in the field direction per unit electric field strength. The conductivity of the material is therefore determined by the product of the number of free electrons per unit of volume and their mobility.

Good conductors are characterized by the fact that their conductivity decreases at higher temperatures. When the temperature is increased the mean velocity of the free electrons also increases and hence their chance of colliding with the atomic crystal lattice. This diminishes the mean velocity in the direction of the electric field.

In the case of the semiconductors germanium and silicon, the four valence electrons are not only strongly bound in the individual atom but also in a crystal lattice. The bond between two adjacent atoms in the cubic lattice of these elements is ensured by the interchange of valence electrons. An energy of about 1 eV is required to liberate such an electron, and their average energy at room temperature is approx. 25 meV. As a consequence there are relatively speaking very few free electrons and the conductivity of these elements is very low, although the mobility of the free electrons in these substances is of the same magnitude as with good conductors. As the number of valence electrons which have sufficient energy to become liberated, increases roughly exponentially with the temperature, the conductivity at high temperatures will increase rapidly in this case.

The specific resistance of Ge and Si at room temperature is approx. 50 Ω cm and 200 kΩ cm respectively, compared with 1.7 $\mu\Omega$ cm for copper. The mobility of electrons in these three elements, in the same order, has the values 4000, 1500 and 30 cm^2/V s. At 100 °C, the specific resistance of germanium has already reduced to 5 Ω cm.

We have mentioned that a "spread out" positive charge remains when an electron has been liberated in a conductor. Not so with semiconductors. Here a concentrated positive charge remains in the lattice in the form of a "hole". These holes have such a great attraction to electrons, that they are "filled" quite easily. However, because of the paucity of free electrons, the chance that a hole will be filled by a free electron is small. This task is preferentially carried out by valence electrons of adjacent atoms, thus creating new holes in those atoms. The result is an apparent displacement of the holes through the material before capturing a free electron, with which they recombine, thus causing the disappearance of both a hole and a free electron. If an electric field is applied, the movement of the holes will become "directional" and we can therefore also speak of the mobility of the holes. The displacement of the holes is so rapid that their mobility μ_p is almost the same as that of the free electrons. The ratio $\mu_p : \mu_n$ has an approximate value of 0.5 for Ge and 0.3 for Si.

The charge transport in semiconductors occurs therefore by two distinct processes, which may be described as a movement of electrons and a movement of holes. In the state of equilibrium, the number of free electrons and holes created by the thermal motion (M/cm^3 s) must equal the number which disappear by recombination. The latter is proportional to the number of free electrons present (n/cm^3) and the number of holes present (p/cm^3), so that we have:

$$M = \lambda\, np \quad \text{or} \quad np = \frac{M}{\lambda} \tag{20.1}$$

where λ is a constant.
With the pure material, a hole will always occur for each free electron, so that here $n = p = \sqrt{M/\lambda}$. At room temperature the constant $\sqrt{M/\lambda}$ is approx. 10^{13}/cm^3 for germanium and approx. 10^{10}/cm^3 for silicon.

An essential point is that it is possible to make the number of holes unequal to the number of free electrons by the introduction of suitable impurities. For example, if a homogeneous impurity of an element with five valence electrons is introduced into the material, its atoms will be easily assimilated in the lattice, and four of the five valence electrons will ensure the bond with the neighbouring atoms. The fifth electron is essentially superfluous and has therefore a weak adherence to its nucleus. It can move as freely through the material as the valence electrons of true conductors and it leaves behind not a hole but a "spread out" charge. Such a "donor" atom therefore produces a free electron but no hole. If the impurity is N atoms per cm^3, and if N is much greater than $\sqrt{M/\lambda}$, we have

$$n \approx N$$

and with $np = M/\lambda$:

$$p \approx M/\lambda N$$

The number of holes can be made greater than the number of electrons by using "acceptor"-elements with three valence electrons as impurity.

According to whether n is larger or smaller than p, we speak of n- or p-material. If $n = p$, the material is called intrinsic. This is not necessarily a pure material, for even when the number of atoms with five valence electrons equals the number of atoms with three valence electrons, we have $n = p$. In other cases it is also possible that both types of impurity are present but to a different degree. The materials used in transistors have n/p and p/n ratios in the order of 10^4–10^6, which, as can be easily established, amounts to an impurity of approx. 1 in 10^8!

We shall now see how diode action can be obtained by making use of these n- and p-materials. Fig. 20-1 shows a piece of n-material and a piece of p-material, as well as the schematic representation of the concentrations of holes and free electrons. It should be borne in mind that it is not possible to give a correct representation of the actual concentration ratios normally encountered in transistor material.

If one imagines the two pieces combined to a single bar, the situation shown in Fig. 20-2 is created. A sharp junction from p- to n-material occurs, which is called a p-n junction. Such a junction is achieved in reality by introducing impurities in one crystal, which creates n- and p-regions, which will merge into each other. The situation sketched in Fig. 20-2 cannot be maintained. Because of the thermal motion, more electrons will diffuse at the junction plane from left to right than from right to left and, vice versa, more holes from right to left than from left to right. This would continue until the concentration of electrons and also the concentration of holes were equal in both pieces, were it not for the fact that an electric field is formed at the junction plane because of the displacement of charge. This field will very soon prevent any further diffusion. A positive charge will be formed on the left-hand side of the junction and a negative charge on the right-hand side. The regions in which these charges occur prove to be extremely small.

If the dimensions of the bar are greater than the width of this junction, the potential difference between the two halves can be easily calculated. In the n-material outside the junction, the normal equilibrium concentrations which we shall call n_n and p_n will prevail; in the same way the equilibrium concentrations n_p and p_p will still prevail in the remainder of the p-material.

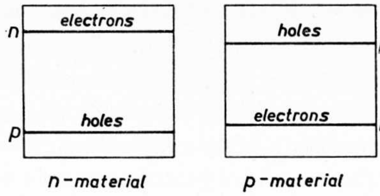

Fig. 20-1

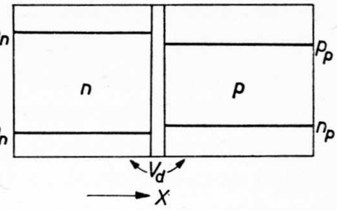

Fig. 20-2

We now have for the junction region, in the final situation, the same number of electrons moved by diffusion to the right, as are forced by the electric field to the left. The number of electrons diffused to the right-hand side per unit of area is given by $-D_n(\mathrm{d}n/\mathrm{d}x)$, where D_n = diffusion constant for the electrons. The corresponding number of electrons forced to the left by the electric field strength E is $+\mu_n E n$, so that

$$-D_n \frac{\mathrm{d}n}{\mathrm{d}x} = \mu_n E n$$

or

$$E\mathrm{d}x = -\frac{D_n}{\mu_n} \cdot \frac{\mathrm{d}n}{n}$$

Integration over the layer thickness gives:

$$V_d = V_n - V_p = \frac{D_n}{\mu_n} \ln \frac{n_n}{n_p}$$

Einstein has derived the following relation between μ and D (Einstein's Diffusion Law):

$$\frac{\mu}{D} = \frac{q}{kT}$$

where q = charge of the charge carrier, i.e. the electron, k = Boltzmann's constant and T = absolute temperature of the medium. The value of q/kT for $T = 300$ °K is about 40 volt^{-1}.
Using Einstein's relation, the above equation can be written as:

$$V_d = \frac{kT}{q} \ln \frac{n_n}{n_p} = \frac{kT}{q} \ln \frac{p_p}{p_n} \tag{20.2}$$

as the same derivation is valid for the holes, and moreover (20.1) gives $n_n p_n = n_p p_p$.

With the usual impurities the value of V_d amounts to a few hundred millivolts. It can be calculated from this value how large the charge displacement in the junction area must be to cause this potential difference. It then appears that the junction width is very small, of the order of 1 micron.

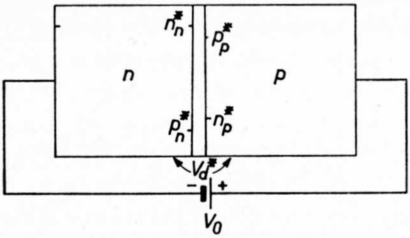

Fig. 20-3

Voltage V_d is not available between the external terminals. As soon as an external connection is established between the terminals of the bar, there occur also potential differences across the junctions of the bar to the terminals and the sum of all potential differences is zero. However, when a voltage supply source V_0 is inserted in the external circuit, the equilibrium will be upset and a current will pass through the circuit (Fig. 20-3). To calculate this current we once again assume that the charge displacements are restricted to an extremely thin layer. In this junction region, the electron current density I_n (=charge displacement by the electrons per unit area per second) must everywhere be equal to the difference between the electron diffusion current density and the electron field current density:

$$I_n = q\left(-D_n \frac{\mathrm{d}n}{\mathrm{d}x} - \mu_n\, n\, E\right)$$

It is difficult to manipulate this equation with exactness but we can apply a justifiable approximation. In view of the thinness of the junction region, $\mathrm{d}n/\mathrm{d}x$ and E are relatively large, so that current densities $qD_n(\mathrm{d}n/\mathrm{d}x)$ and $q\ \mu_n\ nE$ are usually very large compared with the current densities I_n. The previously used equation

$$-D_n \frac{\mathrm{d}n}{\mathrm{d}x} = \mu_n\, n\, E$$

will therefore still apply to a very good approximation. Integration then yields:

$$V_d{}^* = \frac{kT}{q} \ln \frac{n_n{}^*}{n_p{}^*}$$

where $V_d{}^*$ is the potential difference across the junction and $n_n{}^*$ and $n_p{}^*$ are the electron concentrations at the boundaries of the junction region with the n- and p-material.

Since the potential differences at the external connections will not change when the contacts are good, it follows that, neglecting the voltage drop due to the ohmic resistance of the material, this new voltage must differ by an amount $-V_0$ from the original voltage V_d, with the polarity of V_0 indicated in the illustration:

$$V_d{}^* = V_d - V_0 = -V_0 + \frac{kT}{q} \ln \frac{n_n}{n_p}$$

so that
$$V_0 = -\frac{kT}{q} \ln \frac{n_n{}^*}{n_n} \cdot \frac{n_p}{n_p{}^*}$$

For the holes, we can state analogously:

$$V_0 = -\frac{kT}{q} \ln \frac{p_p{}^*}{p_p} \cdot \frac{p_n}{p_n{}^*}$$

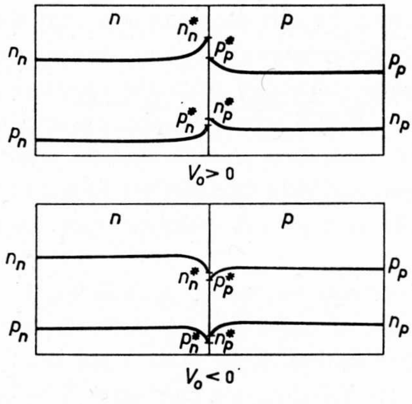

Fig. 20-4

As we have assumed that a charge displacement only occurs in the junction, and therefore outside this region the material to be electrically neutral, we have (Fig. 20-4):

$$n_n{}^* - n_n = p_n{}^* - p_n$$

and

$$p_p{}^* - p_p = n_p{}^* - n_p$$

Dividing the first equation by n_n yields:

$$\frac{n_n^*}{n_n} = 1 + \frac{p_n^* - p_n}{n_n}$$

We have seen that p_n/n_n is small (approx. 10^{-5}). We can therefore neglect $(p_n^* - p_n)/n_n$ with respect to 1 for small disturbances. Putting $n_n^*/n_n \approx 1$ and $p_p^*/p_p \approx 1$, we find:

$$V_0 = -\frac{kT}{q} \ln \frac{n_p}{n_p^*} \quad \text{or} \quad n_p^* = n_p \, e^{\frac{qV_0}{kT}}$$

and

$$V_0 = -\frac{kT}{q} \ln \frac{p_n}{p_n^*} \quad \text{or} \quad p_n^* = p_n \, e^{\frac{qV_0}{kT}}$$

From further calculations then follows that the concentrations show an exponential dependence on the distance to the junction.

For $V_0 < 0$, we obtain the situation shown in the upper part of Fig.20-4. In the case of the "majority charge carriers", i.e. the electrons in the n-region and the holes in the p-region, the relative disturbance is thus much smaller than with the "minority charge carriers".

The advantage gained by this calculation is that the junction with its strong potential and concentration difference no longer needs to be considered. The current can now be calculated from the situation in the "quiescent" regions away from the junction, where the electrons and holes are transported by a much smaller concentration gradient and the weak field due to ohmic voltage. Both diffusion and field currents are found in the neighbourhood of the junction, while at a greater distance from the junction only field currents will occur.

To calculate the current, we reason as follows: In the n-material the concentration drops are of equal size for both electrons and holes in the neighbourhood of the junction. Since there is not much difference between their diffusion constants, the diffusion currents will be of the same order of magnitude. Because there are more electrons than holes (at least for small departures from the equilibrium position) the electron field current at the junction will still be important compared with the electron diffusion current, while the much smaller hole field current does not play any part compared with the hole diffusion current. The latter represents here about the entire hole current and can easily be calculated from the known difference in concentration at the junction and at a greater distance from the junction. Similarly, the electron current in the p-region can be identified with the electron diffusion current, and can be calculated from the known difference in electron concentration. Assuming further that the creation and recombination of electron hole pairs can be neglected in the very narrow junction, it follows that the hole current at the boundary of the p-region is equal to the current at the boundary of the n-region. As this also applies to the electron currents, the total current I equals the sum of the diffusion currents of the "minority charge carriers" at the boundaries of the junction.

We have attempted in Fig. 20-5 to explain the composition of the current in the various regions. The length of the arrows is a measure of the magnitude. This schematic representation corresponds to the situation shown in Fig. 20-4 (upper part), where the concentrations are larger at the junction than in the rest of the materials.

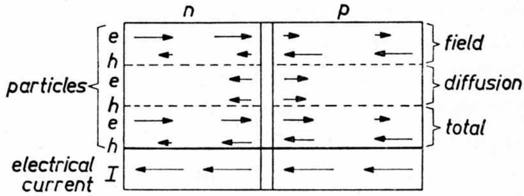

Fig. 20-5

The diffusion current of electrons in the p-region is proportional to the concentration gradient $n_p^* - n_p = n_p(e^{qV_0/kT} - 1)$. Similarly, the hole diffusion current in the n-region will be proportional to $p_n(e^{qV_0/kT} - 1)$, so that we find for the current in a semiconductor diode:

$$I = I_0 \left(e^{qV/kT} - 1\right) \tag{20.3}$$

where V is the voltage across the diode.

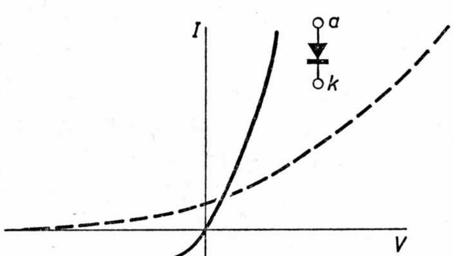

Fig. 20-6

By introducing different impurity concentrations into the materials ($n_p \neq p_n$) we can arrange that one type of charge carrier prevails at the junction and one could then, for instance, speak of the "injection" of holes from the *p*-region into the *n*-region.

The relationship between I and V (continuous line in Fig. 20-6) shows a great similarity to that of a valve diode (broken line). The vacuum diode carries a positive current even at negative anode-cathode voltages, which approximates to zero at large negative voltages. In the case of the semiconductor diode, however, the current reverses its sign at zero voltage, and approximates to a fixed and relatively small value I_0 for increasing negative voltages; this value I_0 is almost reached at approx. -1 volt. With positive voltages, the current increase with the semiconductor diode is approximately a factor e per 25 mV. In the case of the vacuum diode, the relation at small current densities has the form $I = I_0\ e^{qV/kT}$, where T = cathode temperature and where approximately 100 mV is required for such an increase in current.

As for the vacuum diode, we distinguish between a "forward" and a "reverse" direction. It follows from the above that the terminal of the *p*-material must be positive compared to that of the *n*-material with regard to the forward direction.

It is now relatively easy to explain the operation of the transistor. Let us begin by assuming that the diode receives such a voltage that it is reversed (Fig. 20-7).

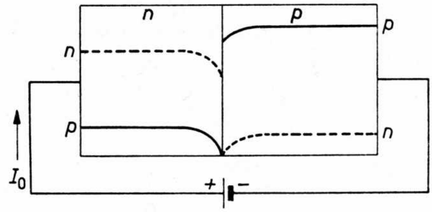

Fig. 20-7

The concentrations of electrons at the boundary of the *p*-region and of holes at the boundarys of the *n*-region are both practically zero, while the small current I_0 passes through the diode (Fig. 20-6).

If it were possible to increase the concentration drops in the *n*-region, the hole current would increase, while the electron current, determined by the concentration drop in the *p*-region, would remain unchanged. The result would be that the total current through the junction would also increase.

Such a change in the concentration drop can be effected by arranging a second conductive *p-n* junction at such a short distance from the first junction that the concentration drops of the two junctions will merge. The

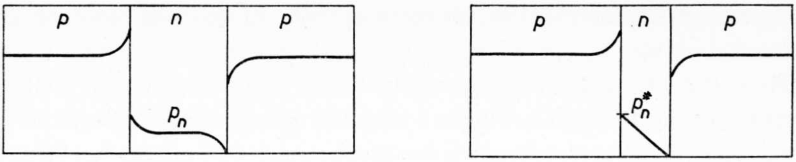

Fig. 20-8

left-hand side of Fig. 20-8 shows the situation when the distance between the two junctions is large compared with the diffusion area. It is obvious that the concentration drop of the holes at the conducting junction will not affect the concentration gradient at the reversed junction. However, this will be the case when the distance is sufficiently small (Fig. 20-8, right-hand side). The hole concentration at the boundary of the conducting junction is dependent on voltage V across this junction ($p_n^* = p_n e^{qV/kT}$). When the distance between the two junctions is small enough, a concentration gradient will occur which is proportional to $e^{qV/kT}$ and a proportional hole current will pass through the reversed junction. It is thus possible to control the current through one junction by varying the voltage across a second junction, similar to what happens in a valve when the anode current is controlled by the grid-cathode voltage. A transistor is a combination of two junctions which makes this control possible (Fig. 20-9).

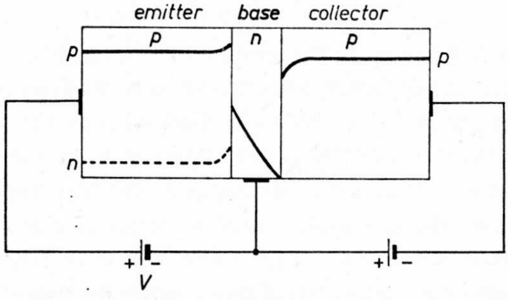

Fig. 20-9

In order to maintain the concentration gradient in the thin layer of n-material (called the *base* of the transistor), we need a continuous supply of holes on the side of high concentration. It is as though the holes were emitted by the adjoining p-material. The latter is therefore called the *emitter*. The holes are transported at the other side of the base to the other p-material, which is called the *collector*.

We have mentioned that the electron current through the base-collector junction does not change when the base-emitter voltage is altered. This is different for the base-emitter junction which is conductive, so that the electron current through it will vary with the voltage across it (broken line in Fig. 20-9). This current flows via the base contact. One attempts, of course, to keep it small compared to the hole current by an appropriate choice of impurities; nevertheless the base current is still a significant fraction of the total current: 0.1–10 per cent are possible values.

The ratio of the changes in the hole current and the total current through the base-emitter junction is called the emitter efficiency γ. Moreover, a number of holes will recombine after entering the base; only a fraction β reaches the collector. For this too it is useful to make the base thin; the base efficiency then

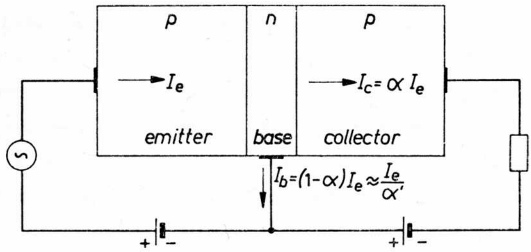

Fig. 20-10

approximates to 1. Changes in the emitter current I_e, Fig. 20-10, occur for a fraction $\alpha = \beta\gamma$ in the collector, while the remaining current $(1-\alpha)I_e$ passes through the base contact. Originally, α was the symbol for the "current amplification factor" of the transistor. However, it proved to be more important to use the ratio of the collector current change to the base current change $\alpha' = \alpha/(1-\alpha)$, which has also the advantage of a greater ease of manipulation as a figure. Nowadays we nearly always mean by the current amplification factor α'. Any possible misunderstanding disappears when the quantity is mentioned: α lies between approximately 0.9 and 1; α' between 10 and 1000.

Apart from the transistor in the configuration of Fig. 20-8 (*p-n-p* transistors) it is also possible to manufacture transistors having *p*-material for base and *n*-material for emitter and collector (*n-p-n* transistors). Their behaviour is similar, except that all d.c. voltages and currents change their sign, while the charge transport mechanism is mainly achieved by free electrons instead of holes.

It may be useful to note that most transistors are not as symmetrical in

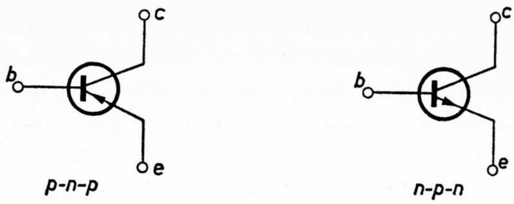

Fig. 20-11

design as would appear from Fig. 20-8. Not only do the dimensions vary but also the degree of impurity of emitter and collector. Although again a transistor will be obtained when changing over emitter and collector, it will have poorer properties for most applications.

The usual symbols for transistors are shown in Fig. 20-11. The arrow in the emitter indicates the direction of the d.c. current in the emitter. Under normal conditions as an amplifying element, the collector is negative with respect to the emitter in the case of a *p-n-p* transistor, and positive with an *n-p-n* transistor. With both types, the base voltage lies between the collector and emitter voltages, which is different to the comparable grid voltage of a valve.

The elements used for transistors are germanium and silicon. The latter's great advantage is that current I_0 (see equation (20.3)) which determines the leakage current between base and collector, has a much smaller value. Typical values for the leakage current at room temperature are 10^{-4}–10^{-6} A for germanium and 10^{-8}–10^{-10} A for silicon. In both cases the leakage currents have a certain dependence on the temperature (Si: 15 per cent and Ge: 10 per cent per °C). This means that in the case of germanium the leakage current will attain the same order of magnitude as the base current for moderate increases in temperature, which may completely spoil the operation. With silicon transistors this will only occur for much larger temperature increases.

Voltages and currents are of the same order of magnitude for both materials. They differ considerably from those met in valves. In most cases the base-emitter voltage is less than 1 volt, and the collector-emitter voltage does not exceed a few tens of volts. The admissible currents vary from a few milliamps with the smallest types to a few amps with power transistors.

The most important constructional difference found in the various manufacturing processes is the one between alloying and diffusion techniques. Alloying was the only possible technique to be used at the advent of transistors. It consists of forming alloys with the desired impurities for the emitter and the collector on both sides of the "base"-material. The effective base thickness is determined here by the penetration of the alloys. The variations in the base thickness are quite considerable and result in large differences between the parameters of various transistors of the same type. Another effect of this great variation is that one has to be content with a relatively thick base slice. This gives the transistor poor high-frequency properties, because any change in the base-emitter voltage necessitates, for the corresponding change in current, a certain change in the concentration of the minority charge carriers in the base. When the base is thick, there is a greater change in the number of charge carriers, corresponding to a greater apparent capacitance between base and emitter. We shall discuss the high-frequency properties of transistors in the next section.

Another effect inherent in the alloy transistor is the "Early effect". In order to obtain a transistor with good amplification properties, the collector material must contain considerably more free charge carriers per unit volume than the base material. When the junction between base and collector is reversed the space charge area is mainly in the base. If the collector voltage is now increased, this junction region will become wider and hence the effective base width will become smaller. The concentration gradient in the base will now increase in the first place, causing a corresponding increase in the emitter current. Furthermore, the base efficiency and therefore the current amplification factor α' will also increase slightly because of the narrowing of the width of the base. Thus, where in an ideal transistor, as long as the base-collector junction is reversed, the collector voltage has practically no effect on the passage of current, is in an alloy transistor the effect of the collector voltage still considerable.

In diffusion techniques one diffuses a thin layer into the collector material. This thin layer forms the base and the method allows one to determine the base thickness more directly than in the alloy process. When we also consider that the diffusion process takes place far more gradually than that of alloying, it will be obvious that a much greater precision in the base thickness can be obtained. This makes it possible to manufacture transistors with very thin base width and hence much better high-frequency characteristics.

The collector of a diffusion transistor usually contains less impurities than the base. Here the advantage is that when the collector voltage is raised, the space charge region at the base-collector junction will spread in the direction of the collector, so that the effective base thickness will not alter to such an extent and the Early effect will be much smaller. Another favourable result of the diffusion process is that the charge transport in the base region occurs largely by an electric field, so that the diffusion current becomes a smaller fraction of the total current. In accordance with what we noted regarding alloy transistors, this means that the effective capacitance between emitter and base becomes even smaller than would directly result from the effect of the thinner base.

A disadvantage of the lower conductivity of the collector material is that a relatively large resistance is present between the collector terminal and the collector-base junction. This can be troublesome especially in high-frequency applications. It can, however, be avoided by making use of the "epitaxial" technique, where a thin layer is grown on the crystal. This layer has properties which permit the base to be diffused into it. The crystal itself can then be more strongly contaminated and therefore be made more conductive. The emitter can be positioned on or in the base in a similar way.

The denominations "planar" and "mesa" transistors refer to certain variations in the techniques described above which give better electrical characteristics for some purposes.

Another component with amplification properties, the "field effect transistor" has been improved so much during recent years that it can now be used with profit in electronics.

The operation of the field effect transistor is based on the capability of affecting the conductance of semiconductor materials by varying the space charge area present with a reversed junction (see Fig. 20-12). Let us first consider the case when there is no potential difference between *S* and *D*. If we now apply voltages across the two *p-n* junctions in the reverse direction, a corresponding space charge is created because free charge carriers are no longer present in the junction area. The width of the area will be greater when the amount of impurity is small, and will increase with the reverse

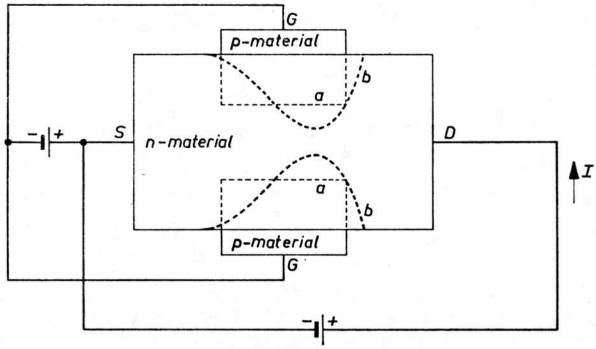

Fig. 20-12

voltage for a given amount of impurity (*a* in the diagram). If a potential difference is applied between *D* and *S*, a current will pass through the material and an electric field will be formed in the direction of the junction. This breaks the symmetry in the distribution of the space charge over the junction area, and the resulting situation is indicated by *b*. It now appears that above a given value of V_{DS}, the "pinch-off" voltage, conductance takes place only through a narrow channel, and there is no significant increase in current when V_{DS} increases. However, the value of this "saturation current" decreases approximately proportional to the square of the reverse voltage, and therefore characteristics as shown in Fig. 20-12 apply to field effect transistors. The symbols used are *S* for source, *D* for drain and *G* for gate.

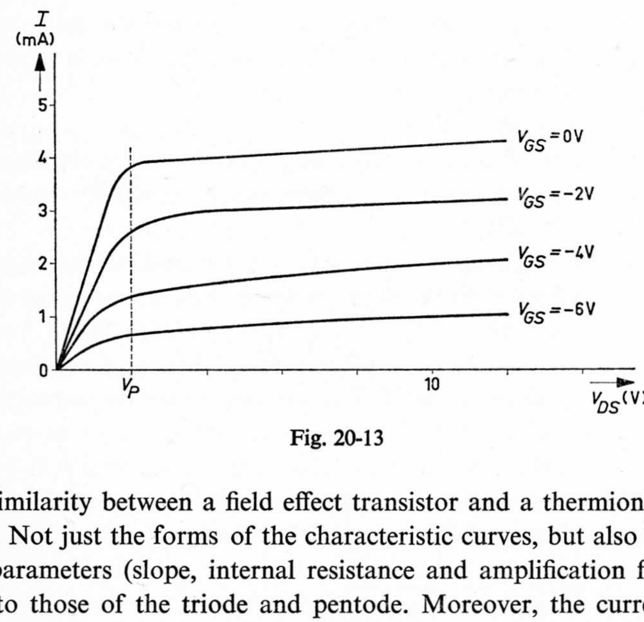

Fig. 20-13

The similarity between a field effect transistor and a thermionic valve is striking. Not just the forms of the characteristic curves, but also the values of the parameters (slope, internal resistance and amplification factor) are similar to those of the triode and pentode. Moreover, the current to the control electrode is the same as the reverse current of a *p-n* diode, which need not exceed 10^{-9} A when silicon is used. This is of the same order of magnitude as the grid current of a good-quality valve. The only difference is that the voltage V_{DS} is considerably smaller than the anode-cathode voltage in a valve. A typical value for this is approx. 10 volts. This is high, however, when compared with the normal collector voltage of an ordinary transistor.

The concept of the field-effect transistor is relatively old. In 1928, Lilienfeld applied for a patent on "a device for controlling electric current" (USP 1,900,018). His principle was illustrated by a sketch, corresponding to Fig. 20-14. The field-effect electrode (*G*) is insulated from a thin layer of semiconductor material (copper sulphide) by means of a layer of insulating material (aluminium oxide). When the voltage at (*G*) is changed, the number of charge carriers induced capacitatively into the semiconductor will also change, and hence the conductance between the two terminals *S* and *D*. In order to achieve a large relative change in conductance, it is necessary that the semiconductor slice is extremely thin. Lilienfeld proposed making a notch in the material for this purpose. However, due to various surface effects, which were not

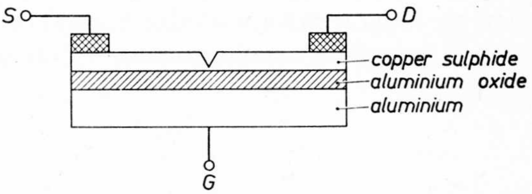

Fig. 20-14

understood until much later, the available control proved to be rather small. In order to avoid these surface effects, Shockley proposed in 1952 the use of a reversed $p-n$ junction, which led to the development of the above-mentioned field-effect transistor. The last few years have witnessed such an improvement in the technology of thin layers that it is now possible to make usable transistors which differ basically very little from Lilienfeld's configuration. The most advanced of this type of transistor is now the MOS (metal-oxide-semiconductor)-transistor where the conductance in a thin channel at the surface of a silicon crystal is controlled by a field-effect electrode, the latter being insulated by means of a thin layer of SiO_2. The current to this electrode is of the order of magnitude of 10^{-12} A.

Because field effect transistors bear a remarkable resemblance to valves as amplifying elements, it will only be necessary to mention them in isolated instances in the following sections.

21. Transistor circuits

Calculations with the transistor as a linear amplification element can be carried out in the same way as for valves. When the collector current I_c and base current I_b can be expressed as functions of the base-emitter voltage V_{be} and the collector-emitter voltage V_{ce}, these functions can be expanded into a power series for a given point of operation. It is again possible to limit ourselves to the linear term for small signal operation, and by only considering deviations from the quiescent position, we obtain an equation for the collector current

$$i_c = S\left(v_{be} + \frac{v_{ce}}{\mu}\right) \tag{21.1}$$

which is analogous to that for a triode.

Because the collector d.c. current changes by approximately a factor e for a change of 25 mV in the base-emitter d.c. potential, it means that the curvature of the corresponding characteristic curve is much stronger than that of the curve for a triode. The signal value for which a linear approximation of the characteristic curve is still allowed, is therefore much smaller than with a triode and does not exceed a few tens of millivolts for most applications.

The exponential relationship between collector current and base-emitter voltage results in the ratio S/I_c being hteoretically independent of the value of the collector current: $I_c = I_{co}\, e^{qV_{be}/kT}$, gives $S = dI_c/dV_{be} = q/kT \cdot I_c$; therefore $S/I_c = q/kT \approx 40$ volts^{-1} at room temperature. With the same collector current, S will decrease by approximately 0.3 per cent for each °C of temperature increase.

As a result of the voltage drop in the base and emitter material, the voltage across the junction is smaller than the voltage V_{be} across the terminals. Since this voltage drop is greater at larger current values, the ratio will also decrease at increasing current values with transistors. Fig. 21-1 shows S/I_c against I_c for a small transistor (OC71). Comparison with the corresponding Fig. 17-7 for a valve shows, that with the same current, this ratio will be 10–30 times larger for a transistor.

As long as the base-collector junction remains reverse biased, the effect of the collector voltage on the collector current is relatively small; in other words, the transistor resembles the pentode in this respect. This is confirmed by the large values which the amplification factor μ has in practice.

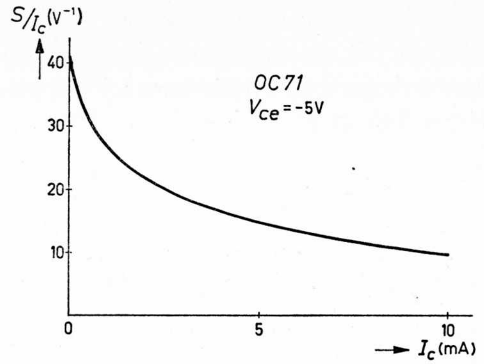

Fig. 21-1

As mentioned in the previous section, the effect of the collector is very small, especially in the case of transistors made with the diffusion technique; μ has values of 1000 and over. With the alloy transistor this value is smaller by one order of magnitude and μ appears to be proportional to $\sqrt{V_{ce}}$ to a good approximation.

The effect of the collector voltage on the collector current can be neglected in many cases. Equation (21.1) can then be simplified

$$i_c = S v_{be} \tag{21.2}$$

The equation for the base current can be written:

$$i_b = \frac{S}{\alpha'}\left(v_{be} - \frac{v_{ce}}{\mu'}\right) \tag{21.3}$$

where α' is the current amplification factor described in the previous section. As follows from the transistor mechanism, α' is hardly dependent on the choice of working point.

The value μ' is of the same order of magnitude as μ. According to the definition (21.3), μ' will be positive with the minus sign in front of the term v_{ce}/μ', because the effect of V_{ce} on I_b is opposite to that of V_{be}. Anyhow, this term can be neglected in almost every case and sufficiently accurate results are obtained with:

$$i_b = \frac{S}{\alpha'} v_{be} \tag{21.4}$$

By applying the simplified expressions (21.2) and (21.4), we can replace one of them by the relation:

$$i_c = \alpha' i_b$$

When equations (21.1)–(21.4) are applied, one should take into account the correct polarity. As the *n-p-n* transistor corresponds in this respect to the valve, the same arrangement is valid here for the positive direction of currents and voltages (Fig. 21-2).

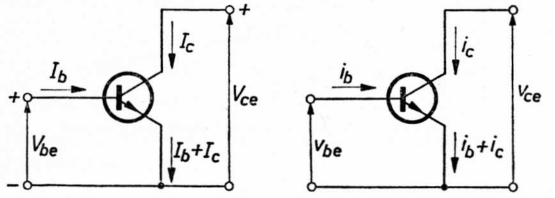

Fig. 21-2

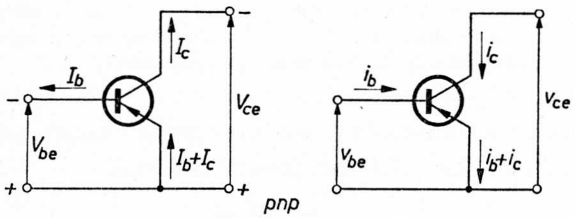

Fig. 21-3

The situation is different for the *p-n-p* transistor, as explained on the left-hand side of Fig. 21-3 where the polarity of the d.c. voltages and the direction of the direct currents are indicated. A change in V_{be} in the positive sense results in a reduction of currents I_b and I_c, that is S would become negative in the above equation if the same direction for the signal voltages and currents was chosen as for the quiescent currents. This makes it advantageous to choose the sign conventions for the signal currents opposite to those for the direct current (right-hand side of Fig. 21-3). In this way, positive values are maintained in the equation, and signal current directions will be analogous to those in the triode and the *n-p-n* transistor, thus obtaining a uniform description for all valves and transistors. This system has been adopted in this book for the calculations that follow.

Much misunderstanding has been caused in recent years because, when considering the transistor as a linear amplification element, all kinds of combinations of the parameters and basic circuits (*h*-, *z*-, *y*-parameters, common base and common emitter circuits, etc.) have been presented as essential for understanding the transistor. This can make it a little inconvenient to use the published transistor characteristics which normally correspond to a few of these parameters or circuits. However, the values used here (S, α', μ and μ') can be easily derived, if necessary, from the "common emitter" characteristic curves, where the voltages have been plotted with respect to the emitter (Fig.21-4).

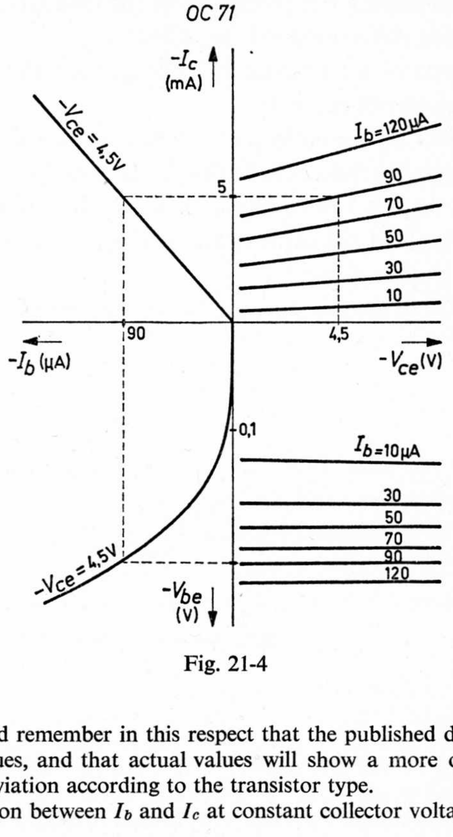

Fig. 21-4

We should remember in this respect that the published data refer to average values, and that actual values will show a more or less pronounced deviation according to the transistor type.

The relation between I_b and I_c at constant collector voltage gives directly α': $\alpha' = \left(\frac{dI_c}{dI_b}\right)_{V_{ce}}$ = constant. It follows from the I_b—V_{be} curve at constant collector voltage: $S/\alpha' = \left(\frac{dI_b}{dV_{be}}\right)_{V_{ce}}$ and therefore S can be found knowing the value α'. We can determine $\mu' = \left(\frac{dV_{ce}}{dV_{be}}\right)_{I_b}$ from the V_{ce}—V_{be} curve, while the I_c—V_{ce} curve yields at constant base current the value $S(1/\mu + 1/\mu') = \left(\frac{dI_c}{dV_{ce}}\right)_{I_b}$, thus permitting one to calculate the still unknown μ. For example, Fig. 21-4 shows that when transistor OC71 is adjusted to $I_c = 5$ mA and $V_{ce} = 4.5$ volts, we find $\alpha' \approx 50$; $S/\alpha' \approx 2$ mA/V, therefore $S \approx 100$ mA/V; $\mu' > 1000$ and $\mu \approx 600$.

Since it is normally possible to neglect the effect of the collector voltage, and to estimate the slope from the quiescent current ($S = 10–30\ I_c$ per volt), it is in many cases sufficient to know α' in order to form an opinion of the overall behaviour of a transistor.

When comparing the properties of the transistor with those of the triode, the following differences will be noted:

1. The slope of a transistor is much greater than that of a triode for the same quiescent current;
2. The effect of the collector voltage on the quiescent current is relatively much smaller than that of the anode voltage, and the transistor behaves in this respect rather like a pentode. The internal resistance $r_a = \mu/S$, however, is of the same order of magnitude as that of a triode, because of the greater slope;
3. Contrary to the grid current, the base current cannot as a rule be neglected.

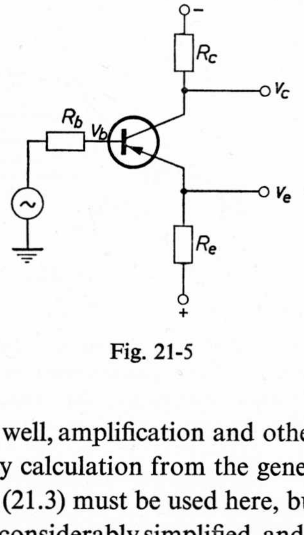

Fig. 21-5

With the transistor as well, amplification and other properties of the basic circuits can be derived by calculation from the general case (Fig. 21-5). The full equations (21.1) and (21.3) must be used here, but we shall see that these calculations can often be considerably simplified, and that it is possible to predict most properties qualitatively. For example, with a.c. amplifiers we usually meet the case that the emitter is decoupled by a very large capacitor, so that the emitter voltage may be assumed to be zero for frequencies in the desired frequency range. This gives for these frequencies the circuit shown in Fig. 21-6, where

$$i_c = Sv_b + \frac{S}{\mu}v_c,\ \ i_b = \frac{S}{\alpha'}v_b - \frac{S}{\alpha'\mu'}v_c,\ \ v_b = v_i - i_bR_b,\ \ v_c = -i_cR_c$$

and therefore:

$$v_c\left[\frac{1}{R_c}+\frac{S}{\mu}+\frac{S}{\mu'}\cdot\frac{1}{1+\dfrac{\alpha'}{SR_b}}\right]=\frac{-S}{1+\dfrac{SR_b}{\alpha'}}v_i \qquad (21.5)$$

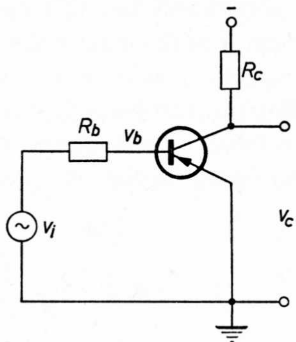

Fig. 21-6

For values of R_b that are small compared with α'/S, the third term in the coefficient for v_c becomes small in comparison with S/μ, and can be neglected. For very large values of R_b this third term is of the same order of magnitude as S/μ'. It follows that the coefficient of v_c always equals $1/R_c$ plus a few times S/μ. R_c is often considerably smaller than μ/S, which gives a further simplification of the coefficient of v_c to $1/R_c$.

The sum of the second and third terms is smaller than the quantity $S(1/\mu + 1/\mu')$ we referred to several times before. A normal value for this with alloy transistors (e.g. OC71) is approximately 0.05 mA/V per mA collector current. If the entire collector d.c. current passes through R_c (thus no d.c. current through a further stage or other load impedance connected to the collector), simplification of the coefficient of V_c to $1/R_c$ is therefore admissible if the d.c. voltage across R_c is considerably smaller than $1/0.05 = 20$ volts. With transistors of the alloy-diffusion type, such as the OC170, the corresponding value will exceed 100 volts, so that the simplification may nearly always be applied here. The d.c. voltage drop across R_c is in practice seldom more than a few tens of volts.

We thus obtain for the amplification:

$$-\frac{v_c}{v_i}=\frac{SR_c}{1+\dfrac{kSR_c}{\mu}}\cdot\frac{1}{1+\dfrac{SR_b}{\alpha'}} \qquad (21.6)$$

with $k\approx 1$.

In the case that $R_c \ll \mu/S$, this may be simplified to:

$$-\frac{v_c}{v_i} = \frac{SR_c}{1 + \dfrac{SR_b}{\alpha'}} \tag{21.7}$$

In the case of valves it is extremely rare that the resistance of the signal source is large in comparison with the input resistance of the grid, so that, practically speaking, the signal comes from a current source. This occurs much more frequently with transistors because of their low input resistance. The relation in that case is found by substituting $R_b i_i$ for v_i, where i_i is the input current in (21.5), and taking the limit $R_b \to \infty$. We then find:

$$-\frac{v_c}{i_i} = \frac{\alpha' R_c}{1 + SR_c\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)} \tag{21.8}$$

and if $R_c \ll \mu/S$

$$-\frac{v_c}{i_i} = \alpha' R_c \quad \text{or} \quad \frac{i_c}{i_i} = \alpha' \tag{21.9}$$

With the emitter follower of Fig. 21-7, analogous to the cathode follower, the difference between base and emitter voltage will become relatively small when the emitter resistance is large with respect to $1/S$. Just how close the emitter voltage follows the signal voltage v_i is also determined by resistor R_b in the base circuit.

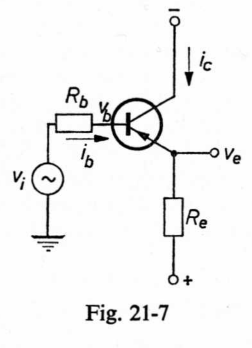

Fig. 21-7

By putting $R_b = 0$, we find for the circuit shown in Fig. 21-7:

$$v_c = 0, \; v_e = (i_b + i_c)R_e, \; v_b = v_i$$

which, when substituted in equations (21.1) and (21.3), yields:

$$\frac{v_e}{R_e} = S\left(1 + \frac{1}{\alpha'}\right)v_i - S\left(1 + \frac{1}{\mu} + \frac{1}{\alpha'} - \frac{1}{\alpha'\mu'}\right)v_e.$$

Therefore: $$v_e\left(\frac{1}{SR_e} + 1 + \frac{1}{\mu} + \frac{1}{\alpha'} - \frac{1}{\alpha'\mu'}\right)\frac{1}{1 + \dfrac{1}{\alpha'}} = v_i$$

or, to a very good approximation with $\alpha' > 10$ and $SR_e \gg 1$:

$$v_e\left(1 + \frac{1}{SR_e} + \frac{1}{\mu}\right) \approx v_i$$

Thus: $$\frac{v_e}{v_i} \approx 1 - \left(\frac{1}{SR_e} + \frac{1}{\mu}\right) \tag{21.10}$$

If R_b is not equal to zero, equation (21.10) is valid for v_e/v_b instead of v_e/v_i.

Thus: $v_e = v_b(1-\varepsilon)$ where $\varepsilon = 1/SR_e + 1/\mu$. We also have: $v_b = v_i - i_b R_b$. The base current is calculated from (21.3) with $v_c = 0$:

$$i_b = \frac{S}{\alpha'}v_b\left(\varepsilon + \frac{1}{\mu'}\right) \quad \text{so that} \quad v_b\left\{1 + \frac{SR_b}{\alpha'}\left(\varepsilon + \frac{1}{\mu'}\right)\right\} = v_i$$

and therefore:

$$\frac{v_e}{v_i} = \frac{1-\varepsilon}{1 + \dfrac{SR_b}{\alpha'}\left(\varepsilon + \dfrac{1}{\mu'}\right)}$$

which can be approximated for values of SR_b/α' which are small with respect to $(\varepsilon + 1/\mu')^{-1}$ by:

$$\frac{v_e}{v_i} = 1 - \varepsilon\left(1 + \frac{SR_b}{\alpha'}\right) - \frac{SR_b}{\alpha'\mu'}$$

thus:

$$\frac{v_e}{v_i} = 1 - \left[\left(\frac{1}{SR_e} + \frac{1}{\mu}\right)\left(1 + \frac{SR_b}{\alpha'}\right) + \frac{SR_b}{\alpha'\mu'}\right]. \tag{21.11}$$

Example: We find for the OC71, when $I_e = 5$ mA (see p. 91): $S \approx 10^{-1}$ A/V; $\mu \approx 600$; $\mu' \approx 1000$; $\alpha' \approx 50$ which gives for $R_e = 10^4\ \Omega$ and $R_b = 10^3\ \Omega$:

$$\frac{v_e}{v_i} = 1 - \left[\left(\frac{1}{1000} + \frac{1}{600}\right)(1+2) + \frac{2}{1000}\right] \approx 1 - \frac{1}{100}$$

Equation (21.11) can often be approximated at the very high values of μ and μ' with the latest types of transistors by:

$$\frac{v_e}{v_i} = 1 - \frac{1}{R_e}\left(\frac{1}{S} + \frac{R_b}{\alpha'}\right)$$

However, when R_e is made very large artificially, this approximation is no longer permissible.

From the above it follows that calculations for transistorized circuits are more involved than those for corresponding valve circuits. It is, however, often possible to simplify the calculations very considerably (without necessarily diminishing the accuracy) by introducing permissible approximations at the correct point in the calculations or, as in the case of the emitter follower, by separating the effect of a number of parameters.

The output impedances of the basic circuits can be calculated in a similar manner to valve circuits. The results can, however, be quite well predicted qualitatively. For example, the part of the collector output impedance determined by the transistor will be as high as with the anode output impedance of a valve, but the effective increase of this impedance by means of resistor R_e in the emitter circuit will be less than with the triode because collector and emitter currents are not equal to each other.

The following equations apply to the circuit shown in Fig. 21-8:

$$i_c = S\,(v_b - v_e) + \frac{S}{\mu}(v_c - v_e)$$

$$i_b = \frac{S}{\alpha'}(v_b - v_e) - \frac{S}{\alpha'\,\mu'}(v_c - v_e)$$

$$v_b = -\,R_b i_b\,,\ v_e = R_e(i_c + i_b).$$

Elimination of all variables except v_c and i_c gives for the internal resistance of the collector $R_{ic} = |\,v_c/i_c\,|$:

$$R_{ic} = \frac{\alpha' + \left[\alpha'\left(1 + \frac{1}{\mu}\right) + 1 - \frac{1}{\mu'}\right] SR_e + SR_b + S^2R_eR_b\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)}{S\left[\frac{\alpha'}{\mu} + S\,(R_e + R_b)\left(\frac{1}{\mu} + \frac{1}{\mu'}\right)\right]}$$

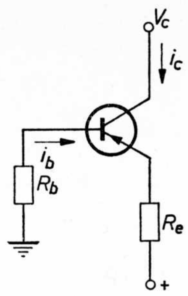

Fig. 21-8

At this stage it is permitted to introduce a number of simplifications. For example, the term $\alpha'(1 + 1/\mu) + 1 - 1/\mu'$ can be approximated to α' at the usual values of μ and μ'.

If it is desired to use the circuit as a quasi-current source (analogous to the corresponding valve circuit) we must choose R_b and R_e in such a way that R_{ic} becomes large. Since dividend and divisor are linear in both R_b and R_e, large values of R_{ic} are found for very large or very small values of R_b and R_e. We can approximate for very large values of R_b by:

$$R_{ic} = \frac{1 + SR_e\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)}{S\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)} = R_e + \frac{1}{S\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)} \approx$$

$$\approx R_e + \frac{1}{S\,\dfrac{2}{\mu}} = R_e + \frac{\mu}{2S}$$

This does not give any advantage worth mentioning against R_e or μ/S. If R_b is made as small as possible:

$$R_{ic} = \frac{\alpha'(1 + SR_e)}{S\left[\dfrac{\alpha'}{\mu} + SR_e\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)\right]}$$

It is obvious that R_e must not be small as otherwise $R_{ic} = \mu/S$. Because S is large we have $SR_e \gg 1$ even for relatively small values of the emitter resistance, and we can then further approximate to:

$$R_{ic} = \frac{\alpha' R_e}{\dfrac{\alpha'}{\mu} + SR_e\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)} = \frac{1}{\dfrac{1}{\mu R_e} + \dfrac{S}{\alpha'}\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)}$$

This means that R_{ic} can be considered as a parallel combination of μR_e and $\dfrac{\alpha'}{S\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)}$, Fig. 21-9.

We have already seen that the quantity $S(1/\mu + 1/\mu')$ can be derived

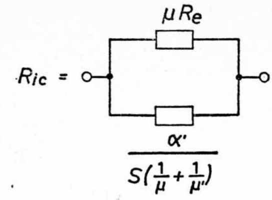

Fig. 21-9

from the $I_c - V_{ce}$ curve at the corresponding constant base current. It is obvious that R_{ic} will be larger when μ and μ' are larger. This therefore applies particularly to the recent types of transistors. For larger values of R_e, R_{ic} approximates to $\dfrac{\alpha'}{S\left(\dfrac{1}{\alpha} + \dfrac{1}{\alpha'}\right)}$. This gives a value of approx. 0.5 MΩ for alloy transistor OC71 at a standing current of 5 mA. In the case of alloy-diffusion transistor OC171, this values becomes approximately 10 MΩ.

When the base is connected to a source of low internal resistance, the emitter follower circuit of Fig. 21-7 has an output resistance of approximately $1/S$ which can be made very small, namely approximately 25 Ω/I_c (I_c in mA). It can be checked quite easily that this output resistance will become larger when R_b increases. We thus find for the internal resistance of the emitter $R_{ie} = \left| v_e/i_e \right|$:

$$R_{ie} = \frac{1}{S} + \frac{R_b}{\alpha'}$$

R_{ie} is calculated from the complete transistor equations with $v_c = 0$ and $v_b = -R_b i_b$:

$$R_{ie} = \frac{1}{S} \cdot \frac{\alpha' + SR_b}{SR_b\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right) + \alpha'\left(1 + \dfrac{1}{\mu} + \dfrac{1}{\alpha'} - \dfrac{1}{\alpha'\mu'}\right)}$$

which can be written to a good approximation as:

$$R_{ie} = \frac{1}{S} \cdot \frac{\alpha' + SR_b}{\alpha' + SR_b\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)}$$

In almost all practical applications

$$R_b \ll \frac{\alpha'}{S\left(\dfrac{1}{\mu} + \dfrac{1}{\mu'}\right)}$$

and the above expression will be equal to $\dfrac{1}{S} + \dfrac{R_b}{\alpha'}$

For extremely large values of R_b, the expression approaches the limiting value $\dfrac{1}{S\left(\dfrac{1}{\mu}+\dfrac{1}{\mu'}\right)}$.

The expression $R_{ie} = \dfrac{1}{S} + \dfrac{R_b}{\alpha'}$ could also be derived from the simplified relations of (21.2) and (21.4).

It follows from the above that the output resistance of the basic circuits with transistors hardly ever differ more than by one order of magnitude from those of the corresponding valve circuits. However, the difference is very much greater in the case of the input resistance.

The input resistance of a valve is determined by the extent to which the grid current changes with the grid voltage. Since the grid current can quite easily be smaller than 10^{-7}–10^{-8} A, and the changes in that current for changes in the grid voltage are at the most of the same order of magnitude, the input resistance of a valve circuit can easily exceed 10–100 MΩ.

The situation with the transistor is entirely different because the base current usually constitutes a non-negligible fraction of the collector current and furthermore varies strongly with the base and emitter voltages. By using the simplified transistor equations for the circuit of Fig. 21-10, where the effect of the collector voltage is neglected, we obtain

$$i_c = Sv_{be}\,,\quad i_b = \frac{S}{\alpha'}\,v_{be}\,,\quad v_e = R_e(i_c + i_b)\quad,\quad v_i = v_e + v_{be}$$

so that

$$R_{ib} = \frac{v_i}{i_b} = \frac{\alpha'}{S} + (\alpha' + 1)\,R_e$$

By putting $R_e = 0$, this yields:

$$R_{ib} = \alpha'/S\,.$$

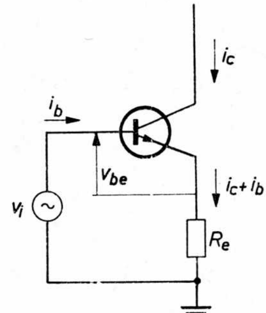

Fig. 21-10

Since $S \approx 40\ I_c$ per Volt, the input resistance will be of the order of 1 kΩ for $\alpha' = 30$–50 and $I_c = 1$ mA. This relation shows, however, that the input resistance increases considerably when a relatively small resistor is inserted in the emitter circuit. With $\alpha' = 100$ and R_e only 1 kΩ, the input resistance has already risen to more than 100 kΩ. The emitter follower will therefore have quite a high input resistance when R_e is large. This, together with the small output impedance, makes the circuit, in common with the cathode follower, therefore suitable for decreasing the impedance level of signal sources.

When two stages are coupled by means of a capacitor or when, for example, an emitter is decoupled, the same considerations as applied to valves are valid. The capacitor values will be much higher because of the lower impedance level of most transistor circuits. For example, if in one stage of an audio amplifier ($\omega_{min} = 100$) the emitter is decoupled with a capacitor C, its impedance must be small with respect to the output impedance of the emitter, which approximately equals $1/S$ for a small source resistance. This gives for C: $C \gg S/\omega_{min}$. At a quiescent current of a few milliamps, S soon reaches a value of 0.1 A/V, and a value of a few thousand microfarads is required for C. If the source impedance (including the resistors in the base circuit) is larger, the output impedance will also be larger, and C can be correspondingly smaller.

Amplifier stages with cascade circuitry can, in the case of transistors, be designed in the same way as with valves. As the collector-emitter potential often needs to be only a small part of the total supply voltage, a base input can be directly connected to the collector output of the previous stage, thus eliminating the use of a coupling capacitor. This is because the required bias across the base-collector junction is only about 0.5 volt, and for signals which are less than, say, 1 volt, we can content ourselves with $V_{ce} = 1.5$ volt, while a value of 10–15 volts is usual for the supply. It is also possible to follow one stage with *p-n-p* transistors by another stage with *n-p-n* transistors.

Fig. 21-11 gives an example of a two-stage amplifier with capacitive coupling at the input and direct coupling between the stages. Once again, the calculation of the total amplification is more complicated than for valve amplifiers because the loading of one stage by the next must be considered. In the example, we first calculate the input impedance Z_{i2} of the second stage and then the voltage amplification A_1 of the first stage, whereby we take as the collector impedance $R_{c1}//Z_{i2}$. The total voltage amplification then equals the product of A_1 and the voltage amplification of the second stage.

When Z_{i2} is much smaller than R_{c1}, almost all the collector signal current

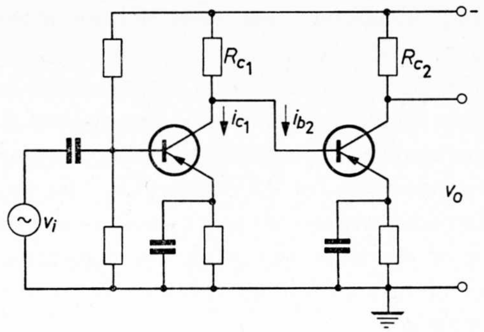

Fig. 21-11

of the first stage will pass into the base of the second stage. In that case $i_{b2}=-i_{c1}$ applies and thus $i_{c2}\approx-\alpha'_2 i_{c1}$, with α'_2 the current amplification factor of the second stage. i_{c1} can now be calculated neglecting the collector feedback action. This gives $i_{c1}=S_1 v_i$, and the total amplification is therefore:

$$\frac{v_o}{v_i} = \alpha'_2 S_1 R_{c2}$$

with possibly a small correction when R_{c2} is not small with respect to μ_2/S_2.

Naturally, the same combination possibilities exist for transistors as for triodes. For example, one can make transistor cascodes (Fig. 21-12). These are, however, rarely used because there is little need for them in view of the already small collector feedback action in the case of a single transistor, and because they lack some of the advantages of valve cascodes, not all of these having been mentioned so far.

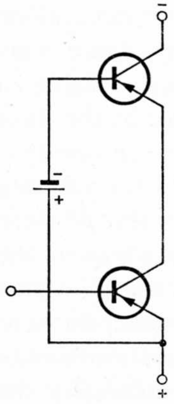

Fig. 21-12

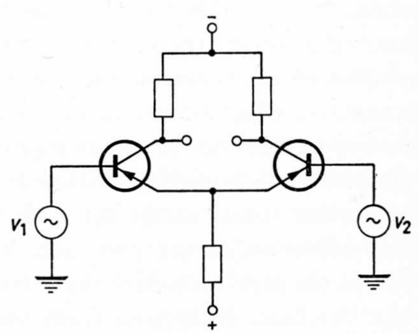

Fig. 21-13

The transistorized "long-tailed pair" (Fig. 21-13) is more valuable in this respect because it possesses the same specific properties as the valve model. It mainly amplifies the voltage difference between both bases and has the same limiting effect in the case of overloading signals. Although the sum of the quiescent emitter currents is stabilized by the common emitter resistor, this is not the complete case for the distribution over the two transistors. Temperature differences between the two transistors exert a great influence. However, because of the balancing properties, considerable compensation does occur for common changes in temperature. We shall refer to this when discussing d.c. amplifiers.

Regarding the frequency dependence of the amplification, we can first mention that the early transistors had large capacitive effects that had to be taken into account in almost all applications. Modern transistors, however, made with the use of diffusion techniques, have such small capacitances. that it is often possible to select a transistor where the effect of capacitance can be neglected. However, in order to enable us to make such a choice, we must also know which effects occur and what can be expected.

We should also note that because the impedance level is so much lower for transistorized circuits than for valve circuits, the effect of parasitic capacitances is usually much smaller, and the limit of the amplification at high frequencies is thus mainly determined by the amplification mechanism in the transistors themselves.

For the transistor we have to deal solely with capacitances whose values, excluding those of the wiring, depend on the point of operation. On changing the bias V across a junction, the space charge present will first have to change accordingly (Poisson's equation). The relevant calculations show that, at least for alloy transistors, the value of the capacitance is inversely proportional to $\sqrt{V}$. This voltage dependence is usually slightly smaller for diffusion transistors. The value is further determined by the dimensions and impurities of the materials. For the base-collector junction this is the only capacitative effect which, in transistors intended for very high frequencies, can be kept to values as small as some tenths of picofarads. Although the effect is larger for the base-emitter junction than for the base-collector junction, it can, except for v.h.f. transistors, still be neglected in comparison with the effective capacitance C_{diff} which results from the process, mentioned in the previous section, for replenishment of the minority charge carriers in the base. It appears from the relevant calculation that this capacitance is proportional to the emitter current. For reasons to be explained, this relation is given in the form

$$C_{\text{diff}} = \frac{1}{\omega_{\text{lim}}} \cdot \frac{qI_e}{kT}$$

where ω_{lim} is the limit or cut-off frequency of the transistor.

We have now listed the most important capacitive effects and one may wonder just how they can be taken into account in design calculations. For this we use the original transistor equations. Especially for h.f. transistors, the feedback action of the collector may be entirely neglected:

$$i_c = Sv_{be}$$
$$i_b = \frac{S}{\alpha'} v_{be}$$

As we have noted, slope S would be proportional to I_c if no voltage losses occur anywhere else than at the junction. The proportionality factor here is $q/kT = 40$ volt^{-1}. Therefore:

$$\begin{cases} i_c = S^* \, v_{b'e'} \\ i_b = \dfrac{S^*}{\alpha'} v_{b'e'} \end{cases} \tag{21.12}$$

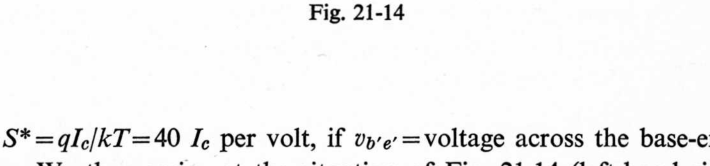

Fig. 21-14

where $S^* = qI_c/kT = 40\ I_c$ per volt, if $v_{b'e'}$ = voltage across the base-emitter junction. We thus arrive at the situation of Fig. 21-14 (left-hand side): a resistance α'/S^* is inserted between internal base b' and emitter e'; a current $S^* v_{b'e'}$ passes between collector and emitter. We can now introduce the capacitances of the base-emitter junction (C_{diff}) and of the base-collector junction ($C_{b'c'}$) into this arrangement, as shown on the right-hand side of Fig. 21-14.

We find for the parallel combination of α'/S^* and $C_{\mathrm{diff}}=S^*/\omega_{\mathrm{lim}}$

$$\frac{\alpha'}{S^*} \,//\, C_{\mathrm{diff}} = \frac{1}{\dfrac{S^*}{\alpha'} + \mathrm{j}\omega\dfrac{S^*}{\omega_{\mathrm{lim}}}} = \frac{1}{\dfrac{S^*}{\alpha'}\left(1 + \mathrm{j}\alpha'\dfrac{\omega}{\omega_{\mathrm{lim}}}\right)} = \frac{\alpha'_\omega}{S^*}$$

where
$$\alpha'_\omega = \frac{\alpha'}{1 + \mathrm{j}\alpha'\dfrac{\omega}{\omega_{\mathrm{lim}}}}$$

It follows that ω_{lim} represents the frequency where $|\alpha'_\omega|$ has almost reached the value 1.

Substitution now gives us Fig. 21-15, left-hand side, which indicates that in those cases where the effect of the very small capacitance $C_{b'c'}$ can be neglected, α' should only be replaced by α'_ω in equation (21.12) in order to take the frequency dependence of the transistor into account.

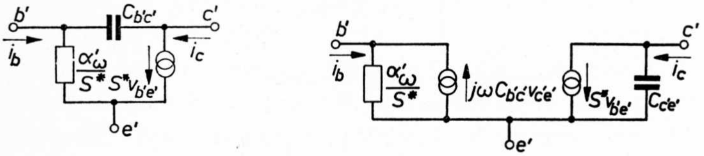

Fig. 21-15

A current $\mathrm{j}\omega C_{b'c'}v_{c'b'}=\mathrm{j}\omega C_{b'c'}(v_{c'e'}+v_{e'b'})$ passes through $C_{b'c'}$. For the base the second term means an increase of C_{diff} by capacitance $C_{b'c'}$, which can nearly always be neglected. The first term gives the influence of the collector-emitter voltage on the base current. As for loading the collector the second term can also be neglected while the first can be taken into account by increasing the already present parasitic capacitance between c' and e' by $C_{b'c'}$ to $C_{c'e'}$. This gives us the situation shown on the right-hand side of Fig. 21-15. The following equations apply:

$$\left\{\begin{aligned} i_c &= S^* v_{b'e'} + \mathrm{j}\omega C_{c'e'} v_{c'e'} \\ i_b &= \frac{S^*}{\alpha'_\omega} v_{b'e'} - \mathrm{j}\omega C_{b'c'} v_{c'e'} \end{aligned}\right. \qquad (21.13)$$

These equations have the same form as the original equations (21.1) and (21.3). The picture can now be completed by incorporating the material resistances situated between the external terminals b, e and c and the internal

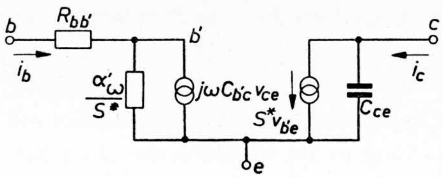

Fig. 21-16

points b', e' and c'. The resistance of the emitter may be neglected here, whilst that of the collector is usually small with respect to the load resistance, so that the diagram of Fig. 21-16 is a sufficiently accurate representation of the situation. The following equations apply:

$$\left\{\begin{aligned} i_c &= S^* v_{b'e} + \mathrm{j}\omega C_{ce} v_{ce} \\ i_b &= \frac{S^*}{\alpha'_\omega} v_{b'e} - \mathrm{j}\omega C_{b'c} v_{ce} \\ v_{bb'} &= i_b R_{bb'} \end{aligned}\right. \tag{21.14}$$

Obviously it is possible to eliminate v_b' from these equations as in fact has been done in the original equations (21.1) and (21.3), but at higher frequencies the above description is to be preferred because of the simplicity of the coefficients.

Resistance $R_{bb'}$ follows directly from transconductance S. We have for low frequencies $i_b = \dfrac{v_{be}}{R_{bb'} + \dfrac{\alpha'}{S^*}}$ (Fig. 21-16) but also, according to equation (21.4): $i_b = \dfrac{S}{\alpha'} v_{be}$. Therefore: $R_{bb'} = \alpha'(1/S - 1/S^*)$

For large currents $1/S^*$ approximates zero and we have $S = \alpha'/R_{bb'}$.

It appears from the formula for α'_ω that the influence of C_{diff} becomes noticeable at frequencies in the neighbourhood of $\omega_{\mathrm{lim}}/\alpha'$. The magnitude of this influence is determined by the circuit; by making the emitter voltage follow the base voltage, the impedance between base and emitter is seemingly increased, and thus C_{diff} apparently decreased.

Let us take the h.f. transistor OC169 as an example. We have here, with the usual tolerances:

$$\omega_{\mathrm{lim}} = 4 \cdot 10^8 \text{ rad/sec.}; \ \alpha' = 100; \ C_{b'c} = 2 \text{ pF and } C_{ce} = 6 \text{ pF}.$$

At 1 mA, S^* has a value of 0.04 A/V, so that we find for C_{diff}:

$$C_{\mathrm{diff}} = \frac{S^*}{\omega_{\mathrm{lim}}} = 100 \text{ pF}.$$

The influence of $C_{\rm diff}$ thus becomes noticeable at frequencies in the neighbourhood of $\omega_{\rm lim}/\alpha' = 4 \cdot 10^6$. If this transistor is designed as an amplifier stage with decoupled emitter, for signals of approx. 1 Mc/s, the effect of $C_{b'c}$ on the input impedance due to the Miller effect will be small compared to that of $C_{\rm diff}$, as long as the amplification does not grossly exceed 10. In order to obtain this amplification we need a collector resistance of a few hundred ohms. The impedance of C_{ce} at 1 Mc/s amounts to approx. $2 \cdot 10^4\ \Omega$ so that at this frequency, which lies in the neighbourhood of $\omega_{\rm lim}/\alpha'$, it will be sufficient to a first approximation to take into account only the frequency dependence of α'. As a rule of thumb we can therefore assume that for frequencies $\omega < \omega_{\rm lim}/10\alpha'$, the transistor will not show any frequency dependence; for frequencies $\omega_{\rm lim}/10\alpha' < \omega < \omega_{\rm lim}/\alpha'$ only the frequency dependence of α' must be taken into account; whilst for frequencies $\omega > \omega_{\rm lim}/\alpha'$ we must use equations (21.14) and Fig. 21-16.

In principle, the same methods as we have described for valves can be used for adjusting the working point of a transistor. However, some factors, such as the greater temperature dependence of certain transistor parameters make some solutions less suitable, whilst on the other hand we have additional possibilities with the transistor for stabilizing the working point because an important part of the current passes through the base circuit. Before giving some examples of practical circuits, we shall first examine the influence of a number of effects on the zero adjustment of a transistor. For these calculations we shall suppose that the influence of the collector voltage on I_b and I_c can be neglected, which allows us to use the simple equations $i_c = Sv_{be}$ and $i_b = (S/\alpha')v_{be}$.

The effect of a change in temperature can be subdivided into several components. Firstly, with a change in temperature, the quiescent currents in the transistor will increase if the voltages are kept constant. It appears that the relative change in the base current is slightly smaller than that in the emitter or collector current (approx. 6 and 7% in Ge, 8 and 9% in Si). It is nevertheless possible to take account of this temperature effect with sufficient accuracy by assuming a change of 2–2.5 mV in the base-emitter voltage for a change in temperature of 1 °C (Fig. 21-17).

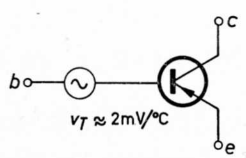

Fig. 21-17

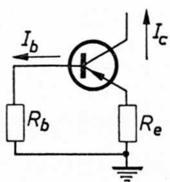

Fig. 21-18

The temperature effect can be taken into account by adding terms to the transistor equations:

$$i_c = S\, v_{be} + \gamma_c I_c \cdot \Delta T$$

$$i_b = \frac{S}{\alpha'} v_{be} + \gamma_b I_b \cdot \Delta T$$

where γ_c and γ_b represent the temperature coefficients and I_c and I_b the quiescent values of the collector and base current. With the help of these equations and $v_b = -i_b R_b$ and $v_e = (i_b + i_c)R_e$, we find for the circuit of Fig. 21-18:

$$v_{be}\left\{\frac{1}{S} + \frac{R_b + R_e}{\alpha'} + R_e\right\} = -\frac{1}{S}\left\{\gamma_b I_b(R_b + R_e) + \gamma_c I_c R_e\right\}\Delta T$$

When $I_c \gg I_b$, this can be reduced to

$$v_{be}\left\{\frac{1}{S} + \frac{R_b}{\alpha'} + R_e\right\} = -\frac{1}{S}\left\{\gamma_b I_b R_b + \gamma_c I_c R_e\right\}\Delta T$$

By introducing the approximation $I_c = \alpha' I_b$ we obtain:

$$v_{be}\left\{\frac{1}{S} + \frac{R_b}{\alpha'} + R_e\right\} = -\frac{I_c}{S}\left\{\gamma_b \frac{R_b}{\alpha'} + \gamma_c R_e\right\}\Delta T$$

Provided $R_b/\alpha' + R_e \gg 1/S$, v_{be} will have a value $-\gamma(I_c/S)\,\Delta T$, where $\gamma_b \leqslant \gamma \leqslant \gamma_c$, if $\gamma_b < \gamma_c$. This corresponds to the above-mentioned change in v_{be} of 2–2.5 mV per °C.

The fact that I_b and I_c do not change in the same proportion means that their ratio α' is also temperature-dependent (1–1.5 per cent per °C). The accompanying changes in amplification characteristics can be negated, if necessary, by applying negative feedback (see Section 22).

The second and often most important effect of an increase in temperature is the increase in the leakage current I_{cb0} ($=I_0$ in Fig. 20-6) which passes from the collector to the base through the reverse-biased base-collector junction. With germanium the increase in current is approx. 10 per cent per °C and with silicon approx. 15 per cent, that is approximately doubling the current for each increase of 5–7 °C.

In the case of germanium transistors, I_{cb0} already amounts to a certain percentage of the base current at room temperature, so that with temperature increases of a few tens of degrees, the effect of this leakage current cannot be neglected. On the other hand, I_{cb0} is very small with silicon transistors. Values here are in the order of magnitude of 10^{-8}–10^{-11} A, so that their effect will only become noticeable at very great increases in temperature.

The effect of leakage current I_{cb0} can be accounted for by a current source between the base and the collector of the transistor, as shown on the left-hand side of Fig. 21-19.

Since the direct effect of I_{cb0} in the collector circuit is so small as to be negligible, this current source may also be envisaged between the base and earth (right-hand side of Fig. 21-19).

The "internal" resistance of the current source is the reverse-biased diode resistance of the base-collector junction, i.e. it will be very high with respect to the usual impedances, so that this current source may actually be replaced by an ideal current source.

The total temperature effect can now be represented by Figs 21-17 and -19 as a good approximation and is shown in Fig. 21-20. The effect of I_{cb0} will become greater with increasing resistance in the base circuit. Since $R_{bb'}$, (the internal resistance of the base material) usually amounts to several tens of ohms or more, this means that in the case of germanium transistors, I_{cb0} will nearly always have a greater effect than v_T. With silicon, this will only be the case at much larger values of R_b.

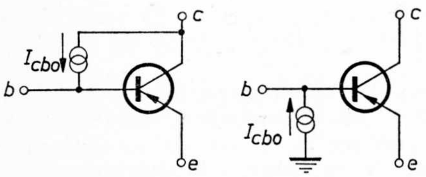

Fig. 21-19

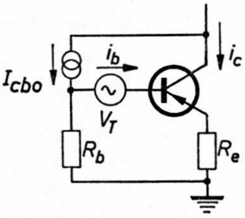

Fig. 21-20

We shall now calculate the effect of I_{cb0} in the circuit of Fig. 21-20 for $v_r = 0$:

$$i_b = \frac{S}{\alpha'} v_{be}, \quad i_c = \alpha' i_b, \quad v_b = -(i_b - I_{cb0})R_b, \quad v_e = (i_c + i_b)R_e.$$

This yields:

$$i_c = \left(\frac{\alpha'}{\dfrac{\alpha'}{SR_b} + 1 + \dfrac{(\alpha'+1)R_e}{R_b}} \right) I_{cb0}$$

or as a good approximation, if $S \gg \alpha'/R_b$

$$i_c = \left(\frac{\alpha'}{1 + \dfrac{\alpha' R_e}{R_b}} \right) I_{cb0}$$

Expressed in another way: for very large values of R_b, or when $R_e = 0$, i_c will be approximately α' times larger than I_{cb0}. This current is easier to measure and is often indicated by I_{ce0}.

Example: Ge-transistor OC71: $I_{ce0} < 325\ \mu A$ at room temperature; $\alpha' > 30$; therefore $I_{cb0} < 11\ \mu A$.

At a collector current of 3 mA, the base current is approx. 100 μA, so that I_{cb0} can amount to 10 per cent of the base current.

Si-transistor BCZ 100: the data of this transistor indicate that both I_{cb0} and I_{ce0} are smaller than 10^{-9} A at room temperature, so that with $\alpha' = 20$ and a collector current of 2 mA, I_{cb0} will be less than 0.001 per cent of the base current.

Apart from the temperature effect, one often wishes to know how the adjustment of the working point of an amplifier stage will change when replacing a transistor. The greatest difference which occurs between two specimens of the same type concerns the current amplification factor α'. It is therefore useful to examine the effect of a change in this parameter on the working point of an amplifier stage.

Let us consider the circuit shown on the left-hand side of Fig. 21-21, which gives the situation for the quiescent currents and voltages.

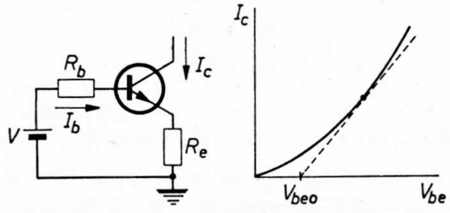

Fig. 21-21

We select the following transistor equations:

$$I_c = S(V_{be} - V_{be0}) \quad \text{and} \quad I_c = \alpha' I_b$$

where V_{be0} is a few tenths of a volt (see right-hand side of Fig. 21-21). We also have:

$$V - V_b = I_b R_b,$$

where V is the source voltage between base and emitter resistance;

$$V_e = (I_c + I_b) R_e$$

This yields:

$$I_b \left\{ \frac{\alpha'}{S} + R_b + (\alpha' + 1) R_e \right\} = V - V_{be0}$$

It follows that

$$\frac{\Delta I_b}{I_b} = -\frac{1}{1 + \dfrac{R_b + R_e}{R_e + S^{-1}} \cdot \dfrac{1}{\alpha'}} \cdot \frac{\Delta \alpha'}{\alpha'}$$

and when putting $I_c = \alpha' I_b$:

$$\frac{\Delta I_c}{I_c} = \frac{1}{1 + \dfrac{R_e + S^{-1}}{R_b + R_e} \cdot \alpha'} \cdot \frac{\Delta \alpha'}{\alpha'}$$

or

$$\frac{\Delta I_b}{I_b} = -\frac{1}{1+k} \cdot \frac{\Delta \alpha'}{\alpha'} \quad \text{and} \quad \frac{\Delta I_c}{I_c} = \frac{1}{1+k^{-1}} \cdot \frac{\Delta \alpha'}{\alpha'}$$

where

$$k = \frac{R_b + R_e}{R_e + S^{-1}} \cdot \frac{1}{\alpha'}$$

It follows from these formulae that only when R_b is large with respect to R_e and $1/S$, will the relative change in I_b be much smaller than the one in α'. In all other cases, the relative change in I_c is, on the contrary, rather small.

Examples: Transistor OC71 with $\alpha' = 30$ is used in the circuit of Fig. 21-21: $R_b = 2$ kΩ; $R_e = 1$ kΩ and $V = 10$ volts. I_e will be approximately 10 V/1 kΩ = 10 mA; S is about 125 mA/V in this case. When applying the above formulae, we find for the currents $I_b \approx 320$ μA and $I_c \approx 9.6$ mA. With $R_e \gg 1/S$, factor k becomes:

$$k = \frac{R_b + R_e}{R_e} \cdot \frac{1}{\alpha'} = \frac{1}{10}$$

so that:

$$\frac{\Delta I_b}{I_b} \approx -\frac{\Delta \alpha'}{\alpha'} \quad \text{and} \quad \frac{\Delta I_c}{I_c} \approx \frac{1}{11} \cdot \frac{\Delta \alpha'}{\alpha'}$$

It is now easy to examine how a number of circuits will behave regarding the selection of the working point.

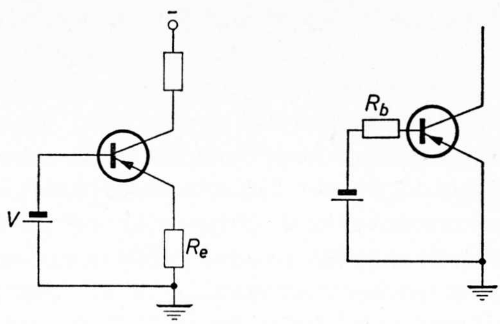

Fig. 21-22

For circuits having little or no base resistance (left-hand side of Fig. 21-22) the temperature sensitivity is determined by v_T. Its effect depends on the values of R_e and V. In the limiting case when no emitter resistance is present, and V is only a few tenths of a volt, v_T will exercise a maximum effect. In this case, I_b and I_c will also both vary strongly with α', while at greater values of R_e, only the stability of I_c will improve.

Because the quiescent voltage of the base lies between collector and emitter, it is not possible to apply the valve circuit with "automatic" bias voltage, Fig. 13-3, here.

With the schematic circuit shown on the right-hand side of Fig. 21-22, at sufficiently large values of R_b, the temperature effect will be determined by I_{cb0}, while I_b will then be insensitive to variations in α'.

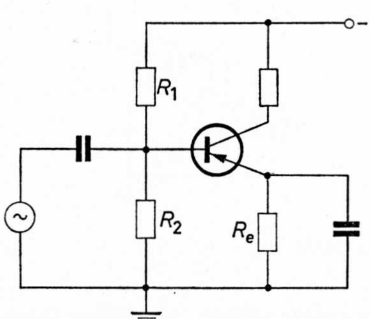

Fig. 21-23

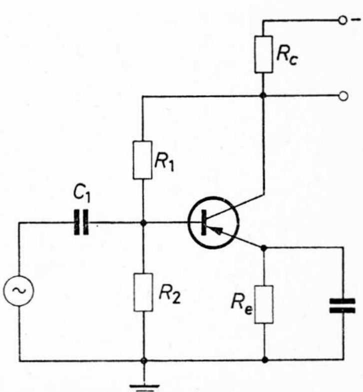

Fig. 21-24

The circuit of Fig. 21-23 can be used with a.c. amplifiers. In normal operation, R_1 and R_2 will be so large that the temperature effect is determined by I_{cb0}. Factor k is here usually about 1 (i.e. $\alpha' R_e \approx R_1//R_2$) so that I_b and I_c are both sensitive to changes in α'.

Fig. 21-24 shows a variation of this circuit in which a change in collector voltage is transferred to the base. The consequent change in the collector current causes a change in collector voltage in the opposite direction so that the whole action is self-opposing. In other words: due to negative feedback (Section 22) a certain amount of stabilization of collector current and voltage occurs. However, because of the small "loop gain", the efficiency of this circuit is not very great.

The effect of feedback in the circuit shown on the left-hand side of Fig. 21-25 is calculated for $R_e = 0$. (At large values of R_e the collector current will already be stabilized, so that there would be no sense in using the circuit.) This feedback action can be taken into account as indicated on the right-hand side of Fig. 21-25, where $r = R_2/(R_1 + R_2)$ and $R_i = R_1R_2/(R_1 + R_2)$.

With $\quad i_c = \alpha' i_b = S v_b, \quad v_c = -\, i_c R_c, \quad i_b = \Delta I + i_t$ and

$$i_t = (r v_c - v_b)/R_i$$

we find:

$$i_b \left\{ 1 + \frac{1}{R_i}\left(\alpha' r R_c + \frac{\alpha'}{S} \right) \right\} = \Delta I$$

or, with $rR_c > \dfrac{1}{S}$

$$i_b \left(1 + \frac{\alpha' r R_c}{R_i} \right) = \Delta I$$

so that the reduction factor is $1 + \alpha' r R_c / R_i$. This shows that with $r < 1$, this feedback is only effective when R_i is of the same order of magnitude as R_c. However, in this case the input resistance of the circuit will be rather low.

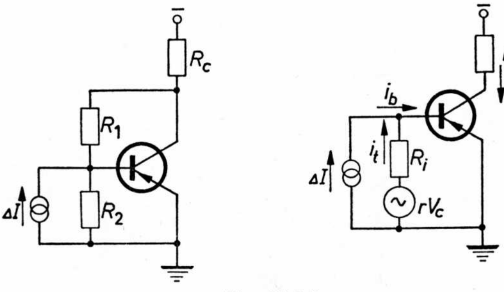

Fig. 21-25

Finally a few notes on the maximum permissible ratings. Exceeding these ratings with valves, particularly values concerning voltages, does not necessarily invite disaster, but for transistors it is absolutely essential to keep within the indicated ratings.

Exceeding the maximum temperature (Ge: approx. 85 °C; Si: approx. 150 °C) is especially fatal because this may cause a "thermal runaway" effect. The high temperature will result in an increase in the leakage current from the collector to base and therefore also in the collector current. At sufficiently high collector voltages and small collector resistance, the power dissipated in the collector will increase, thus causing an increase in temperature and hence a further increase of the leakage current. Whether this really causes a thermal runaway effect will be determined by the values of the resistors in base, emitter and collector circuits, the value of I_{cb0} and the thermal resistance of the collector. This last point often poses a problem, particularly with power transistors.

We can illustrate this by considering the simple case presented in Fig. 21-26. We have for the changes in current: $i_b = i_{cb0} - v_b/R_b$, therefore with $i_b = (S/\alpha')v_b$:

$$i_b = \frac{i_{cb0}}{\left(1 + \dfrac{\alpha'}{SR_b}\right)}$$

For values of $R_b \gg \alpha'/S$ this reduces to $i_b = i_{cb0}$ and thus $i_c = \alpha' i_{cb0}$. The dissipation at the collector increases by $\alpha' \, i_{cb0} V_c$, and if the heat resistance of the collector is k (°C per watt), this will produce an increase in temperature of $k\alpha' i_{cb0} \, V_c$ and hence a change in I_{cb0} of the magnitude $\gamma \, k\alpha' V_c I_{cb0} \cdot i_{cb0}$ where γ is the temperature coefficient of I_{cb0}. In order to limit this effect, the multiplication factor $\gamma \, k\alpha' V_c I_{cb0}$ must be smaller than unity.

Taking transistor OC71 as an example, with $\alpha' = 30$, $V_c = 10$ volts, $\gamma = 0.08$, $k = 400$ °C/W, we obtain for I_{cb0} a maximum permissible value of 0.1 mA which is about 8 times greater than the maximum value of I_{cb0} at room temperature. This corresponds to an increase in temperature of about 30 °C, so that at room temperature the collector dissipation must not exceed 30/400 = 75 mW, and the collector current at 10 volts must not exceed 7 mA.

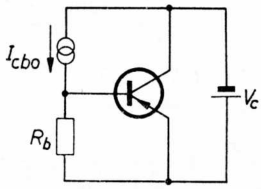

Fig. 21-26

22. Feedback

After having explained the principal characteristics of passive and active components and their most simple combinations, we can now discuss what is probably the most important principle applied in electronics, namely feedback. The application of this principle arises from the need to eliminate some of the disadvantages of the circuits discussed above, as well as to obtain properties which are difficult to achieve by the methods so far mentioned. An example of the first requirement is the fact that amplification depends on values which can change with age or replacement of components, whilst active filters (as discussed in Section 33) belong to the second class of requirements.

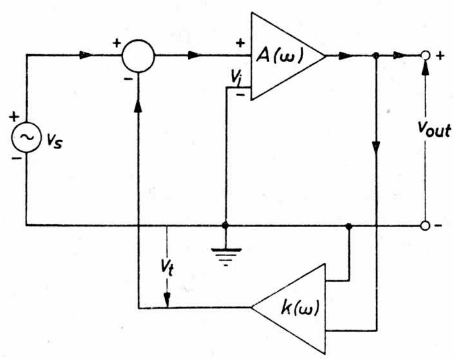

Fig. 22-1

The principle of feedback consists in transferring part of the output signal back to the input. Fig. 22-1 illustrates this. Triangle $A(\omega)$ represents a linear signal transfer system, e.g. an amplifier, and for brevity's sake we shall call it that. The transfer (amplification) of this system, also indicated by $A(\omega)$, will in general be frequency-dependent. We have for the relation between input and output signals of this amplifier, v_i and v_o respectively, at each frequency:

$$v_o = A(\omega) v_i$$

In theory, triangle $k(\omega)$ represents a corresponding system which does not usually have to amplify and therefore, in practice, frequently consists of a passive circuit. It has a transfer $k(\omega)$, which gives us $v_t = k(\omega) v_o$. The circle with the plus and minus signs indicates that the signal v_i is the equal to

$+v_s$ $-v_t$ acquired in a manner which will be explained later. The choice of sign is not essential, but we shall keep to the indicated notation. The arrows in the interconnection paths indicate the signal direction.

It is now easy to calculate the relation between v_o and v_s because

$$v_o = A(\omega)v_i = A(\omega)\,(v_s - v_t) = A(\omega)v_s - A(\omega)k(\omega)v_o$$

so that:

$$[1 + A(\omega)k(\omega)]v_o = A(\omega)v_s$$

and

$$v_o = \frac{A(\omega)}{1 + A(\omega)k(\omega)}\,v_s$$

or

$$A_{fb}(\omega) = \frac{v_o}{v_s} = \frac{A(\omega)}{1 + A(\omega)k(\omega)} \qquad (22.1)$$

where $A_{fb}(\omega)$ = amplification of the system including feedback through $k(\omega)$.

While the above considerations apply to all kinds of signals and combinations of various signals, we shall simplify the discussion by assuming, for the time being, that all signals are voltages.

In many cases one is most interested in the absolute value of the amplification, and sometimes also in the argument indicating the phase shift between input and output voltages. The characteristic of the absolute gain against the frequency is called the amplitude curve; that of the argument, the phase characteristic; and their combination, the frequency characteristic of a system.

Before considering the consequences of the established relation we should have a look at a different aspect. We have seen in Section 7 that the equation in ω is an algebraic simplification of the differential equation valid for any random signal form. This differential equation is re-obtained by replacing $j\omega$ by d/dt. If we make this substitution in (22.1), so that $A(\omega)$ becomes $A^*(d/dt)$ and $k(\omega)$ becomes $k^*(d/dt)$, we have in the feedback system the differential equation:

$$\left\{1 + A^*\left(\frac{d}{dt}\right)\,k^*\left(\frac{d}{dt}\right)\right\}\,v_o = A^*\left(\frac{d}{dt}\right)v_s$$

This means that natural modes of oscillation may occur in the system with "frequencies" p_i which satisfy:

$$1 + A^*(p)k^*(p) = 0 \qquad (22.2)$$

When this equation possesses roots with a positive real part, the amplitudes of the corresponding natural modes will grow. A limiting action will put an end to this growth, but the system will then produce signals even in the absence of an input signal. In this case we speak of an oscillating or generating system. When all roots of equations (22.2) have a negative real part, the natural modes will be damped and the system is called non-oscillating or stable.

The word "stable" is used in electronics both with the above meaning of non-oscillating and in the sense that an element or quantity is almost unaffected by the prevailing circumstances. We shall indicate which sense is meant whenever a misunderstanding could occur through this double meaning.

Let us demonstrate with an example that even simple amplifiers can be made to oscillate by means of feedback. A single-stage amplifier with anode resistance R and stray anode capacitance C has at frequency ω the gain:

$$A(\omega) = \frac{A(0)}{1 + j\omega RC} = \frac{A(0)}{1 + j\omega\tau}$$

We therefore have for a three-stage amplifier:

$$A(\omega) = \frac{A_0}{(1 + j\omega\tau_1)(1 + j\omega\tau_2)(1 + j\omega\tau_3)}$$

with A_0 is the amplification for $\omega = 0$.

If feedback is applied by means of a voltage divider with:

$$k(\omega) = \text{constant} = k_0,$$

we find:

$$A_{fb}(\omega) = \frac{A_0}{(1 + j\omega\tau_1)(1 + j\omega\tau_2)(1 + j\omega\tau_3) + A_0 k_0}$$

Whether oscillation takes place or not is determined by the roots of

$$(1 + p\tau_1)(1 + p\tau_2)(1 + p\tau_3) + A_0 k_0 = 0$$

The sum of the roots is $-(1/\tau_1 + 1/\tau_2 + 1/\tau_3)$ and is therefore independent of $A_0 k_0$. We can make use of this for determining the roots, by means of a simple graph, as the x-values of the points of intersection of curve $y = (x\tau_1 + 1)(x\tau_2 + 1)(x\tau_3 + 1)$ and the straight line $y = -A_0 k_0$ (left-hand side of Fig. 22-2, assuming $\tau_1 > \tau_2 > \tau_3$). The curve intersects the x-axis at points $-1/\tau_1$, $-1/\tau_2$ and $-1/\tau_3$ and the y-axis at point $y = +1$. If $A_0 k_0$ is negative and $|A_0 k_0| > 1$, one of the roots will be positive and real, and will cause a

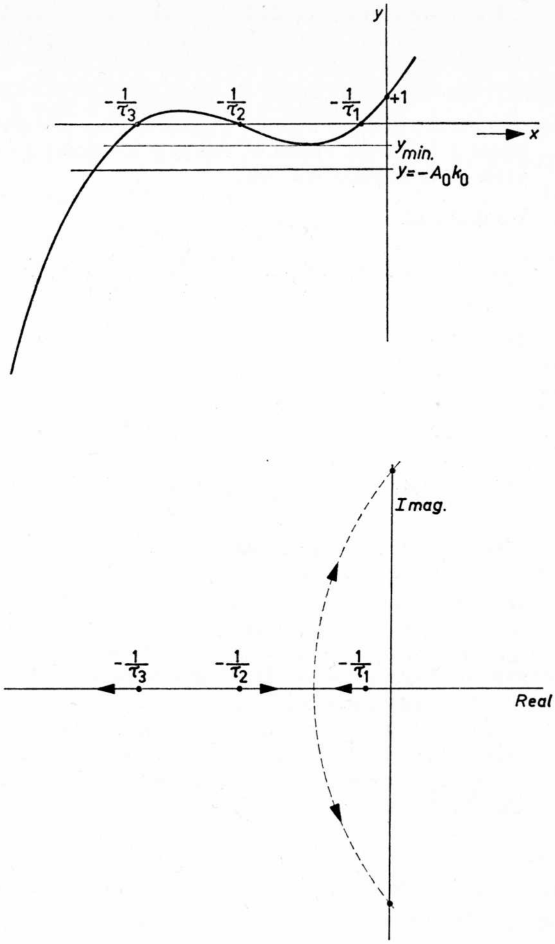

Fig. 22-2

signal to grow exponentially until the amplifier "blocks". For A_0k_0 negative and $|A_0k_0|<1$, all three roots will have a negative real part, so that the natural modes will then be damped. This will also be the case for small positive values of A_0k_0. When the value of A_0k_0 is increased, the left-hand side point of intersection shifts to the left, i.e. it becomes more negative, so that the sum of the other two roots becomes less negative. This remains valid even when these roots become complex conjugate, which is the case for $A_0k_0 > -y_{\min}$, and beyond a given value of A_0k_0 the real part of these roots will become positive and oscillation will occur.

The lower part of Fig. 22-2 shows the path of the roots when A_0k_0 increases. It is easy to calculate the value of A_0k_0 at which the roots pass through the imaginary axis. For we can then write:

$$\tau_1\tau_2\tau_3(p + ja)(p - ja)(p + b) = 0$$

where $\pm ja$ are the values of the pure imaginary roots and b is the value of the negative real root.

We thus find:

$$(p^2 + a^2)(p + b) = \left(p + \frac{1}{\tau_1}\right)\left(p + \frac{1}{\tau_2}\right)\left(p + \frac{1}{\tau_3}\right) + \frac{A_0k_0}{\tau_1\tau_2\tau_3}$$

Therefore:

$$b = \frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}, \quad a^2 = \frac{1}{\tau_1\tau_2} + \frac{1}{\tau_1\tau_3} + \frac{1}{\tau_2\tau_3},$$

and $$A_0k_0 + 1 = (\tau_1 + \tau_2 + \tau_3)\left(\frac{1}{\tau_1} + \frac{1}{\tau_2} + \frac{1}{\tau_3}\right).$$

This shows that the maximum value of A_0k_0 which is permissible without giving rise to oscillation, can be increased either by increasing the largest time constant or decreasing the smallest time constant. The latter course is normally not possible and one of the best known remedies against oscillation is increasing the largest time constant. However, the resultant disadvantage is that the amplification will already commence to fall off at low frequencies.

In control techniques often use is made of graphic representations such as the one shown in the lower part of Fig. 22-2, which indicate the relation between amplification A_0k_0 and the values of the roots. This is called the "root locus"-method.

We shall discuss oscillation and stability of amplifiers in greater detail in Sections 37 and 38. In the rest of this section we shall assume that no oscillation occurs.

We have three possibilities with a non-oscillating system. Firstly, $|A(\omega)k(\omega)|$ is large with respect to 1 for the frequency considered. Basic equation (22.1), for which we can also write

$$\frac{1}{A_{fb}(\omega)} = k(\omega) + \frac{1}{A(\omega)}$$

can then be simplified to:

$$A_{fb}(\omega) \approx \frac{1}{k(\omega)}$$

This shows one of the advantages of feedback: the amplification is no longer determined by A_0, which may strongly depend on prevailing circumstances, but by $k(\omega)$ which should be essentially constant if suitably chosen. Therefore we usually take a simple passive circuit, such as a voltage divider consisting of two resistances. In that case the new amplification will be both theoretically independent of frequency and also real, so that there will be no phase shift between output and input signals. However, these advantages are achieved at the cost of amplification, because when $|A(\omega)k(\omega)| \gg 1$, we have $|A_{fb}(\omega)| \ll |A(\omega)|$.

In the second case too, when $|A(\omega)k(\omega)|$ is small with respect to 1, it is easy to indicate the effect of feedback. It will be zero: $A_{fb}(\omega) \approx A(\omega)$, which is hardly an interesting result.

The situation becomes more complicated if the value $|A(\omega)k(\omega)|$ is nearly unity at the frequency considered. The amplifier's behaviour will then be determined by the exact value of $A(\omega)k(\omega)$, i.e. by both modulus and argument of this product. In all cases where the feedback is used to obtain favourable properties which are valid for $|Ak| \gg 1$ we shall, because of the decrease in amplification at high frequencies, always meet the case $|Ak| \approx 1$ too, with its associated problems. The applicability of a certain feedback method will be strongly determined by the behaviour in this region and we shall consider the latter problem in the remainder of this section. In the following sections we shall discuss some of the favourable effects valid for $|Ak| \gg 1$.

To illustrate what can be expected in the transition region $|Ak| \approx 1$, we shall now discuss in detail the feedback of a two-stage amplifier, each stage having an amplification of the form $A/(1+j\omega\tau)$. The reason for choosing this type of amplifier is that the calculations remain relatively simple, whilst at the same time the phenomena are representative of what can occur in this region.

The relation for the overall amplification of the two stages is:

$$A(\omega) = \frac{A_0}{(1+j\omega\tau_1)(1+j\omega\tau_2)}$$

where $A_0 > 0$ and τ_1 and τ_2 are assumed to be of the same order of magnitude.

Let us first consider the case when the feedback voltage has been obtained by means of a frequency-independent voltage divider, i.e. $k(\omega) = \text{constant} = k_0$, when $0 < k_0 < 1$ (Fig. 22-3). We also assume $A_0 k_0 \gg 1$.

The amplification of the amplifier with feedback is

$$A_{fb}(\omega) = \frac{A_0}{(1+j\omega\tau_1)(1+j\omega\tau_2) + A_0 k_0} \tag{22.3}$$

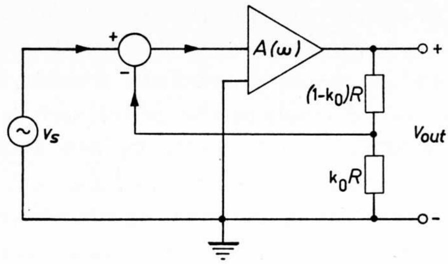

Fig. 22-3

As can be easily checked, this amplifier cannot oscillate because the roots of the equation

$$(1 + p\tau_1)(1 + p\tau_2) + A_0k_0 = 0$$

have a negative real part for each value $A_0k_0 > 0$. When the amplitude characteristic is measured, it will be seen that a peak occurs of which the height increases with $\sqrt{A_0k_0}$ (Fig. 22-4).

The reasoning that the peak in amplification is due to too high a gain for higher frequences seems to be promising since the insertion of additional capacitance across one of the anode resistance will cause the peak to decrease.

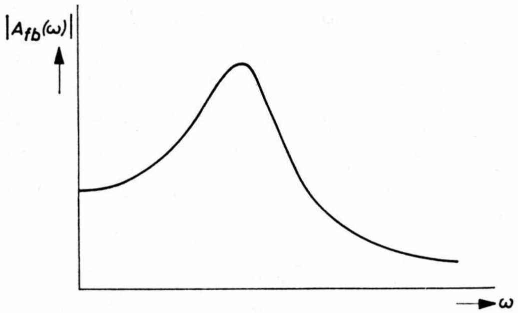

Fig. 22-4

The incorrectness of this reasoning is however proved by continuing the process and inserting additional capacitance across the other anode resistance as well. The peak will then not decrease any further, but will in fact increase. This can be explained by the following calculation:

Equation (22.3) gives:

$$|A_{fb}(\omega)|^2 =$$

$$= \frac{A_0^2}{(1+A_0k_0)^2 + \{(\tau_1+\tau_2)^2 - 2(A_0k_0+1)\tau_1\tau_2\}\,\omega^2 + \tau_1^2\tau_2^2\omega^4} \tag{22.4}$$

In the case of $2(A_0k_0+1)\tau_1\tau_2 > (\tau_1+\tau_2)^2$ or $A_0k_0 > \frac{1}{2}(\tau_1/\tau_2+\tau_2/\tau_1)$, the second term of the diviser is negative and the right-hand side has a maximum $|A_{fb}|_{max}$:

$$|A_{fb}|_{max} = A_{fb}(0) \cdot \frac{1}{\sqrt{1 - \left\{1 - \dfrac{(\tau_1+\tau_2)^2}{2\tau_1\tau_2} \cdot \dfrac{1}{A_0k_0+1}\right\}^2}}$$

with

$$A_{fb}(0) = \frac{1}{k_0 + A_0^{-1}}$$

This can be approximated for values of A_0k_0 which are large with respect to $\frac{1}{2}(\tau_1/\tau_2+\tau_2/\tau_1)$ by:

$$|A_{fb}|_{max} = A_{fb}(0) \cdot \sqrt{\frac{\tau_1\tau_2}{(\tau_1+\tau_2)^2}} \cdot \sqrt{A_0k_0}$$

The factor $\tau_1\tau_2/(\tau_1+\tau_2)^2$ can be written as $1/(\tau_1/\tau_2+2+\tau_2/\tau_1)$ and has a maximum value of $\frac{1}{4}$ for $\tau_1=\tau_2$. If we now increase the ratio between the time constants, for example by increasing the largest time constant, the value of the peak will decrease. If the other time constant is subsequently increased, the height of the peak will once more increase.

The amplitude characteristic of Fig. 22-4 is often undesirable: a flat response is preferred. This can be achieved either by decreasing A_0k_0 or by varying the time constants so that the coefficient of ω^2 in (22.4) becomes zero:

$$A_0k_0 = \frac{1}{2}\left(\frac{\tau_1}{\tau_2} + \frac{\tau_2}{\tau_1}\right) \tag{22.5}$$

With this last method it is necessary only to increase the largest time constant by adding extra capacitance across the anode resistance, so that (22.5) is satisfied. The amplitude characteristic then becomes, to a good approximation:

$$\left|A_{fb}(\omega)\right| = \left|A_{fb}(0)\right| \cdot \frac{1}{\sqrt{1 + 4\tau^4\omega^4}} \tag{22.6}$$

where τ = the smaller of the two time constants.

This amplitude characteristic is shown in Fig. 22-5, where the broken line represents an amplifier having two equal time constants τ and no feedback.

In order to obtain such a flat amplitude characteristic where the amplification in the flat region is independent of amplification A, it is necessary that A_0k_0, the value of the "loop gain" at low frequencies, is large with respect to unity. It follows from (22.5) that the ratio between the two time constants must then be large too, namely approx. $2A_0k_0$, which gives us for the largest time constant the value $2A_0k_0\tau$. Nevertheless, the amplitude characteristic of Fig. 22-5 is determined exclusively by the smallest time constant, and has a bandwidth which is even better than that of a amplifier with two

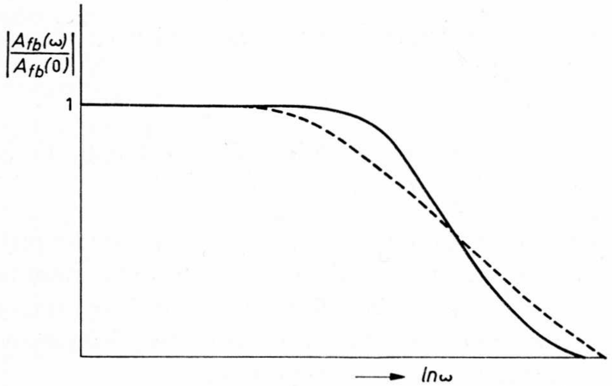

Fig. 22-5

of the smaller time constants and no feedback. We can use this effect to give a flat amplitude characteristic up to high frequencies to a system which inherently comprises one large time constant.

An example of such a case is found in the mass spectrometer. We have to measure here currents of the order of $10^{-12}-10^{-15}$ A which, as we shall see on p. 240 are preferably transposed into voltages by means of a large resistor (e.g. $10^{11}\ \Omega$). This is shown in Fig. 22-6. If the total capacitance of the collector plate of the mass spectrometer and the input of the amplifier is, say, 20 pF, the time constant of this input will be $10^{11} \cdot 20 \cdot 10^{-12} = 2$ seconds, which means that only very slow

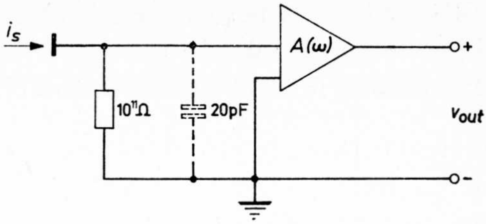

Fig. 22-6

changes in the measuring current can be amplified and measured if no special precautions are taken. By using feedback, the bandwidth of the input plus amplifier can be made to correspond to the smallest time constant of the amplifier. For example, if the latter is 50 μ s, it is possible to amplify signals up to frequencies of approx. 1 kc/s with almost constant amplification. In that case, the necessary loop gain is $\frac{1}{2} \cdot \frac{\tau_1}{\tau_2} = \frac{1}{2} \cdot \frac{2}{50 \cdot 10^{-6}} = 20{,}000$ times. When using the full output signal as feedback voltage ($k_0 = 1$), A_0 must be 20,000 (Fig. 22-7).

In reality we need an amplifier with more stages in order to obtain the required amplification and thus with a greater number of small

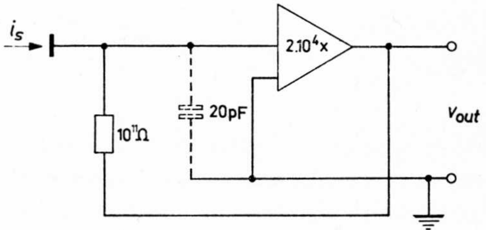

Fig. 22-7

time constants. In that case the critical loop amplification is, of course, smaller than with only one small time constant; the gain in speed is therefore also smaller than would follow from the ratio of large to small time constants. The example takes this into consideration by assuming that the amplifier possesses a single time constant of 50 μ s, which is a rather large value for an amplifier stage. Under normal circumstances, the time constant of an amplifier which gives an amplification of approx. 10,000 times can be easily restricted to approx. 10 μ s, so that the mass spectrometer can cope with signals of several tens of kc/s.

If we start from an amplifier with time constants originally of the same order of magnitude and a flat amplitude characteristic which has been obtained by making one of them $2A_0k_0$ times as large, Fig. 22-5 shows that the improvement does not always justify the loss in amplification. In this respect the loss in amplification would have been more profitable if it had been obtained by decreasing the anode resistances and thus the time constants. In doing this we might also expect that by our procedure the amplification would become less dependent on A_0. Whether this is true can be checked by calculating from (22.4) the dependence of $A_{fb}(\omega)$ on changes in A_0:

$$\frac{\mathrm{d}|A_{fb}(\omega)|}{|A_{fb}(\omega)|} =$$

$$= \frac{\mathrm{d}A_0}{A_0} \cdot \frac{1 + A_0k_0 \quad + \{(\tau_1 + \tau_2)^2 - \quad (A_0k_0 + 2)\tau_1\tau_2\}\omega^2 + \tau_1{}^2\tau_2{}^2\omega^4}{(1 + A_0k_0)^2 + \{(\tau_1 + \tau_2)^2 - 2(A_0k_0 + 1)\tau_1\tau_2\}\omega^2 + \tau_1{}^2\tau_2{}^2\omega^4}$$

In the special case of maximum flat response (equation 22.5) this reduces to

$$\frac{\mathrm{d}|A_{fb}(\omega)|}{|A_{fb}(\omega)|} = \frac{\mathrm{d}A_0}{A_0} \cdot \frac{1 + A_0k_0 + A_0k_0\tau_1\tau_2\omega^2 + \tau_1{}^2\tau_2{}^2\omega^4}{(1 + A_0k_0)^2 + \tau_1{}^2\tau_2{}^2\omega^4}$$

With $A_0k_0 \gg 1$ and $\tau_1\tau_2 = 2A_0k_0\tau^2$, where τ is again the smaller of the two time constants, this approximates to

$$\frac{\mathrm{d}|A_{fb}(\omega)|}{|A_{fb}(\omega)|} = \frac{\mathrm{d}A_0}{A_0} \cdot \frac{\dfrac{1}{A_0k_0} + 2\tau^2\omega^2 + 4\tau^4\omega^4}{1 + 4\tau^4\omega^4}$$

For small values of ω, the relative variation in amplification is thus reduced A_0k_0 times, but this factor will already decrease at low frequencies. It follows from (22.6) that although the amplification is almost flat up to frequencies $\omega \approx 1/2\tau$, the reduction has already become a factor 2 smaller at the frequency $\omega = 1/\tau\sqrt{2A_0k_0}$. This proves that the stabilization of the amplification is only effective in a small part of the frequency band in which the amplitude is flat.

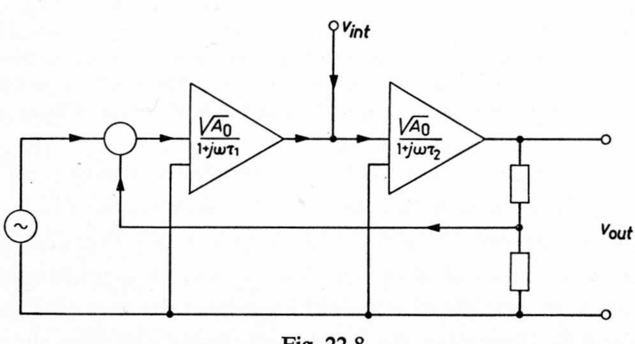

Fig. 22-8

Fig. 22-8 shows a second disadvantage of increasing one of the time constants. In a two-stage amplifier with equal stages, the flat characteristic can be theoretically achieved by increasing the time constant of either the first or the second stage. If we choose the first possibility, the first stage will only slightly amplify the high frequencies of the ultimately transmitted band, namely $\sqrt{A_0/(1+\omega^2\tau^2)}$ times. Thus an interference signal v_{int} containing these high frequencies and appearing at the input of the second stage will give a correspondingly larger contribution to the output voltage.

On the other hand, increasing time constant τ_2 of the second stage will render the amplification of this stage small for the high frequencies in the transmitted band. If it is desirable to achieve a large output signal for these frequencies, the input signal of the second stage must consequently be large. A large part of the I_a–V_{gk} characteristic is then used, which may cause unwanted phenomena such as non-linear distortion (see Section 23).

It is obvious that we go from one compromise to another with this configuration. Despite of this, however, this method of feedback is simple and in many cases it leads to useful results.

Let us now assume, with reference to equation (22.1), that we had followed a different line of reasoning and had attributed the occurrence of the peak at high frequencies not to an exaggerated amplification in this region, but to the fact that the divisor of (22.1), i.e. $|A(\omega)k(\omega)|$ is too small.

There is not very much we can do about $A(\omega)$, but $k(\omega)$ can be increased at high frequencies by inserting a capacitor across the upper resistor R_1 of the voltage divider (Fig. 22-9).

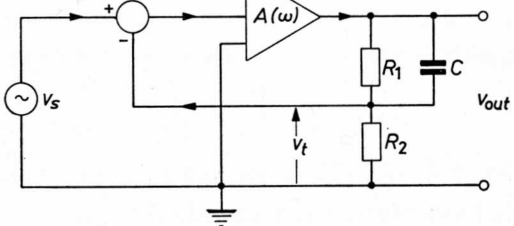

Fig. 22-9

In order to simplify the calculation we assume that $v_t \ll v_{out}$, so that the current passing through the voltage divider is determined by the impedance of the upper branch:

$$k(\omega)=k_0(1+j\omega\tau_0)$$

where $\tau_0=R_1C$.

We now find for the amplification of the feedback amplifier:

$$A_{fb}(\omega) = \frac{A_0}{(1 + j\omega\tau_1)(1 + j\omega\tau_2) + A_0k_0(1 + j\omega\tau_0)}$$

and for the absolute value:

$$|A_{fb}(\omega)|^2 = \frac{A_0^2}{(1 + A_0k_0)^2 + \{(\tau_1 + \tau_2 + A_0k_0\tau_0)^2 - 2(A_0k_0 + 1)\tau_1\tau_2\}\omega^2 + \tau_1^2\tau_2^2\omega^4} \tag{22.7}$$

The coefficient of ω^2 in the divisor must be zero for a characteristic of optimum flatness:

$$(\tau_1 + \tau_2 + A_0k_0\tau_0)^2 = 2(A_0k_0 + 1)\tau_1\tau_2$$

Thus:

$$\tau_0 = \frac{\sqrt{2(A_0k_0 + 1)\tau_1\tau_2} - \tau_1 - \tau_2}{A_0k_0} \tag{22.8}$$

This can be approximated for $\tau_1 \approx \tau_2$ and $A_0k_0 \gg 1$ to:

$$\tau_0 = \sqrt{\frac{2\tau_1\tau_2}{A_0k_0}}.$$

The time constant τ_0 will therefore usually be very small with respect to τ_1 and τ_2.

We find for the amplitude characteristic:

$$|A_{fb}(\omega)| = \frac{A_0}{A_0k_0 + 1} \cdot \frac{1}{\sqrt{1 + \left(\frac{\tau_1\tau_2}{1 + A_0k_0}\right)^2 \omega^4}}$$

When comparing this with (22.6), we find that the frequency range has become $\sqrt{2A_0k_0}$ times wider, a very considerable gain.

To find in this case the effect of variations in A_0 on amplification A_{fb} we derive the relative variations from equation (22.7) by differentiation:

$$\frac{\mathrm{d}|A_{fb}(\omega)|}{|A_{fb}(\omega)|} \Big/ \frac{\mathrm{d}A_0}{A_0} = \frac{1 + A_0k_0 + \{(\tau_1 + \tau_2)(\tau_1 + \tau_2 + A_0k_0\tau_0) - (A_0k_0 + 2)\tau_1\tau_2\}\omega^2 + \tau_1^2\tau_2^2\omega^4}{(1 + A_0k_0)^2 + \{(\tau_1 + \tau_2 + A_0k_0\tau_0)^2 - 2(A_0k_0 + 1)\tau_1\tau_2\}\omega^2 + \tau_1^2\tau_2^2\omega^4}$$

and this becomes in the case of maximum flat response (22.8):

$$\frac{\mathrm{d}|A_{fb}(\omega)|}{|A_{fb}(\omega)|} \Bigg/ \frac{\mathrm{d}A_0}{A_0} =$$

$$\frac{1 + A_0k_0 + \{(\tau_1 + \tau_2)\sqrt{2(A_0k_0 + 1)\tau_1\tau_2} - (A_0k_0 + 2)\tau_1\tau_2\}\omega^2 + \tau_1{}^2\tau_2{}^2\omega^4}{(1 + A_0k_0)^2 + \tau_1{}^2\tau_2{}^2\omega^4}$$

Once again we find a reduction factor of $1 + A_0k_0$ at very low frequencies, but because of the negative term in the dividend this factor decreases much more slowly and even increases initially. This makes it possible, with correct choice of parameters, to obtain useful frequency bandwidths which exceed those of an amplifier without feedback. Moreover, the other disadvantages of increasing the time constants do not occur here, so that the change to a frequency-dependent feedback loop offers many advantages.

The reader may wonder why we derived the solution with an increased time constant as we have seen that this has poorer results than that with frequency-dependent feedback. Apart from the fact that the first method can be most useful in many cases, we would also have ignored one of the most fascinating aspects of electronics. For although it would appear quite reasonable to assume that frequency-independent feedback offers the best guarantee for a frequency-independent amplitude characteristic, the contrary comes out to be true in practice. Furthermore, although we can rely on calculations to tell us all we wish to know about a given configuration, there are usually no rules for arriving at the configuration which would lead to the best results. One consequence of this is that it is still possible to obtain surprising improvements even in apparently exhausted fields.

The amplitude and phase characteristics generally give a reliable indication of the properties of an amplifier for signals consisting of one or more sinusoidal components. Apart from this type of signal, as already mentioned in Section 7, we often deal with transient signals, of which the step function, pulse function and the square wave are much used waveforms. (Fig. 22-10).

It is thus important to know in these cases how the amplifier behaves for

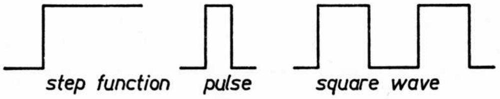

Fig. 22-10

such a waveform. Theoretically any signal can be expressed as the sum of various sinusoidal components derived by means of Fourier analysis, (Section 36). Each of these components can then be treated separately, with due regard to the frequency characteristic of the amplifier, and then added together at the output. Clearly therefore the frequency characteristic also determines the result for arbitrary signals. Needless to say, such a method is often very tedious to apply in practice.

If we are dealing with the waveforms of Fig. 22-10, the solution will be obtained more quickly by starting from the differential equation valid for the system. Particularly when the input signal is a step function, it is easy to find a specific solution of this equation for $t>0$. The additional conditions can then be taken care of by means of the natural modes. In the case of a step function these conditions are determined by the values of the currents and voltages at the beginning of the step. The switching-on phenomenon is then completely accounted for in the amplitude of the natural modes. Since both the pulse and the square wave may be considered as sum signals of positive and negative step functions that have different origins in time, the solutions for these types of waveforms are easily derived from those for a step function.

If the system is at rest for $t<0$, the coefficients of the natural modes can be determined directly by means of Heaviside's formulae.

Let us suppose that the relation between input and output signals is given by the equation:

$$\{B(p)\}v_o = \{p\,A(p)\}v_i$$

where $A(p)$ and $B(p)$ only contain positive powers of the operator $p=\mathrm{d}/\mathrm{d}t$. With a step function of unit height as input signal the output signal for $t>0$ is then given by

$$v_o = \sum_1^n i\,\frac{A(p_i)}{B'(p_i)}\cdot e^{p_i t}$$

where p_i is a single root of the equation

$$B(p)=0$$

and $B'(p_i)$ is the value of $\mathrm{d}B(p)/\mathrm{d}p$ for $p=p_i$. The solution is slightly more complicated for multiple roots of $B(p)$. If p_i is an n-fold root, this contributes to the solution:

$$\sum_{j=0}^{n-1}\frac{t^j e^{p_i t}}{(n-j-1)!\,j!}\cdot\left[\left(\frac{\mathrm{d}}{\mathrm{d}p}\right)^{n-j-1}\left\{\frac{A(p)\cdot(p-p_i)^n}{B(p)}\right\}\right]_{p=p_i}$$

Examples:

We find for Fig. 22-11:

$$v_o = \frac{RCp}{RCp + 1} v_i$$

therefore: $A(p) = RC$; $B(p) = RCp + 1$.

Since $p = -1/RC$ is a root of $B(p) = 0$ and $B'(p) = RC$, we find that for $t > 0$

$$v_o = e^{-t/RC}$$

For fig. 22-12 we find:

$v_o = \dfrac{1}{RCp + 1} v_i$ or $p(RCp + 1)v_o = pv_i$, so that $A(p) = 1$, $B(p) = p(RCp + 1)$ and $B'(p) = 2RCp + 1$.

The roots of $B(p) = 0$ are $p_1 = 0$ and $p_2 = -1/RC$, hence $B'(p_1) = 1$ and $B'(p_2) = -1$.

For a unit step input we thus find for $t \geqslant 0$:

$$v_o = 1 - e^{-t/RC}$$

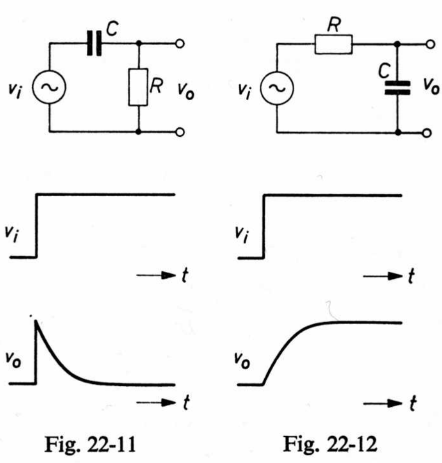

Fig. 22-11 Fig. 22-12

A square wave can be considered as the superpositioning of alternating positive and negative step functions, each spaced by an interval time T. Depending on the ratio between T and RC we obtain for the circuits shown in Figs. 22-11 and 22-12 with a square wave as input voltage signal, the output signals given in Fig.22-13.

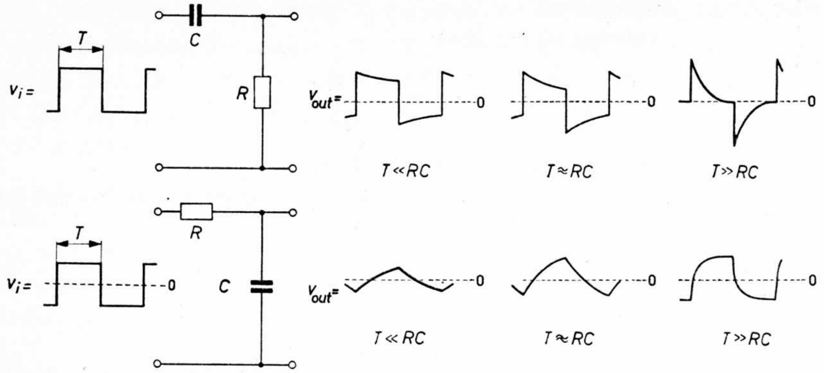

Fig. 22-13

In common with the frequency characteristic, the "step function response", i.e. the curve of the output signal plotted against time with a unit step as input signal, completely determines an amplifier's properties. Any arbitrary input signal can also be considered as the superposition of a large number of infinitely small steps (Fig. 22-14). The step function response is usually

Fig. 22-14

obtained by connecting a square-wave signal generator to the input and observing the output signal on an oscilloscope. The repetition frequency of the square wave should be chosen so that the response has reached its final value well within half a period.

Once the flattest possible amplitude characteristic has been obtained for an amplifier with two time constants, either by correctly adjusting the feedback loop gain or the ratio of the time constants, or by applying frequency-dependent feedback, an output signal as shown in Fig. 22-15 is obtained when a step function is applied to the input. The effect shown is termed "overshoot". This phenomenon is undesirable for a number of applications and needs to be eliminated if the amplifier is to be used for such purposes. Suitable measures can be taken when the cause is known and in this case this is not difficult to find; it is the damped natural modes

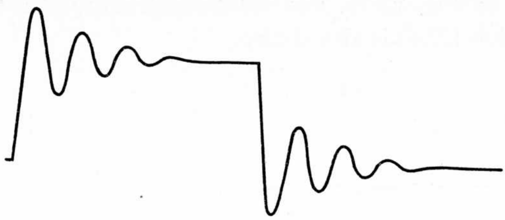

Fig. 22-15

of the system which are "pulsed" by the step. Apparently in a system with a flat frequency characteristic the natural modes are still damped sinusoids and the roots of the corresponding equation complex. The remedy therefore lies in making these roots negative real.

In the case of an amplifier with frequency-independent feedback the natural frequencies can be determined by putting the denominator of (22.3) equal to zero:

$$(1 + p\tau_1)(1 + p\tau_2) + A_0k_0 = 0$$

The roots are:

$$-\frac{\tau_1 + \tau_2}{2\tau_1\tau_2} \pm \sqrt{\frac{(\tau_1 + \tau_2)^2 - 4(1 + A_0k_0)\tau_1\tau_2}{4\tau_1{}^2\tau_2{}^2}}$$

which becomes, by applying equation (22.5) for the flat characteristic:

$$-\frac{\tau_1 + \tau_2}{2\tau_1\tau_2}[1 \pm j]$$

This proves that damped sinusoidal vibrations occur.

However this is no longer true if

$$(\tau_1 + \tau_2)^2 \geqslant 4(1 + A_0k_0)\tau_1\tau_2 \quad \text{or} \quad 4A_0k_0 + 2 \leqslant \frac{\tau_1}{\tau_2} + \frac{\tau_2}{\tau_1}$$

In the limiting case where the equality sign is valid, the largest time constant should be made about twice as large as was necessary for a flat amplitude characteristic. We find in that case to a good approximation:

$$|A_{fb}(\omega)| = |A_{fb}(0)| \frac{1}{1 + 4\tau^2\omega^2}$$

This is shown in Fig. 22-16. For comparison purposes, the flat characteristic of equation (22.6) is also shown.

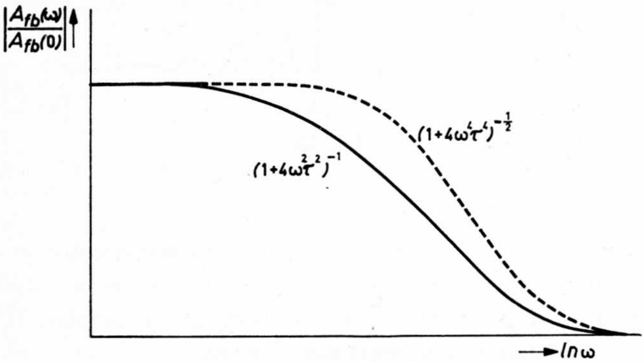

Fig. 22-16

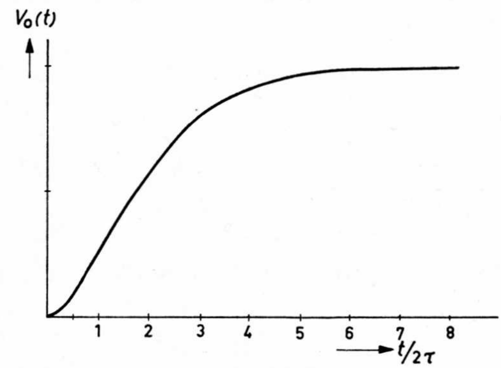

Fig. 22-17

For the step function response we obtain in this case

$$v_o(t) = \frac{A_0}{1 + A_0 k_0}\left\{1 - \left(\frac{t}{2\tau} + 1\right)e^{-\frac{t}{2\tau}}\right\}$$

and overshoot no longer occurs (Fig. 22-17).

The process occurs completely similarly for the frequency-dependent feedback. With the amplifier which just does not show overshoot, we find for the time constant τ_0:

$$\tau_0 = \frac{2\sqrt{(A_0 k_0 + 1)\tau_1 \tau_2} - (\tau_1 + \tau_2)}{A_0 k_0},$$

for the amplitude characteristic:

$$\left|A_{fb}(\omega)\right| = \left|A_{fb}(0)\right| \frac{1}{1 + \dfrac{\tau_1\tau_2}{1 + A_0k_0}\,\omega^2}$$

and for the step function response

$$v_o(t) = \frac{A_0}{1 + A_0k_0}\left[1 - \left(\sqrt{\frac{A_0k_0 + 1}{\tau_1\tau_2}}\,t + 1\right)e^{-\sqrt{\frac{A_0k_0+1}{\tau_1\tau_2}}\,t}\right]$$

With $\tau_1 = \tau_2 = \tau$ thus an increase in bandwidth and speed of $2\sqrt{A_0k_0+1}$ times compared to the frequency-independent feedback, results.

It is not surprising that with a feedback amplifier showing a peak in its amplitude characteristic, a step function as input signal gives an output signal containing oscillations, because there are frequencies for which the amplification is greater than at low frequencies. However, this is not the case with a flat characteristic and at first sight there does not seem to be any reason for the overshoot. One therefore often finds the statement that this is due to certain shortcomings of the phase characteristic, such as deviations from a linear relation between phase and frequency (see section 36). This is incorrect for two reasons; firstly, as we shall see later, the amplitude and phase characteristics of a transfer system are inter-dependent and it is obviously not possible to ascribe a certain disadvantage to one of two dependents; secondly, this statement is not even valid in the case (which can only be realized mathematically, not physically) where amplitude and phase characteristics can be chosen independently of each other. Taking the ideal case for the latter that the phase shift is zero at all frequencies, we find that even in this case the phenomenon of overshoot can occur.

According to Fourier analysis (see Section 36), a square wave with a period $T = 2\pi/\omega_0$ and an amplitude $= 1$ can be considered as the sum of sinusoidal voltages of frequencies ω_0, $3\omega_0$, $5\omega_0$, etc.:

$$v_i = \frac{4}{\pi}\left(\sin\omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \frac{1}{5}\sin 5\omega_0 t + \ldots\right)$$

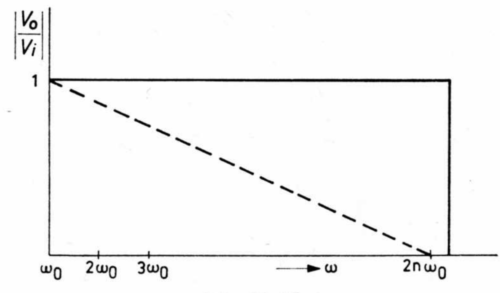

Fig. 22-18

When this voltage is fed through a system having an amplitude characteristic equal to 1 for $\omega < 2n\omega_0$, and equal to zero for $\omega > 2n\omega_0$ (the continuous line in Fig. 22-18), and also possessing the above mentioned ideal phase characteristic, the output voltage will be:

$$v_o = \frac{4}{\pi} \sum_{1}^{n} k \frac{1}{2k-1} \cdot \sin (2k-1)\omega_0 t$$

If this summation is carried out, it appears that overshoot occurs at the zero crossing points. When n is increased, the width of the overshoot will contract towards zero but not the height (Fig. 22-19). This phenomenon is known as Gibbs's phenomenon. On the other hand,

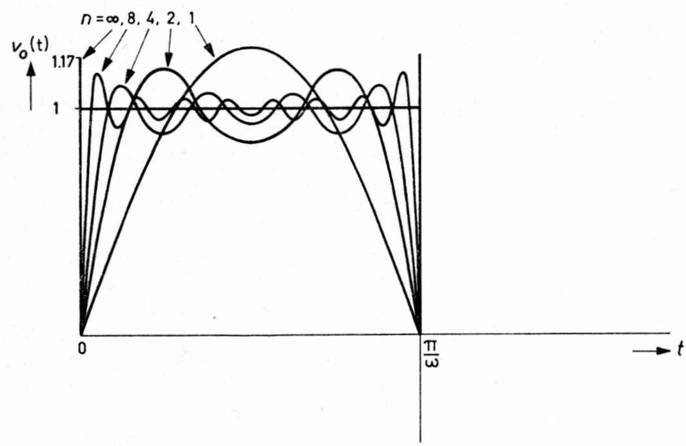

Fig. 22-19

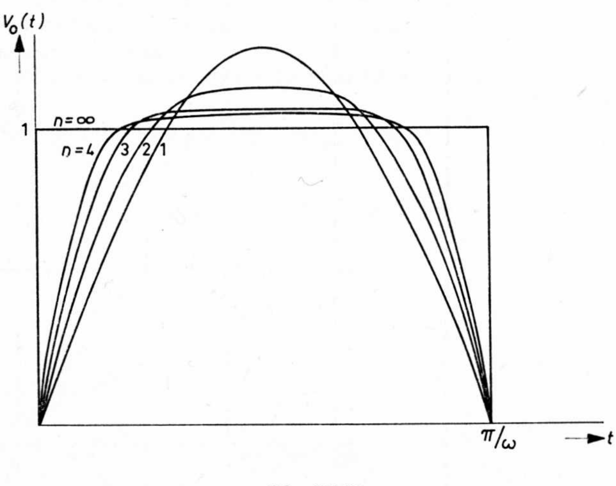

Fig 22-20

assuming a system whose phase characteristic once again equals zero everywhere, but where the amplitude characteristic equals 1 for $\omega = \omega_0$, and drops linearly to zero at $\omega = 2n\omega_0$ (the broken line in Fig. 22-18). The output voltage in this case will be:

$$v_o = \frac{4}{\pi} \sum_1^n k \frac{2n - 2k + 1}{2n - 1} \cdot \frac{1}{2k - 1} \cdot \sin(2k - 1)\omega_0 t.$$

This sum is one of those occurring in a Cesaro summation. Thus Gibbs's phenomenon does not occur with the Cesaro summation of Fourier series. The partial sums do not show any overshoot, Fig. 22-20. Mathematically speaking, the Fourier sum is not uniformly convergent but the Cesaro sum is. Whether overshoot will occur in this case is obviously connected with the slope of the fall in the amplitude characteristic.

Returning to the behaviour of a feedback system with sinusoidal signals, we should note that, in order to assess the amplifier's quality, it is essential to know the phase characteristic, as well as the amplitude characteristic. Theoretically, the various calculations used for, $A_{fb}(\omega)$, should be repeated for the argument of $A_{fb}(\omega)$. It is, however, less usual to pay particular attention to the phase characteristic. This results from the fact that with most transfer systems, the necessary data on the curve of the phase characteristic can be derived from the behaviour of the amplitude characteristic. Bode has derived a relationship between the phase at a certain frequency and the frequency-dependence of the amplification for those systems which have a transfer function of the form of equation (6.5).

$$v_o = \left\{ \frac{A(p)}{B(p)} \right\} \cdot v_i$$

in which the roots of $A(p)$ and $B(p)$ have a negative real part.

In all stable systems, the roots of $B(p)$ will in effect have a negative, real part, whilst roots of $A(p)$ with a positive real part only occur when bridges, or compensation circuits are used. Particularly in the case of linear amplifiers, there will be little need for such circuits, and the condition concerning the roots is therefore nearly always satisfied.

From the relation postulated by Bode it follows that in a frequency range $0 < \omega \leqslant \omega_h$, where the logarithmic amplification undergoes little change (Fig. 22-21), the phase will become practically proportional to the frequency $\varphi_c = -\tau\omega_c$, so that for signals which exclusively contain components in this frequency range, the output signal will be delayed by a time τ with respect to the input signal, but otherwise will have almost exactly the same form. In the region $\omega > \omega_h$, where the amplification changes greatly the phase will also be subject to large changes. However, for a linear amplifier this range is

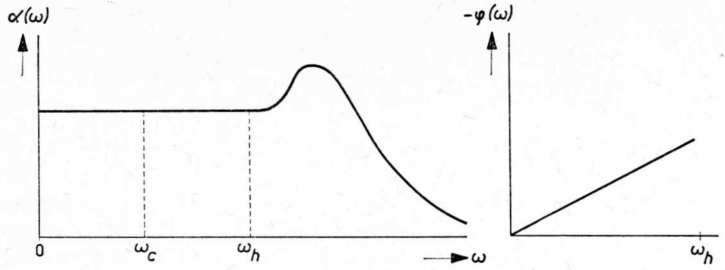

Fig. 22-21

naturally unsuitable because of the distortion caused by the difference in amplification.

As we have in fact tried in the foregoing to obtain a flat amplitude characteristic, it follows that the phase characteristic will also be of a smooth form.

The transfer of a signal with frequency ω is described by:

$$v_o = \frac{A(\mathrm{j}\omega)}{B(\mathrm{j}\omega)} \cdot v_i$$

If we write:

$$\alpha(\omega) = \ln \left| \frac{A(\mathrm{j}\omega)}{B(\mathrm{j}\omega)} \right| \quad \text{and} \quad \varphi(\omega) = \arg \frac{A(\mathrm{j}\omega)}{B(\mathrm{j}\omega)}$$

we find for the phase at frequency ω_c:

$$\varphi(\omega_c) = \frac{2\omega_c}{\pi} \int_0^\infty \frac{\alpha(\omega) - \alpha(\omega_c)}{\omega^2 - \omega_c^2} \mathrm{d}\omega$$

This relation can be derived by making use of the properties of a complex variable and is only valid if all roots of $A(p)$ and $B(p)$ have a negative, real part.

Assuming that the logarithmic amplification $\alpha(\omega)$ is constant up to the frequency $\omega_h \gg \omega_c$, and then for frequencies above ω_h possibly at first increases but decreases at still higher frequencies, the term $\alpha(\omega) - \alpha(\omega_c)$ will be zero, for $0 < \omega < \omega_h$, and therefore also the contribution to the integral. For $\omega > \omega_h$, the denominator can be simplified to ω^2 and the integral

$$\int_{\omega_h}^\infty \frac{\alpha(\omega) - \alpha(\omega_c)}{\omega^2} \mathrm{d}\omega$$

will then have the same value P for all frequencies $\omega_c < \omega_h$. We have therefore

$$\varphi(\omega_c) = \frac{2P}{\pi} \omega_c$$

which corresponds to our previous assumption.

An interesting variant of the formula for $\varphi(\omega_c)$ is obtained by the substitution of $u = \ln(\omega/\omega_c)$. This gives:

$$\varphi(\omega_c) = \frac{2}{\pi} \int_{-\infty}^{\infty} \frac{\alpha(\omega) - \alpha(\omega_c)}{e^u - e^{-u}} \, du$$

or, after partial integration

$$\varphi(\omega_c) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{d\alpha}{du} \cdot \ln \coth \frac{|u|}{2} \, du$$

This formula establishes a relation between the phase at frequency ω_c and the derivative of the logarithmic amplification. The weighting factor $\ln \coth |u|/2$ (see Fig. 22-22) is called Bode's weighting function, and an amplitude characteristic where the logarithmic amplification α is plotted against $\ln \omega$ is called a Bode diagram.

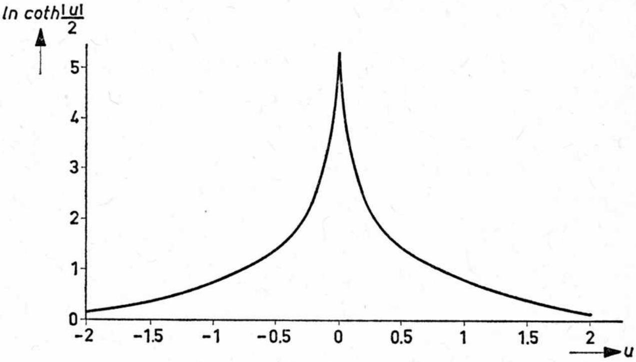

Fig. 22-22

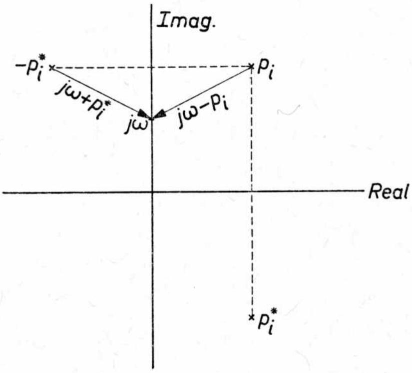

Fig. 22-23

If we have a transfer function where the numerator possesses a root $p_i = \alpha_i + j\beta_i$ with positive real part, a function can be obtained by multiplying by $(p + p_i^*)/(p - p_i)$, Fig. 22-23, of which all roots will have a negative, real part and for which the above-mentioned relation between amplitude and phase applies. The modulus of the multiplication factor equals 1 for purely imaginary values of p, so that the amplitude characteristics of both functions will be equal. The phase of the original function equals that of the new one, minus the argument of the multiplication factor.

Let us conclude our considerations on the behaviour of feedback amplifiers in the frequency range for which $|A(\omega)k(\omega)| \approx 1$ is valid, by saying a few words about systems with more than two time constants. The characteristics can be calculated in the same manner as for an amplifier with two time constants. However, with an increasing number of time constants, the calculations rapidly become more complicated, apart from a few special cases. We can, however, indicate some rules for these systems in direct relation to those we derived for a two-stage amplifier, namely:

1. Oscillation can be avoided by making the greatest time constant sufficiently large.
2. Peaks in the amplitude characteristic will disappear when making the greatest time constant still greater by a certain amount.
3. Overshoot will disappear with a further increase in this time constant.

These rules are valid for time constants which cause the amplitude characteristic to drop at high frequencies. For time constants which cause a drop at lower frequencies (RC-coupling, cathode decoupling) the opposite usually applies:

1. Oscillation in this frequency range is eliminated by reducing the smallest of the relevant time constants.
2. Peaks in the amplitude characteristic in this range will disappear by further reduction of this time constant.
3. Overshoot does not normally occur in this case.

These solutions are usually not the best possible for a given system and often they are not even admissible for certain reasons. Fortunately, however, the systems in measurement electronics with feedback are generally rather simple, so that the theory of more complex systems as used in control techniques is seldom necessary.

When feedback is used, we often distinguish between positive and negative feedback. But, as we have seen above, feedback is usually complex, so that such a simple subdivision is not possible. These terms are, however, useful in practice to indicate the feedback method used.

When we speak of negative feedback, $|Ak+1|$ must be large with respect to 1 over the frequency range which is important; moreover, the argument of Ak must not lie in the neighbourhood of 180°. We then have a non-oscillating amplifier which possesses the favourable properties which exists for $|Ak| \gg 1$. In this sense, most feedback systems can be said to have negative feedback.

In the case of positive feedback, $|Ak+1| < 1$ applies and such an amplifier will possess properties which are valid for $|Ak+1| \approx 0$. This means that any differences in behaviour at various frequencies will be increased by feedback. We shall give an example of positive feedback when discussing active filters in Section 33.

We should also note that, although the various rules we have given above were derived for a system exclusively containing signal voltages, the same rules apply, with due alteration of details, to signal currents or combinations of signal voltages and currents.

In conclusion, we can state that feedback makes it possible to render the amplitude and also the phase characteristic of a transfer system less dependent on basic parameters. However, this is accomplished at the cost of amplification, and there is a danger that the amplitude and phase characteristics exhibit an undesirable behaviour in a certain frequency range, or that oscillations occur. When feedback is being used, one should therefore always check whether there is any danger of these effects. We shall study this matter in more detail in Section 38 after discussing oscillation.

In the following sections we shall discuss some other aspects of that form of feedback, which, according to the definitions given above, could be called negative feedback.

23. Distortion

The highest demand one can make on a transfer system is that at any instant the output signal is equal to the input signal, multiplied by a constant. Because the system would have to operate without any inertia it is not possible to realize this ideal transfer (trivialities, such as voltage dividers consisting of resistors, excepted). The best approximation to this ideal is one where the output signal is shifted in time with respect to the input signal, but otherwise represents a true reproduction of it. A first condition for this is the linearity of the system. Fourier analysis teaches us that the amplitude characteristic of such a linear system must then be independent of frequency, and that the phase must be proportional to frequency (except a few less important possibilities). This, too, cannot be realized in electronics, but the deviation does not have to be large if we restric ourselves a little in the generality of the signals. For example, we have learnt in the previous section, that in a frequency range where the amplitude characteristic is flat, the phase is proportional to the frequency in good approximation. This means that only small deviations will occur for signals which exclusively contain components in this frequency range. We shall discuss this subject in more detail in Section 36.

The distortion which occurs in linear systems, caused by shortcomings in the amplitude and phase characteristics, is called linear distortion. As we have seen in the foregoing this may be reduced in certain cases by incorporating negative feedback.

As soon as more stringent demands are made on the linearity of the relation between the amplitudes of the input and output signals of an amplifier, it is no longer possible to ignore the non-linear behaviour of various elements, particularly the active ones. This will generally cause a certain amount of distortion, which in this case is called non-linear distortion. In this section we shall discuss the effect of feedback on this non-linear distortion.

When non-linearities occur in the transfer, we can write in the simplest case the relation between input and output voltage as follows:

$$v_o = A_0(v_i + a\,v_i^2 + b\,v_i^3 + \ldots) \qquad (23.1)$$

The distortion which occurs in the output signal will not only depend on the magnitude, but also on the form of the input signal. If the input signal of an amplifier to which equation (23.1) applies is sinusoidal, the output signal will only contain harmonics. If the input signal consists of the sum of two sinusoidal signals of different frequencies, we also obtain sum and difference

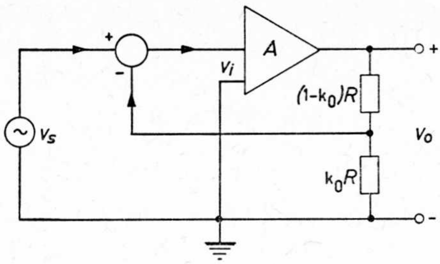

Fig. 23-1

frequencies which can be particularly objectionable, especially in sound reproduction.

In order to consider the effect of feedback on this non-linear distortion, let us look at Fig. 23-1, where the feedback k_0 is not dependent on frequency. In order to calculate this effect we first transform (23.1) to:

$$v_i = \frac{1}{A_0}(v_o + \alpha v_o^2 + \beta v_o^3 + \ldots) \tag{23.2}$$

In this transformation, which is always desirable for the calculation of distortion, the coefficients can be easily determined by substitution and equating of coefficients. For example, we find:

$$\alpha = \frac{a}{A_0} \quad \text{and} \quad \beta = \frac{2a^2 - b}{A_0^2}$$

Combining (23.2) with $v_i = v_s - k_0 v_o$ gives:

$$v_s = \frac{A_0 k_0 + 1}{A_0}\left(v_o + \frac{\alpha}{1 + A_0 k_0}\, v_o^2 + \frac{\beta}{1 + A_0 k_0}\, v_o^3 + \ldots\right) \tag{23.3}$$

When comparing (23.3) with (23.2) we observe that the coefficients of the higher-power terms, and therefore also the influence of these terms for a given output signal v_o, are multiplied by $(A_0 k_0 + 1)^{-1}$ due to feedback. This means for $A_0 k_0 > 0$ that the distortion becomes smaller and the deviation from a linear relation between v_o and v_s therefore decreases with increasing feedback.

Analogous to (23.1), but this time for an amplifier with negative feedback, equation (23.3) also yields:

$$v_o = A_0 \left[\frac{v_s}{1 + A_0 k_0} + \frac{a}{1 + A_0 k_0} \left(\frac{v_s}{1 + A_0 k_0} \right)^2 + \right.$$

$$\left. + \left(\frac{b}{1 + A_0 k_0} - 2a^2 \frac{A_0 k_0}{(1 + A_0 k_0)^2} \right) \left(\frac{v_s}{1 + A_0 k_0} \right)^3 + \dots \right] \quad (23.4)$$

We first observe that the input signal must be multiplied by (A_0k_0+1) to obtain the same output signal, and secondly that the coefficient of the third-power term, which can be approximated by $(b-2a^2)/(1+A_0k_0)$ for $A_0k_0 \gg 1$ is also determined by the coefficient of the quadratic term in the original equation (23.1).

In the foregoing, we have assumed for these considerations that the amplifier satisfied equation (23.1). In practice, however, we always deal with amplifiers where the amplification depends on the frequency, so that the coefficients will be complex. A subdivision into frequency ranges where $|A(\omega)k(\omega)|$ acquires certain values, as with the linear situation, has little sense because other frequencies are usually created by distortion. It is also difficult to define a measure for distortion which is adequate under all circumstances. Existing definitions usually depend on certain signal forms, e.g. sinusoidal ones. The often heard statement that distortion decreases with the use of feedback, for which $|A(\omega)k(\omega)| \gg 1$ is valid in the frequency range of interest, must therefore be considered as an untidy extrapolation of the relations derived above. Because of parasitic capacitances, the relation between input and output voltage will always be given by a differential equation or by an equation with complex coefficients. Other capacitive or inductive effects will also often be present in the circuit. Nevertheless, even in these cases, the foregoing conclusions frequently prove to be applicable, particularly when the following conditions are satisfied:

1. The phase differences in the system can be neglected;
2. The signals do not exceed the limits of the region where the expansion into a power series is valid;
3. The working points of the active components are not changed by the signal.

As one is primarily interested in the effect of feedback on systems which have little inherent distortion, the above conditions are then usually fairly well satisfied, and to a good approximation we can use the results that we have derived for the theoretical case.

It is obvious that distortion will also decrease when smaller input signals are applied to the non-linear components. For a voltage amplifier, this can be achieved for the same output signal by increasing the impedance through which the output current passes. With normal amplifiers, this is usually tied up with a number of other complications, such as the effect of parasitic capacitances, but there is no difficulty with cathode or emitter followers where the total output voltage is used for feedback.

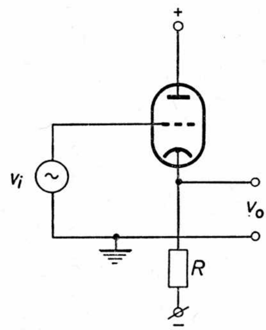

Fig. 23-2

Let us assume that we have for the triode of Fig. 23-2:

$$v_s = \frac{1}{S} i_a + \alpha i_a^2 + \beta i_a^3 + \ldots \tag{23.5}$$

where v_s is the control voltage $v_{gk} + v_{ak}/\mu$. We then have for the circuit: $v_s = v_i - (1 + 1/\mu)v_o$ and $v_o = i_a R$, so that:

$$v_i - \left(1 + \frac{1}{\mu}\right) v_o = \frac{1}{SR} v_o + \frac{\alpha}{R^2} v_o^2 + \frac{\beta}{R^3} v_o^3 + \ldots$$

or

$$v_i = \left(1 + \frac{1}{\mu} + \frac{1}{SR}\right) v_o + \frac{\alpha}{R^2} v_o^2 + \frac{\beta}{R^3} v_o^3 + \ldots$$

For $SR \gg 1$ the coefficient of v_o is nearly 1. The higher-power terms decrease as $1/R^2$, $1/R^3$, etc. Roughly speaking, the distortion decreases with the square of the cathode resistance, and for the emitter follower with the square of the emitter resistance. This is the additional phenomenon mentioned during discussion of the maximum signal magnitude permissible with a cathode follower (Section 14).

For the ECC81-valve applies at a quiescent current of 1 mA and an anode voltage of 100 V:

$$i_a = S\,v_s + a\,v_s^2 + b\,v_s^3 + \dots$$

with $S = 1.5$ mA/V, $a = 0.8$ mA/V^2, $b = 1.6$ mA/V^3,
from which follows for the coefficients α and β from (23.5):

$$\alpha = -\,0.24{\cdot}10^6 \text{ V/A}^2 \quad \text{and} \quad \beta = -\,0.15{\cdot}10^9 \text{ V/A}^3.$$

With a cathode resistance of 100 kΩ, the relation between v_i and v_o will be, to a good approximation:

$$v_i = v_o - 0.24{\cdot}10^{-4}\,v_o^2 - 0.15{\cdot}10^{-6}\,v_o^3$$

or

$$v_o = v_i + 0.24{\cdot}10^{-4}\,v_i^2 + 0.15{\cdot}10^{-6}\,v_i^3$$

When v_i is a pure sinusoidal voltage with an amplitude of, say, 10 volts, it follows from $\cos^2\omega t = \frac{1}{2} + \frac{1}{2}\cos 2\omega t$ and $\cos^3\omega t = \frac{1}{4}(\cos\omega t + \cos 3\omega t)$ that the ratios d_2 and d_3 of the second and third harmonic to the fundamental component in v_0 will be:

$$d_2 = 1.2 \cdot 10^{-4} \quad \text{and} \quad d_3 = 4 \cdot 10^{-6}.$$

24. Output impedance with feedback

So far we have assumed that the amplifier circuits are not loaded at their outputs by external current drains. If this is not the case, connecting a load impedance Z_l across the output terminals will change the output voltage. According to Thévenin's theorem the amplifier can be considered as a voltage source, where the source voltage is equal to the open-circuit voltage, in our case $A(\omega)v_i$, and the internal impedance equals the output impedance Z_o of the amplifier (Fig. 24-1).

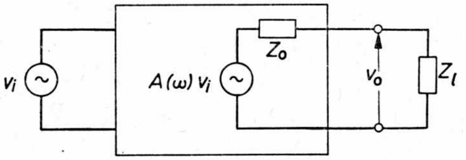

Fig. 24-1

We thus have:

$$v_o = A(\omega)v_i \cdot \frac{Z_l}{Z_l + Z_o} \tag{24.1}$$

If a fraction $k(\omega)v_o$ is fed back to the input (Fig. 24-2) we have:

$$v_i = v_s - k(\omega)v_o$$

and

$$v_o = A(\omega)\,\frac{Z_l}{Z_l + Z_o}\{v_s - k(\omega)v_o\}$$

which gives:

$$v_o = \frac{A(\omega)v_s}{1 + A(\omega)k(\omega)} \cdot \frac{Z_l}{\dfrac{Z_o}{1 + A(\omega)\,k(\omega)} + Z_l} \tag{24.2}$$

Comparison of (24.2) with (24.1) shows that the feedback amplifier behaves as a voltage source of potential $A(\omega)v_s/\{1+A(\omega)k(\omega)\}$, which we already knew, and internal impedance $Z_o/\{1+A(\omega)k(\omega)\}$. The output impedance has therefore become $\{1+A(\omega)k(\omega)\}$ times smaller. The result of applying feedback, where $|A(\omega)k(\omega)| \gg 1$, is that the amplifier more closely resembles an ideal voltage source.

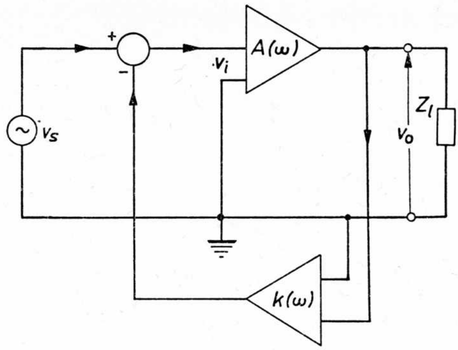

Fig. 24-2

Example: If the output stage of an amplifier consists of a cathode follower with an output impedance of $\approx 1/S$ and this amplifier is fed back with a loop gain Ak, the new output impedance will be $1/AkS$. With $S = 5$ mA/V and $Ak = 1000$, this gives 0.2 Ω, which is very low when compared to the usual impedance level for valve circuits. Particularly in the case of stabilized power supplies, one often uses this type of circuit. We shall refer to this in detail in Section 29.

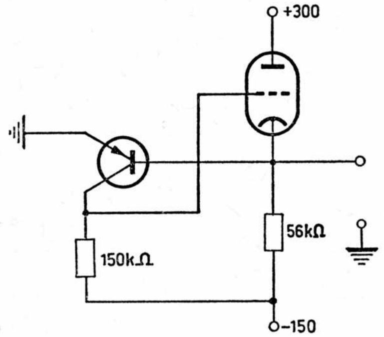

Fig. 24-3

An example of the above which occurs less frequently is the circuit of Fig. 24-3, which consists of the "hybrid" combination of a valve and a transistor, and where the collector-emitter voltage of the transistor is also the grid voltage of the valve. An output impedance of 2 Ω was measured for the given values. This circuit can, for example, replace the triode in the differentiating circuit of Fig. 12-5 as well as in the summing circuit of Fig. 12-7 if still lower impedances are required than already present in those cases.

We thus have for the output impedance with feedback:

$$(Z_o)_{fb} = \frac{Z_o}{1 + A(\omega)k(\omega)} \tag{24.3}$$

and for the output admittance:

$$(Y_o)_{fb} = \frac{1}{(Z_o)_{fb}} = \frac{1 + A(\omega)k(\omega)}{Z_o} = \frac{1}{Z_o} + \frac{A(\omega)k(\omega)}{Z_o}$$

This shows that as a result of feedback an apparent impedance is connected in parallel with the output impedance.

As we have mentioned, the feedback voltage can be derived not only from the output voltage but also from the output current. This may be achieved by inserting a resistor R_t in series with the load impedance and feeding back the voltage developed across R_t (Fig. 24-4).

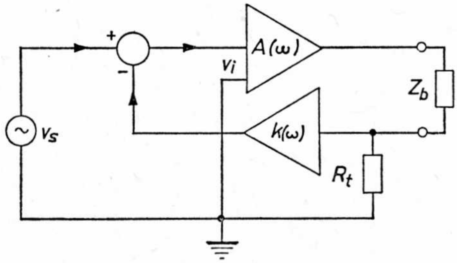

Fig. 24-4

Analogous to the previous case we find:

$$v_o = A(\omega)v_s \cdot \frac{Z_l}{Z_o + \{A(\omega)k(\omega) + 1\}R_t + Z_l}$$

From this follows that the output impedance has increased by $(Ak+1)R_t$ as a result of the feedback. The output thus shows a closer resemblance to a current source.

An intermediate behaviour is found when a feedback signal is derived partly from the output voltage and partly from the output current.

25. Input circuits with feedback

The above symbolically indicated combination of the input signal v_s and the fed-back signal $k(\omega)v_o$ causes practical difficulties, because one terminal of both is usually grounded, and this makes it impossible to connect v_s and $k(\omega)v_o$ in series without special precautions.

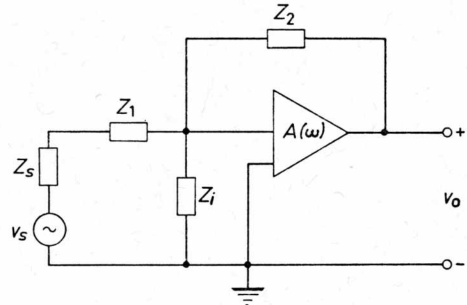

Fig. 25-1

Fig. 25-1 shows a frequently applied method. The feedback fraction k is in this case dependent on the internal impedance Z_s of the signal source and therefore the fraction is only defined when this impedance is either constant or sufficiently small. If this is the case, Z_s can be accounted for in Z_1. We then find for the relation between v_o and v_s:

$$-v_o\left\{\frac{1}{Z_2}+\frac{1}{A}\left(\frac{1}{Z_1}+\frac{1}{Z_2}+\frac{1}{Z_i}\right)\right\}=\frac{v_s}{Z_1}$$

Oscillation of the system can be prevented by selecting, in the frequency range where $|A|$ must be large, the argument of A in the neighbourhood of $-180°$. A is then almost real and negative. By making A large, and selecting impedances Z_1, Z_2 or Z_i so that

$$\frac{1}{A}\left(\frac{1}{Z_1}+\frac{1}{Z_i}\right) \ll \frac{1}{Z_2}$$

applies, we obtain:

$$v_o=-\frac{Z_2}{Z_1}\cdot v_s$$

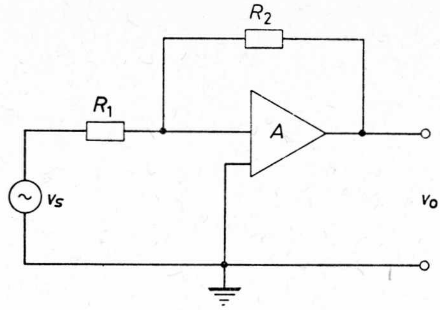

Fig. 25-2

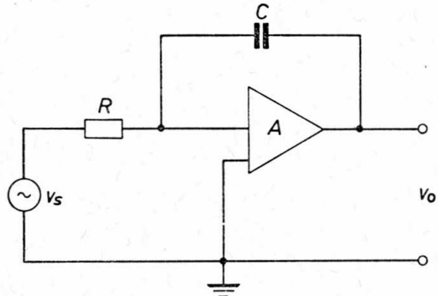

Fig. 25-3

This principle is often used in analogue computers in order to obtain certain mathematical relations between input and output voltages. For example, by using resistors for Z_1 and Z_2, one obtains a fairly accurate multiplication by a real constant (Fig. 25-2). When using a capacitor for Z_2 and a resistor for Z_1, Fig. 25-3, we obtain:

$$v_o = -\frac{1}{pRC} \cdot v_s$$

which is the same as:

$$\frac{\mathrm{d}v_o}{\mathrm{d}t} = -\frac{1}{RC} \cdot v_s \quad \text{or} \quad v_o = -\frac{1}{RC}\int v_s \cdot \mathrm{d}t$$

This means that this circuit integrates the input signal. We shall discuss the accuracy of this type of circuits in Section 42.

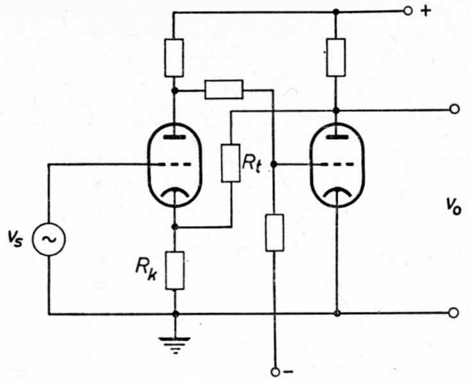

Fig. 25-4

In many amplifiers the feedback is applied to the cathode or emitter of the first stage; Fig. 25-4 gives an example. Here it is not necessary for the second stage to have a cathode resistance because the feedback ensures correct adjustment. The cathode of the first stage should not be decoupled and the cathode resistance R_k must be made small so that the amplification given by this stage is not reduced too much. This often results in a rather small value for feedback resistor R_t and a rather heavy drain on the output by both signal and quiescent currents. For this reason, a cathode follower is often used as the output stage (Fig. 25-5).

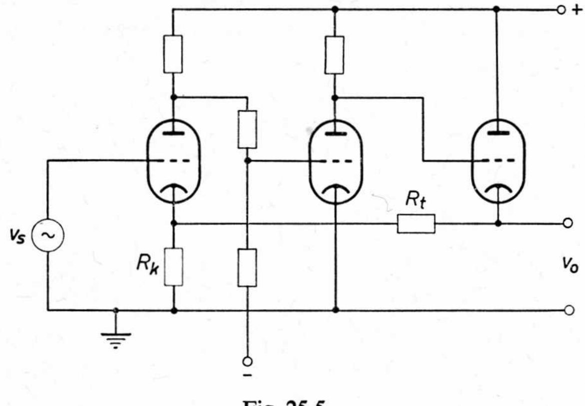

Fig. 25-5

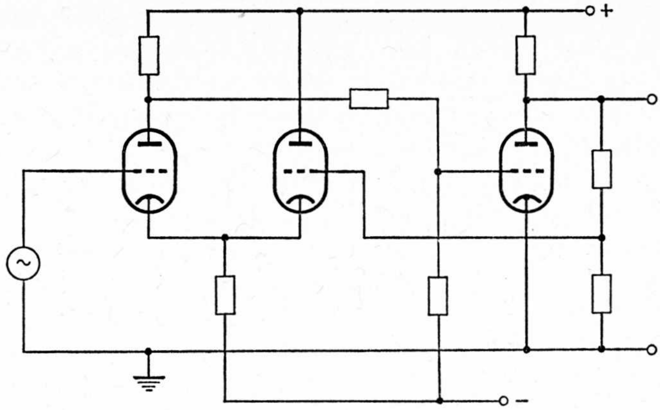

Fig. 25-6

Another possibility for feedback of the output voltage is shown in Fig. 25-6. We can choose a high value for the cathode resistor of the long-tailed pair because the amplification of the stage is almost independent of it. Since feedback is applied to a control grid, the voltage divider can in this case be designed with large resistors; thus the load on the output remains small.

In the basic feedback circuits shown in Fig. 22-1 *et seq.* the input and feedback voltages exert the same effect on the output signal:

$$v_o = A(v_s - v_t)$$

where $v_t = kv_o$.

It is obvious that the practical circuits discussed here, only give an approximation of the exact substaction which was assumed. Examples are found in the circuits of Figs 25-4 and -5. The output signal is here proportional to the current i_a passing through the first valve:

$$i_a = \frac{1}{S}\left\{v_g + \frac{v_a}{\mu} - v_k\left(1 + \frac{1}{\mu}\right)\right\}$$

where v_g equals the input signal v_s and v_k contains the feedback voltage

$$v_k\left(1 + \frac{R_t}{R_k}\right) = R_t i_a + v_o$$

This means therefore that the feedback factor equals the product of the factors $(1+R_t/R_k)^{-1}$ and $(1+1/\mu)$. From this it follows that the amplification is sensitive to possible variations in the first valve and particularly to replacements of the valve. Although the variations thus occurring are not very great, they are still considerably greater than those which may occur in the feedback circuit and they must certainly be taken into account for accurate amplifiers. In the circuit of Fig. 25-6, the ratio of the influences of the grids of the balanced stage is the determining factor. As we have seen when discussing the balanced stage in Section 19, this ratio is almost unity, but it nevertheless still depends on various parameters. When discussing the difference amplifier in Section 28, we shall indicate a few methods to make this deviation from unity (and also the fluctuations in the deviation) smaller than in a simple long-tailed pair. We shall see that the deviation can easily be reduced to less than 10^{-4}, and the effect of fluctuations in this amount is nearly always negligible.

In general, the input impedance of an amplifier will also be affected when feedback is used. For example, the input impedance of the circuit of Fig. 25-1 will be almost equal to Z_1 because for large values of the loop gain Ak the input signal v_i will be very small, so that the signal source must supply a current v_s/Z_1. The feedback in the circuits of Figs. 25-4 and -5 will ensure that the cathode of the first valve follows the grid almost completely. The portion of the input impedance between these electrodes will therefore take less current and hence be apparently increased. Among others this results in an apparent reduction of the parasitic capcitance C_{gk} of the valve.

Because of the small amplification of the first stage, the anode signal will also be reduced, and thus the apparent increase in the grid-anode capacitance (Miller effect) will become less. However, impedances between the input terminal and fixed points, such as leakage resistors and capacitances to earth, are not affected by this kind of feedback.

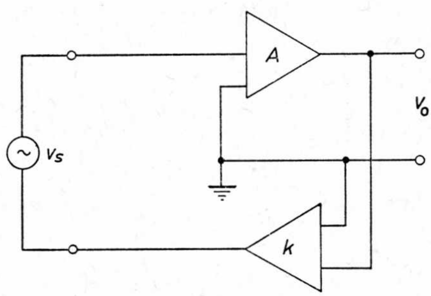

Fig. 25-7

Although some components of the input impedance are thus increased by feedback and input impedances *can* certainly be increased by feedback, the frequently heard statement that the input impedance will always be increased $(Ak+1)$ times, is incorrect in its generality.

The statement is, however, valid in the basic circuit of Fig. 25-7 for the terminals to which the signal source is connected, if no account is taken of parasitic impedances, and the output impedance of the "amplifier" k is neglected. The majority of practical designs, however, are not of this type, if only because v_s must be "floating".

26. Impedance transformations

We have seen several times in the foregoing that impedances are changed when using feedback. Especially in measurement electronics, we often utilize the facility of creating apparent impedances by means of active circuits, which would not have been possible by means of passive components alone. Circuits which realize this can generally be reduced to one of the basic circuits of Fig. 26-1 or 26-2, or to the equivalent circuits of Fig. 26-3 or 26-4.

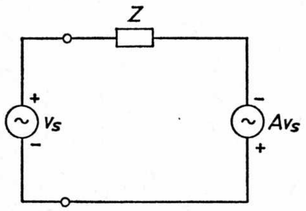

Fig. 26-1

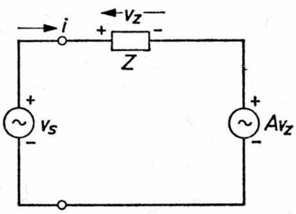

Fig. 26-2

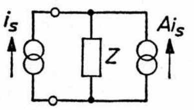

Fig. 26-3

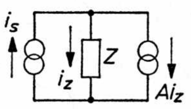

Fig. 26-4

Fig. 26-1 shows an impedance Z between signal source v_s and the derived signal source Av_s, which has one terminal in common with v_s. At the indicated polarity, the current passing through Z has the value $\{(A+1)/Z\}v_s$, so that it appears as though an impedance $Z/(A+1)$ is directly connected to the terminals of the signal source. The circuit on the right of the terminals therefore behaves as an impedance $Z/(A+1)$. This means for a positive value of A a reduction of the impedance Z. An example of this is furnished by the Miller effect, where Z is the anode-grid capacitance. For a negative A and $|A| < 1$, the transformed impedance becomes larger than the original one, whilst when $A < -1$, the new impedance gets the opposite sign to the original one. In this way we can create negative resistances and other negative elements. To obtain constant values for the transformed impedances, A must be given a fixed value, for example by feedback.

In the circuit of Fig. 26-2 the auxiliary source voltage is derived from the voltage across the impedance Z:

$$v_s = v_z + Av_z = (A+1)v_z = (A+1)\,Z\,i$$

or

$$i = \frac{v_s}{(A+1)Z}$$

The apparent impedance is then $(A+1)Z$.

Similarly we find that the impedance Z in Fig. 26-3 is multiplied by $(A+1)$ by means of the auxiliary current source Ai_s and that in Fig. 26-4 divided by $(A+1)$.

By making A frequency-dependent according to a certain function, it is possible to create various apparent impedances in this manner. We shall illustrate its utility by a few simple examples.

For instance, it is often a disadvantage in a low-frequency amplifier that the grid leak resistor must not exceed a given maximum value (e.g. 1 MΩ). This means that the coupling capacitor to the previous stage must have a very large capacitance to prevent a drop in amplification at low frequencies. Since this capacitor must usually be able to withstand a d.c. potential of several hundred volts, it will necessarily become rather bulky as well as

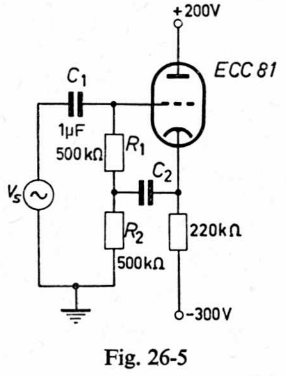

Fig. 26-5

expensive. In this case we can use the cathode follower circuit of Fig. 26-5, where the resistance R_1 is apparently multiplied μ times provided the capacitor C_2, which only carries a small d.c. voltage, is made sufficiently large.

For $v_o \ll v_s$ in the circuit of Fig. 26-6, the current i is practically equal to v_s/R and therefore:

$$v_o = \frac{1}{pRC} v_s$$

This means that v_o is proportional to the time integral of v_s. However, if v_o can no longer be neglected with respect to v_s, the current will be:

$$i = \frac{v_s - v_o}{R}$$

The proportionality then no longer applies. If the current through the capacitor became larger than i by an amount v_o/R, therefore again v_s/R, the circuit would continue integrating perfectly even when v_o was no longer small compared to v_s. This additional current can be supplied by means of the circuit shown on the left-hand side of Fig. 26-7, where a resistor of the

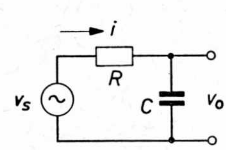

Fig. 26-6

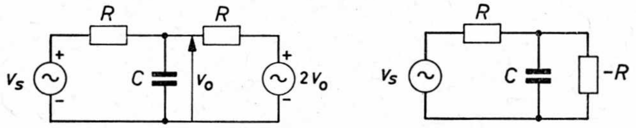

Fig. 26-7

value R is connected between the upper terminal of the capacitor and a voltage source 2 v_o. We have seen that this can be considered as though a resistance $-R$ were placed across the capacitor. A practical design for this circuit which in principle gives perfect integration, is shown in Section 42.

We shall meet many examples of feedback and impedance transformations in the following sections.

27. Adding what is lacking

Although feedback offers many advantages, there are also a number of disadvantages. The most important one is the danger of oscillation, which occurs particularly when a high degree of perfection is desired, which implies a very large value of Ak. As we have seen above, this oscillation is the result of the occurrence of a system with feedback, in which a portion of the output voltage is fed back to the input, and subsequently returns to the output in an amplified form. By applying what we shall call "adding what is lacking", no loop is created and oscillation is excluded in principle.

The avoidance of a loop system is a result of the fact that the correcting signal is not added to the input signal (as with feedback) but to the output signal.

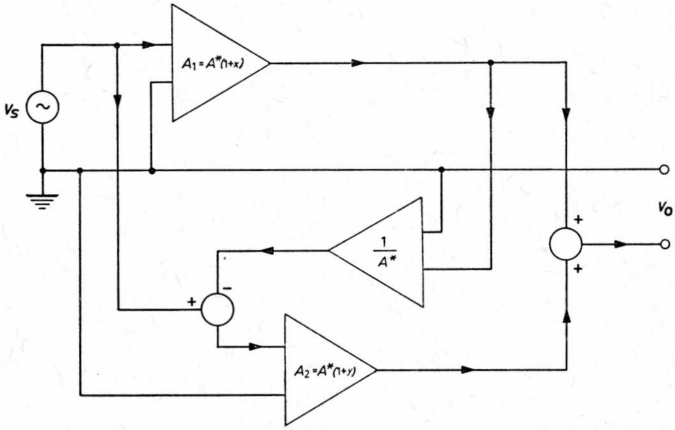

Fig. 27-1

The principle of the method of "adding what is lacking" is shown in Fig. 27-1. Let us assume that we desire to obtain an amplification A^* of the input signal v_s. We take care that the amplifications of the amplifiers used are about A^*, for example $A_1 = A^*(1+x)$; $A_2 = A^*(1+y)$; where x and y are 0.1 or smaller. The following situation is then obtained: The signal $A_1 v_s = A^*(1+x)v_s$ will occur at the output of A_1. From this output signal the exact fraction $1/A^*$ is taken and deduced from v_s. The difference signal $v_s - (1+x)v_s = -xv_s$ is offered to A_2 and appears at the output as $-A_2 xv_s = -A^*(1+y)xv_s$. The sum of the output signals of A_1 and A_2 is then $A^*(1-xy)v_s$.

The relative error in the gain of the total amplifier is therefore the product xy of the relative errors of both constituent amplifiers; in our example, of the order of 0.01. This process can be repeated with a third amplifier, and the relative error will then be the product of the three relative errors; in our example, of the order of 0.001. Especially when each constituent amplifier has been designed for a reasonably stable amplification by means of a safe feedback, this process converges rapidly, and a very high accuracy can be obtained with two amplifiers without the danger of oscillation. There is no real disadvantage in using two or more amplifiers since the second and possibly further amplifiers carry very little signal and can therefore be kept small. Moreover, there is no amplification loss as with feedback. This method is little known, and its usefulness is often underestimated.

A very simple example of the principle of "adding what is lacking" is shown in Figs 27-2, -3 and -4.

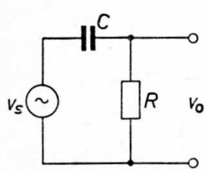

Fig. 27-2

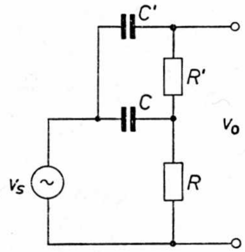

Fig. 27-3

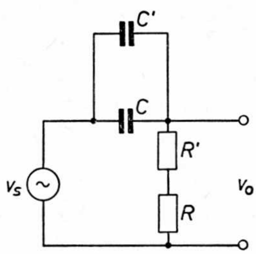

Fig. 27-4

With an a.c. amplifier, the loss in amplification at low frequencies due to the *RC*-coupled circuits is often a nuisance (Fig. 27-2). What is "lacking" here at the output voltage is the voltage lost across the coupling capacitor. A large part of this missing voltage can be added to the original output voltage by means of a second *RC*-circuit (Fig. 27-3), which ensures that the amplitude characteristic will continue to be flat at the lower frequencies. This could also be achieved by using the same elements in the way, shown in Fig. 27-4. However, even disregarding the fact that this does not affect the principle, the improvement is smaller than with the correct design of Fig. 27-3.

A peculiarity worth mentioning of the configuration shown in Fig. 27-3 is that the amplitude characteristic first rises before dropping as the low frequency region is approached and thus shows a maximum (Fig. 27-5, where it is assumed that $R=R'$ and $C=C'$). This makes it possible in an amplifier with more than one coupled stage, to compensate partly for the drop in one stage by the gain in another.

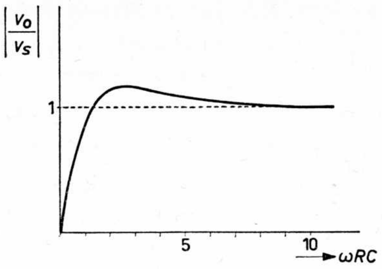

Fig. 27-5

There is yet another side to this same peculiarity, this time mainly theoretical. The output voltage is greater than the input voltage in the neighbourhood of the maximum. Although the gain is not large (maximum about 15 per cent), the property is maintained at a sufficiently small output load. In theory, the process can be repeated until any desired amount is reached. It is therefore possible to transform a voltage to a higher value not only by using transformers, but also by *RC*-circuits.

Fig. 27-6 gives an example of an active circuit where the principle of "adding what is lacking" has been applied. We want to derive accurately from voltages v_1, v_2 and v_3 the voltage $v = -(c_1v_1 + c_2v_2 + c_3v_3)$ where c_1, c_2 and c_3 are positive constants.

By means of reasonably accurate resistors r/c_1, r/c_2 and r/c_3 and a summation circuit I such as that shown in Fig. 12-7, we produce voltage v_P which is approximately equal to $-(c_1v_1 + c_2v_2 + c_3v_3)$, assuming that the input signal of amplifier II is zero. The voltage at point Q, the junction of the

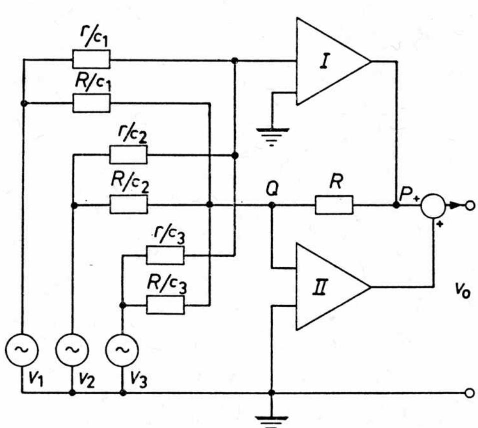

Fig. 27-6

precision resistors R/c_1, R/c_2, R/c_3 and R, is then almost zero. By connecting the input of amplifier II to point Q, a possible deviation of voltage v_Q from zero is reduced by the loop gain of this amplifier, so that v_P will satisfy the requirement more accurately with the same factor. A very high accuracy can be achieved with this circuit. A single feedback amplifier is usually employed for such a summation circuit (Fig. 27-7), but to reach the same accuracy, a much greater loop gain is required in this case, which easily gives rise to oscillation or requires a drastic restriction in bandwidth.

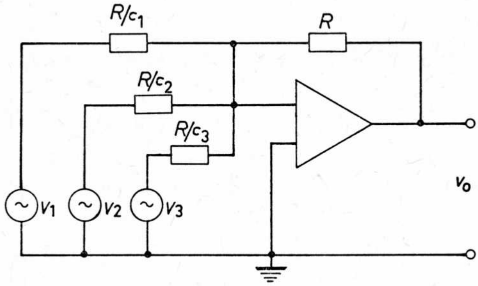

Fig. 27-7

We should say a few words about the summation circuit at the output of the basic circuit shown in Fig. 27-1. Particularly if the first amplifier also has feedback from the output voltage, the addition of the output signal must be done in such a way that the total output signal is not applied to the output of the first amplifier, because this would render the feedback of this amplifier ineffective. This is why simple parallel combination of two identical feedback amplifiers does not result in an improvement regarding them both. In principle therefore, the addition of output signals should be effected so that none of the outputs is affected by the output signal of the other amplifiers, for example, by using transformers or bridge circuits. It appears, however, that addition is possible without these devices if certain asymmetries exist between the feedbacks to the inputs of the various amplifiers. We thus obtain a "higher order" feedback which is called "parallel plant feedback" and which combines high accuracy with freedom from oscillation. It would, however, lead us too far from our main purpose to discuss this subject in greater detail here.

28. Difference amplifiers

Several times in the preceding sections we have met the situation that we must amplify a voltage that is presented between two terminals which both have a potential with respect to earth. The voltage to be amplified is then the difference between these two voltages. For example, we have tacitly assumed in the previous section when discussing the principle of 'adding what is lacking' that with amplifiers such as A_2 of Fig. 27-1, only differences of the various voltages were amplified. If this were not so, it would mean that the final accuracy is determined by the asymmetry in the amplification. In order to ensure a good execution of this principle under all circumstances, we must have amplifiers which only amplify the difference between two voltages. In measurement electronics, one often meets such a situation, where the difference between the two voltages may relatively become exceedingly small. For example, with a bridge circuit as shown in Fig. 28-1 it is not possible to earth simultaneously a terminal of voltage source *B* and a terminal of indicating instrument *M*, because this would mean short-circuiting one of the legs of the bridge. If one of the terminals of the indicating

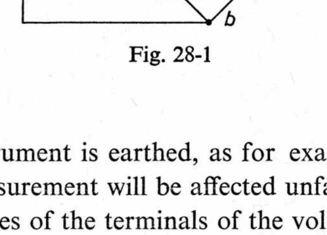

Fig. 28-1

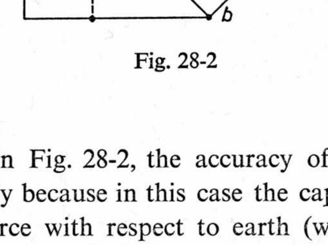

Fig. 28-2

instrument is earthed, as for example *a* in Fig. 28-2, the accuracy of the measurement will be affected unfavourably because in this case the capacitances of the terminals of the voltage source with respect to earth (which usually cannot be neglected) are connected in parallel to the impedances Z_1 and Z_2. This complication does not occur when one of the terminals of the voltage source is already earthed (*b* in Fig. 28-3). However, in this case it is necessary to measure the voltage between points *a* and *c* of the bridge with a non-earthed indicating instrument. When the bridge is approximately balanced, the measurement voltage between *a* and *c* will be small with respect to the almost equal voltages v_{ab} and v_{cb}. In this case there is again a need for an amplifier which exclusively amplifies the difference between two voltages. Its output signal must not be affected by the common part of the two voltages.

To avoid the situation which occurs in the case of the "floating" power supply, the impedances between the input terminals of this amplifier and earth must be sufficiently large. However, it is much easier to make an amplifier rather than a voltage source satisfy this requirement.

We shall see later that application of these "difference amplifiers" also facilitates the solving of many other measurement problems.

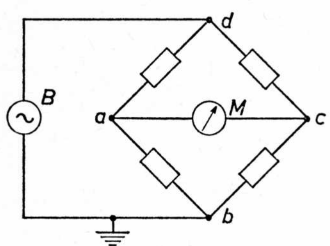

Fig. 28-3

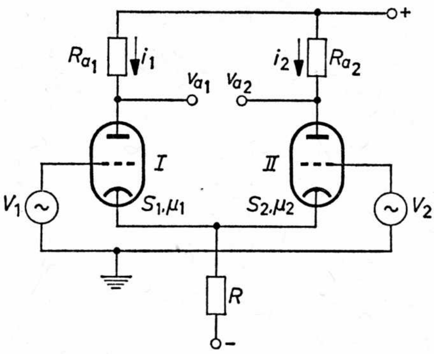

Fig. 28-4

It is obvious that the quality of a difference amplifier is first of all determined by the extent to which only the actual difference between the two voltages is being amplified, thus excluding the effect of that part which both voltages have in common. An example of an amplifier which possesses very favourable properties in this respect is the long-tailed pair circuit which was discussed in Section 19. We have seen that the following equations apply (19.2).

$$\left.\begin{aligned} Di_1 &= (R + \varrho_2)v_1' - Rv_2' \\ Di_2 &= -Rv_1' + (R + \varrho_1)v_2' \end{aligned}\right\} \qquad (28.1)$$

where
$$D = (R\varrho_1 + R\varrho_2 + \varrho_1\varrho_2),$$

$$\varrho_1 = \frac{r_{a1} + R_{a1}}{\mu_1 + 1}, \quad \varrho_2 = \frac{r_{a2} + R_{a2}}{\mu_2 + 1}$$

$$v_1' = \frac{\mu_1}{\mu_1 + 1} v_1, \quad v_2' = \frac{\mu_2}{\mu_2 + 1} v_2$$

In order to render i_1(or v_{a1}) proportional to the difference $v_1 - v_2$, we must satisfy:

$$(R + \varrho_2)\frac{\mu_1}{\mu_1 + 1} = R\frac{\mu_2}{\mu_2 + 1} \tag{28.2}$$

The same consideration for i_2 (or v_{a2}) gives the similar condition:

$$(R + \varrho_1)\frac{\mu_2}{\mu_2 + 1} = R\frac{\mu_1}{\mu_1 + 1} \tag{28.3}$$

It is obvious that both conditions cannot be satisfied at the same time, but it is possible to satisfy exactly at least one of these conditions by varying one of the components, e.g. μ_1 or μ_2.

The amplification factor of a valve can be varied by means of a resistor between anode and cathode (Fig. 28-5):

$$i = i_b + i_R = S(v_g - v_k) + \left(\frac{S}{\mu} + \frac{1}{R}\right)(v_a - v_k)$$

therefore:

$$\text{“}\mu\text{”} = \frac{S}{\dfrac{S}{\mu} + \dfrac{1}{R}} = \frac{1}{\dfrac{1}{\mu} + \dfrac{1}{SR}}$$

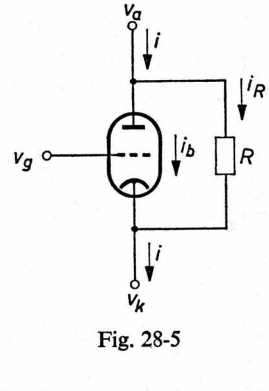

Fig. 28-5

However, the use of such a method is not satisfactory because, apart from the disadvantage of an unbalanced output signal, there is also the disadvantage that re-adjustment may be required because of ageing of the components or replacement of the valve. It is much better to design a circuit which can guarantee that it will be possible to measure the difference between the two voltages with a given accuracy, provided the tolerances of valves and resistors do not exceed certain maximum values. Conditions

(28.2) and (28.3) can be satisfied simultaneously to a very good approximation by making the amplification factors μ_1 and μ_2 very large with respect to unity, and resistor R very large with respect to ϱ_1 and ϱ_2.

At first sight it would seem that a large amplification factor could be obtained by replacing the triodes in Fig. 28-4 by pentodes. However, this is not possible without taking some precautions, for one finds that in this case the amplification factor $\mu_{g2g1}=(\partial i_a/\partial v_{g1})/(\partial i_a/\partial v_{g2})$ will appear in the formulae, and this is of the same order of magnitude as the amplification factor of a triode. It is only possible to use the large amplification factor $\mu_{1a}=(\partial i_a/\partial v_{g1})/(\partial i_a/\partial v_a)$ by providing special decoupling methods for the screen grids to the common cathode.

The cascode is better suited for this purpose. Here the amplification factor is approximately equal to the product of the amplification factors of the individual triodes, provided the voltage on the grid of the upper valve completely follows that of the cathode or the grid of the lower valve. Since no current flows through these control grids, the supply resistors may be large which greatly facilitates decoupling the grids.

The condition $R \gg \varrho_1, \varrho_2$ is satisfied by making R preferably larger than $10^6\ \Omega$, because ϱ_1 and ϱ_2 have approximately the value of $1/S$ (100–1000 Ω) at high values of μ_1 and μ_2. The use of an ordinary resistor of such a value in the common cathode circuit would require an extremely high negative supply voltage. A better solution is achieved by inserting a quasi-current source as discussed in Section 12. We now obtain the circuit of Fig. 28-6 where the decoupling of the two upper grids to the common cathode is indicated symbolically, for the time being, by a battery.

We can write the ratio of the coefficients of v_1 and v_2 in equation (28-1) for i_1 (or v_{a1}) as:

$$\frac{R+\varrho_2}{R}\cdot\frac{\mu_1}{\mu_1+1}\cdot\frac{\mu_2+1}{\mu_2}$$

which for large value of R, μ_1 and μ_2 can be approximated by:

$$\left(1+\frac{\varrho_2}{R}\right)\left(1-\frac{1}{\mu_1}+\frac{1}{\mu_2}\right)\approx\left(1+\frac{1}{SR}\right)\left(1+\frac{\mu_1-\mu_2}{\mu_1\mu_2}\right)\approx$$
$$1+\frac{1}{SR}+\frac{\mu_1-\mu_2}{\mu_1\mu_2}.$$

The deviation from unity is therefore $1/SR+(\mu_1-\mu_2)/\mu_1\mu_2$ where the most unfavourable case occurs when $\mu_1-\mu_2$ is positive, so that the maximum

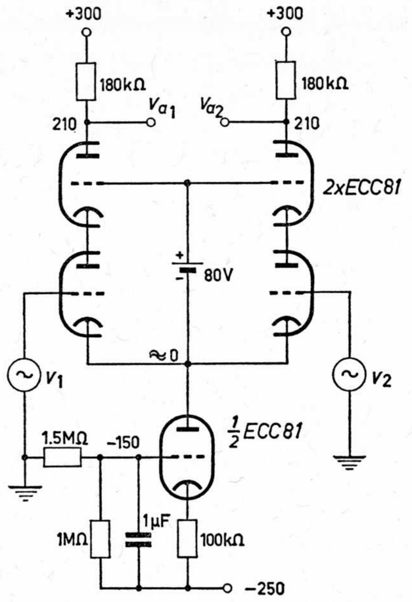

Fig. 28-6

deviation can be approximated by $1/SR + \Delta\mu/\mu^2$, where $\Delta\mu = |\mu_1 - \mu_2|$ and $\mu^2 = \mu_1\mu_2$.

With a possible tolerance of 10 per cent in the amplification factors of the triodes, this can amount to 20 per cent for the amplification factor of the cascode, so that the second term becomes $1/5\mu$ and the maximum deviation from unity $1/SR + 1/5\mu$.

The same values are found for the ratio of the influences of v_1 and v_2 on i_2 or v_{a2}.

For example, we have in Fig. 28-6: $R = 60 \times 100\ \text{k}\Omega = 6\ \text{M}\Omega$, $1/S = 10^3\ \Omega$, therefore $SR = 6{,}000$; $\mu_{\text{casc}} \approx \mu^2{}_{\text{triode}} \approx 3{,}600$, therefore $5\mu_{\text{casc}} \approx 18{,}000$.

In this case the cathode resistor is thus the limiting factor for the overall accuracy. A still smaller deviation from unity could be guaranteed by inserting a cascode current source in the cathode circuit.

However, in general, not v_{a1} or v_{a2} but the difference voltage $v_{a1} - v_{a2}$ is taken as the output voltage. This improves the situation provided that $R_{a1} \approx R_{a2}$.

According to (28.1) the following equation applies:

$$-D(v_{a1} - v_{a2}) =$$

$$[R_{a1}(R + \varrho_2) + R_{a2}R]v_1' - [R_{a1}R + R_{a2}(R + \varrho_1)]v_2' \qquad (28.4)$$

which gives for the ratio of the coefficients of v_1' and v_2':

$$\frac{R_{a1}R + R_{a1}\,\varrho_2 + R_{a2}R}{R_{a1}R + R_{a2}\,\varrho_1 + R_{a2}R}$$

and for those of v_1 and v_2, after substitution of the expressions for ϱ_1 and ϱ_2:

$$\frac{\mu_1(\mu_2+1)\ R(R_{a1}+R_{a2}) + \mu_1 R_{a1}(r_{a2}+R_{a2})}{\mu_2(\mu_1+1)\ R(R_{a1}+R_{a2}) + \mu_2 R_{a2}(r_{a1}+R_{a1})} =$$

$$1 + \frac{(\mu_1-\mu_2)[R(R_{a1}+R_{a2}) + R_{a1}R_{a2}] + \mu_1\mu_2\left[\dfrac{R_{a1}}{S_2} - \dfrac{R_{a2}}{S_1}\right]}{\mu_2(\mu_1+1)\ R(R_{a1}+R_{a2}) + \mu_2 R_{a2}(r_{a1}+R_{a1})}$$

As we always work with very large values of the cathode resistor, the second term in the divisor can be neglected. We thus find a value for the deviation from unity which is in any case smaller than:

$$\frac{\Delta\mu}{\mu^2}\left(1 + \frac{R_a}{2R}\right) + \frac{1}{2SR}\left(\frac{\Delta R_a}{R_a} + \frac{\Delta S}{S}\right) \tag{28.5}$$

where

$$\begin{aligned}
\Delta\mu &= |\mu_1 - \mu_2|, \quad \mu^2 = \mu_1\mu_2 \\
\Delta R_a &= |R_{a1} - R_{a2}|, \quad R_a = \tfrac{1}{2}(R_{a1} + R_{a2}) \\
\Delta S &= |S_1 - S_2|, \quad S^2 = S_1 S_2
\end{aligned}$$

If we now assume a maximum value of 10 per cent for all relative tolerances, this gives for the maximum possible deviation a value of:

$$\frac{1}{10\,SR} + \frac{1}{5\mu} \tag{28.6}$$

Compared with the previous case, where one of the anode currents or voltages was used as the output signal, the influence of the first term has become 10 times smaller, which represents a considerable improvement of the guaranteed value of the deviation. This is one of the reasons why it is usual with a difference amplifier stage to measure the output voltage difference across the anodes. Other reasons will become apparent later.

The following equation applies to the amplification A_d of the voltage difference $v_1 - v_2$ according to (28.4), if the amplification factors and R are all large:

$$|A_d| = \frac{R(R_{a1}+R_{a2})}{D} = \frac{2R_aR}{R(\varrho_1 + \varrho_2)} \approx \frac{2R_aR}{R}\cdot\frac{S}{2} = SR_a.$$

This result follows directly from the fact that under control of a voltage difference the voltage of the common cathode hardly changes, and the cascodes therefore represent the "pentode amplification": transconductance × anode resistor. As both $v_{a1}-v_{a2}$ and $v_{a2}-v_{a1}$ can be used as an output, A_d can be either positive or negative.

If $v_{a1}-v_{a2}$ is the output signal in Fig. 28-6, the maximum deviation from unity in the ratio of the coefficients of v_1 and v_2 will be $60{,}000^{-1}+18{,}000^{-1}=14{,}000^{-1}$. With $S\approx 0.7$ mA/V we find for the amplification of the input voltage difference: $180\cdot 10^3 \times 0.7\cdot 10^{-3}\approx 125$.

We can thus write for the output signal v_o of a difference amplifier:

$$v_o = A_d\{v_1 - (1+\delta)v_2\} = A_d\{(1+\delta/2)(v_1-v_2) - (\delta/2)(v_1+v_2)\}$$
$$\approx A_d\{(v_1-v_2) - (\delta/2)(v_1+v_2)\} =$$
$$A_d\{(v_1-v_2) - H^{-1}(v_1+v_2)\}.$$

The value $H=2/\delta$ is a measure for the quality of the difference amplifier, and is called the rejection factor or common mode rejection ratio. For the guaranteed minimum rejection factor of a balanced circuit with cascodes, at possible differences of 10 per cent between corresponding values, follows from (28.6):

$$H_{\min} = \frac{20}{\dfrac{1}{SR} + \dfrac{2}{\mu_{\text{casc}}}} \tag{28.7}$$

In the last example, the rejection factor had therefore a guaranteed minimum value of $2 \times 14{,}000=28{,}000$.

Fig. 28-7 shows the rejection factors as measured for various ECC81 valves in the circuit of Fig. 28-6. It follows that the assumption of 10 per cent for the relative variations agrees well with practice.

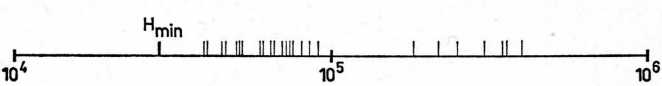

Fig. 28-7

The term $2/\mu$ in (28.7) was caused by the 20 per cent tolerance assumed for the amplification factors of the cascodes. For a balanced stage with triodes or pentodes, at possible differences of 10 per cent between all corresponding values, equation (28.7) becomes:

$$H_{\min} = \frac{20}{\dfrac{1}{SR} + \dfrac{1}{\mu}} \tag{28.8}$$

The amplification given by a single balanced stage will usually not be sufficient, so that more stages will normally be used. We shall see that the favourable properties of the first stage can be spoiled in this case and we must therefore take precautions against this happening.

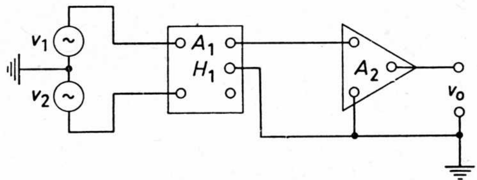

Fig. 28-8

If one of the anode voltages is used as the output voltage of the first balanced stage, the following stages can in principle be single-sided. The rejection factor H_{tot} of the complete amplifier is then the same as that for the first stage (Fig. 28-8). We have, however, seen that a single-sided output does not guarantee optimum rejection. Therefore, if the voltage difference between the two anodes is used as the output voltage of the first stage, we should take into account the fact that these anodes also carry a common signal (hence the smaller rejection factor with a single-sided output).

At sufficiently large values of the amplification factors and R, we find for Fig. 28-4 that the circuit operates as a cathode follower for the common component $(v_1+v_2)/2$ of the input voltages v_1 and v_2, so that the current through the common cathode resistance will change by $(v_1+v_2)/2R$. Both cascodes will each carry half of this current to a first approximation, so that the anode voltages will together change by about $(R_a/4R)(v_1+v_2)$. The amplification for the common input voltage $(v_1+v_2)/2$ is thus $R_a/2R$.

An amplification SR_a was found for the input voltage difference v_1-v_2. The ratio of the amplifications of the difference signal and of the common signal is called the discrimination factor (F), which is therefore $2SR$. We find for the circuit of Fig. 28-4: $F=2\times 0.7\cdot 10^{-3}\times 6\cdot 10^6 \approx 8\cdot 10^3$, i.e. 6 times smaller than the minimum possible rejection factor. The common input voltage thus gives a common signal on the anodes which is many times larger than the contribution in the difference voltage. Thus, if it is required that the guaranteed rejection factor does not become appreciably smaller than that of the first stage, the second stage must have a rejection factor of 100 or more, so that the common voltage on the anodes of the first stage has a negligible effect on the output (difference) signal of the second stage. Such a rejection factor can be easily guaranteed by using a simple long-tailed pair for this stage. With a design as shown in Fig. 28-9, the minimum rejection factor according to equation (28.8) will be:

$$H_{\min} = \frac{20}{\frac{1}{140} + \frac{1}{27}} \approx 450$$

The discrimination factor of such a simple stage is $2 \times 0.7 \cdot 10^{-3} \times 200 \cdot 10^{3} \approx 300$, so that the common voltage on the anodes of the first stage will be reduced by this factor with respect to the difference voltage between the anodes.

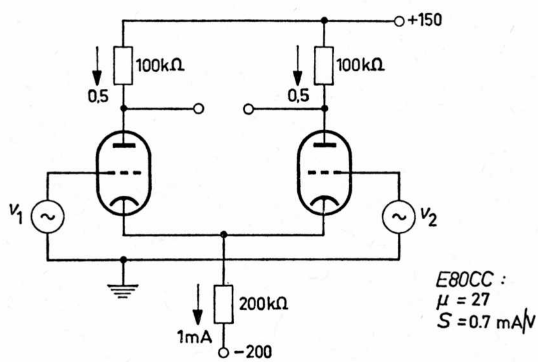

Fig. 28-9

With a two-stage amplifier consisting of a cascode circuit according to Fig. 28-6 and a second stage according to Fig. 28-9, we obtain the situation shown in Fig. 28-10, in which the figures indicate the order of magnitude of the relevant factors.

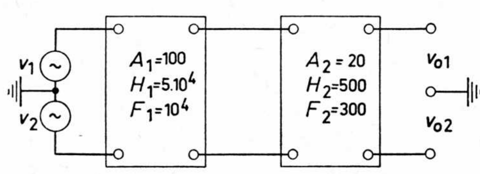

Fig. 28-10

A difference voltage $v_1 - v_2$ at the input yields a difference voltage $v_{o1} - v_{o2}$ at the output, which is A_1A_2 times larger. In Fig. 28-10 this product is 2,000.

The common voltage $(v_1+v_2)/2$ yields a common voltage at the output of:

$$\frac{v_{o1} + v_{o2}}{2} = \frac{A_1A_2}{F_1F_2} \cdot \frac{(v_1 + v_2)}{2}$$

This makes the total discrimination factor equal to F_1F_2, i.e. the product of the discrimination factors of the individual stages. In Fig. 28-10, this gives $F_1F_2 = 3 \cdot 10^6$ and the factor A_1A_2/F_1F_2 has a value of $7 \cdot 10^{-4}$.

The difference voltage $v_{o1} - v_{o2}$ at the output, which is caused by the common signal $(v_1+v_2)/2$ at the input, is calculated as follows. $(v_1+v_2)/2$ gives at the output of the first stage a common voltage $A_1(v_1+v_2)/2F_1$ and a difference voltage $A_1(v_1+v_2)/2H_1$. These give at the output of the second stage the difference voltages

$$\frac{A_1A_2}{F_1H_2}\left(\frac{(v_1 + v_2}{2}\right) \text{ and } \frac{A_1A_2}{H_1}\left(\frac{v_1 + v_2}{2}\right)$$

respectively, and a total of

$$v_{o1} - v_{o2} = A_1A_2\left(\frac{1}{F_1H_2} + \frac{1}{H_1}\right) \cdot \left(\frac{v_1 + v_2}{2}\right)$$

We find therefore for the rejection factor H_{tot} of the entire amplifier:

$$H_{\text{tot}} = \frac{1}{\dfrac{1}{H_1} + \dfrac{1}{F_1H_2}}$$

We have in Fig. 28-10: $F_1H_2 = 5 \cdot 10^6$ and $H_1 = 5 \cdot 10^4$, so that H_1^{-1} is the determining factor. In other words, the rejection factor of the amplifier approximately equals the rejection factor of the first stage. Theoretically, since the total discrimination factor is $F_1F_2 = 3 \cdot 10^6$ times smaller than the difference amplification, no balanced amplifier need be used after the second stage, but we shall see that in practice it is often easier, and even desirable, to design also the following stages as balanced ones.

Because the discrimination factor of a multi-stage amplifier equals the product of the discrimination factors of the individual stages, the total sum amplification will become exceedingly small in the case of an amplifier completely designed with balanced stages. In the above example, this amounted to $7 \cdot 10^{-4}$. This means that large common input signals will not cause overloads anywhere in the circuit. The maximum permissible value of the common signal is therefore entirely determined by the adjustment of the first stage.

Let us take as an example the circuit of Fig. 28-6. The cathode resistance is approximately 6 MΩ and the anode resistance 180 kΩ, so that the sum amplification is less than 0.02. As the grids of the upper valves follow the common cathode, the working point of the lower valves will hardly change at all. The maximum value of a positive common signal is therefore determined by the voltages of the upper triodes. If it is necessary for the valve's anode voltage to be 50 volts or more in order to make the grid bias sufficiently negative at a current of 0.5 mA, the allowed value will be 210 − (80 + 50) = 80 volts at the given conditions.

The maximum negative value is determined by the minimum permissible anode voltage of the ECC81 valve in the common cathode lead. If this is 75 volts, this gives 150 − 75 = 75 volts.

Because the rejection factor of the first stage determines the quality of the complete amplifier, it is essential that a good first stage is not preceeded by circuits that spoil the rejection factor before amplification takes place. For example, in the case of an a.c. amplifier where the inputs must be coupled indirectly to the signal sources (Fig. 28-11), one must ensure that the voltage divider factors $R_g/(R_g+R_b+1/j\omega C)$, where R_b = source resistance, for the two inputs should differ so little from each other, that the common component of the signal voltages gives a negligible difference signal between the inputs. Any adjustment can be avoided by selecting a very large value for R_g with respect to the impedance $R_b+1/j\omega C$ so that the voltage divider factors approximate to unity. If necessary, R_g can be apparently increased by means of the cathode follower circuit of Fig. 26-5.

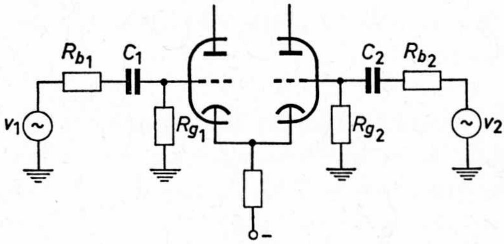

Fig. 28-11

We take as an example the case that $1/\omega C \ll R_b$, which makes the voltage dividers $R_{g1}/(R_{g1} + R_{b1})$ and $R_{g2}/(R_{g2} + R_{b2})$ respectively. With $R_{g1}, R_{g2} \gg R_{b1}, R_{b2}$ this becomes: $1-(R_{b1}/R_{g1})$ and $1-(R_{b2}/R_{g2})$ so that the difference between these factors becomes (R_{b1}/R_{g1}) — (R_{b2}/R_{g2}). With $R_{g1} \approx R_{g2}$, this reduces to $\Delta R_b/R_g$. For example, if the difference between the two source resistances is 1 kΩ and it is required that the difference between the voltage dividers is less than 10^{-5}, the grid leakage resistors must be 10^8 Ω.

An additional advantage of using a cathode follower circuit for the apparent increase of R_g is that this can also cause an apparent decrease in the parasitic capacitance of possible input leads (as found in cardiography, etc.) in the manner indicated in Fig. 28-12.

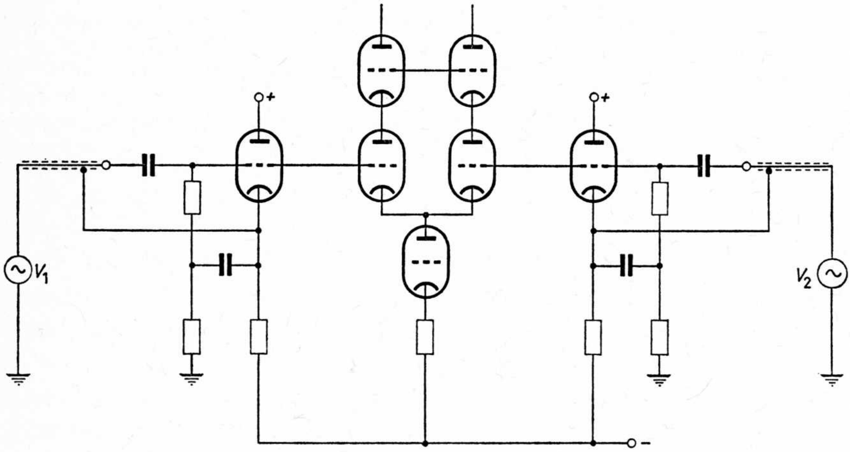

Fig. 28-12

If possible, the inputs should be coupled directly to the signal sources. We then obtain a voltage division $Z_i/(Z_i+R_b)$, where Z_i=input impedance of the cascode stages. This amounts to a few picofarads in parallel with a resistor, the value of which is mainly determined by the grid leakage, and which can easily be made to exceed 10^{11} Ω. Z_i, therefore, is represented by a very large resistance at low frequencies, and the voltage division will closely approximate to unity.

Since the high rejection factor in the discussed circuits is obtained by increasing the value of certain resistances (internal resistance of the valves, common cathode resistance), the value of the guaranteed rejection factor will decrease at high frequencies. For example, the heater-cathode capacitance will be in parallel with the cathode resistor, and short-circuit the latter at high frequencies.

The cathode resistor in the circuit of Fig. 28-6 amounted to 6 MΩ. If the cathode has a capacitance of 5 pF to earth, its impedance will be 6 MΩ at a frequency of 5000 c/s.

In principle it would be possible to obtain an apparent increase in cathode impedance by using feedback, because a decreasing cathode resistance makes itself felt by an increase in the currents passing through the cascodes when controlled by a common signal on the grids. Therefore, we can apparently increase the cathode impedance by measuring the change in current in the common anode circuit and feed it back after amplification to the cathode (Fig. 28-13). This makes it possible to design difference amplifiers

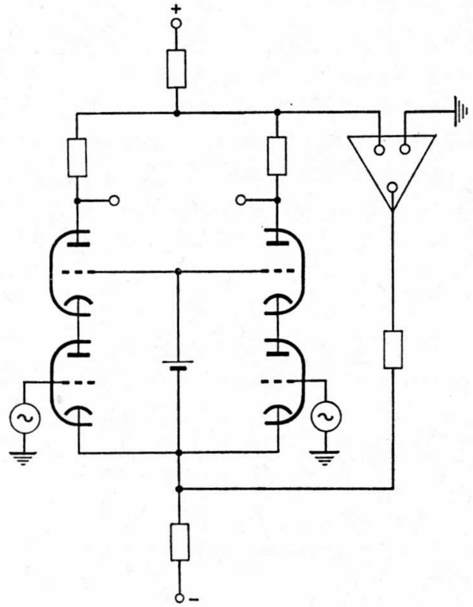

Fig. 28-13

with large rejection factors even at high frequencies. There is, however, little need for these in practice. Measurements where a rejection factor of 10^4 or more is required often occur at frequencies below 10 kc/s, but a requirement for a measurement which needs a rejection factor of, say, 10^3 at 1 Mc/s only occurs very rarely. This is not only due to the fact that measurements do not usually have to be very accurate at these frequencies, but also that at these frequencies common signals already cause a difference voltage at the input, because of parasitic capacitances, which cannot be distinguished anymore from the required signal as described here. An amplifier with a pentode or cascode balanced first stage, and a common cathode resistance of several tens of kilo-ohms is usually sufficient at these frequencies.

In the literature one often comes across the statement that the difference amplifier acquires better properties because of feedback of the common voltage. This is, however, only valid for circuits where the amplification of the common signal has not yet been drastically reduced, i.e. where the discrimination factor is small. This type of circuit also has, however, a small guaranteed rejection factor. For low frequency difference amplifiers, where both the discrimination factor and the guaranteed rejection factor can be given large values, this type of feedback, with its small "loop"-gain, is for this purpose of little use.

At low measurement frequencies there is sometimes a need for a still higher rejection factor than can be guaranteed by the circuits given here. In such a case, the solution can be found by using an inherently good difference amplifier and making every point of it follow the common signal e_g (Fig. 28-14). If the common signal is followed completely but by a fraction $1/P$, the common signal for the amplifier is apparently decreased P

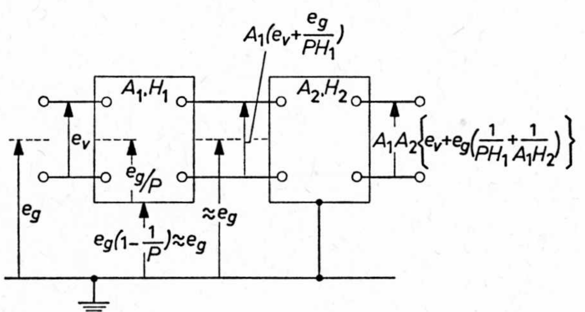

Fig. 28-14

times and will thus yield a correspondingly smaller difference voltage at the output of this amplifier. A second, non-following, difference amplifier can easily render the effect of the still present common signal e_g sufficiently small with respect to the difference signal A_1e_v. As can be seen from the figure, the rejection factor H_{tot} of the entire amplifier will be to a good approximation: $(1/H_1P + 1/A_1H_2)^{-1}$. For frequencies below 10 kc/s, including d.c., guaranteed rejection factors of 10^6 or even 10^7 can be achieved in this manner. The following of the common signal by an amplifier plus its supply obviously necessitates many precautions, so that this method will only be applied in very exceptional cases.

To take a practical case a difference amplifier with high input impedance at frequencies of approx. 100 c/s was required to measure the Hall effect with semiconductor probes. The amplifier had to cope with a common signal of approx. 400 V_{pk-pk}. The rejection factor had to be larger than 10^6, and a maximum amplification of the difference signal of approx. $3 \cdot 10^6$ was required.

These requirements could be met by applying the above described principle, where $A_1 = 50{,}000$, $(H_1)_{min} = 20{,}000$, $P = 200$, $(H_2)_{min} = 500$, $A_2 = 70$; so that the total amplification was approx. $3.5 \cdot 10^6$ and the total guaranteed rejection factor $4 \cdot 10^6$. The measured values of the rejection factors differed between 8 and $15 \cdot 10^6$ for five different input valves. The total input capacitance was less than 1 pF per metre of cable, thus retaining a high value for the rejection factor even when the difference between the internal resistances of the signal sources is considerable.

One may wonder if it is possible to eliminate the effect of the common signal by means of an input transformer (Fig. 28-15). Apart from the fact that this method cannot be applied at d.c. voltage and very low frequencies, it also possesses the following disadvantages:

a. Low input impedance, at least at low frequencies, which is difficult to increase by feedback because of the presence of the common signal on the primary of the transformer.
b. Easily occurring capacitive cross-coupling of the large common signal to the secondary of the transformer, which may cause a voltage difference at the input of the amplifier.
c. A flat amplitude characteristic over a frequency band is difficult to obtain because of the inherent resonances of the transformator.
d. Because of the normally very low signal levels, these transformers must be extremely well screened against extraneous magnetic fields (mains).

Despite all these disadvantages it can happen that conditions of measurement make the use of a transformer indispensable, for example, if the common signal is very large and yet a transistorized design is required. It is obvious that the elimination of the above-mentioned disadvantages is a rather extensive and expensive process.

The use of transformers in input circuits shall be dealt with in Section 32.

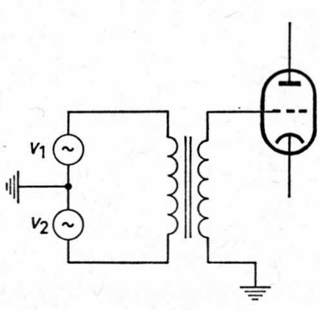

Fig. 28-15

Before discussing a number of properties and applications of difference amplifiers, we should indicate how it can be ensured that the grids of the upper triodes of a cascode difference amplifier (or the screen grids of a pentode difference amplifier) acquire the correct d.c. voltage and follow the common cathode, without excessively reducing the total impedance of the cathode to earth. If the latter happened both the guaranteed rejection factor and the discrimination factor would be reduced.

In theory, decoupling to the common cathode could indeed be achieved as indicated in Fig. 28-6, i.e. with a battery. In practice, however, this method is undesirable. For an a.c. cascode amplifier, decoupling according to Fig. 28-16 (left-hand side) can be applied. The parallel combination of R_1 and R_2 (in the example 2/3 R_2) should preferably be very much larger than the resistance in the common cathode circuit. Since no current passes through the control grids of the triodes, this requirement is easily satisfied. By connecting the lower terminal of R_2 to the cathodes (right-hand side of Fig. 28-16), only $R_1(=2R_2)$ is connected in parallel to the common cathode

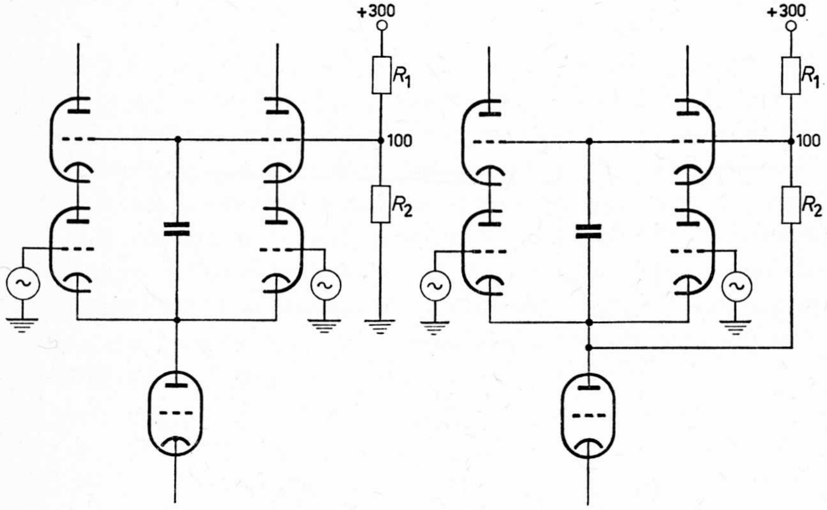

Fig. 28-16

resistance. Therefore, this last method is to be preferred. The impedance of the capacitor must then be small compared with R_1 at the lowest frequency to be amplified.

With d.c. amplifiers the decoupling of the grids is more difficult. Most often a stabilizing valve is used. The current-voltage characteristic of such an element has approximately the form, shown in Fig. 28-17. The nearly horizontal portion is due to the fact that with a glow discharge in an inert gas, the

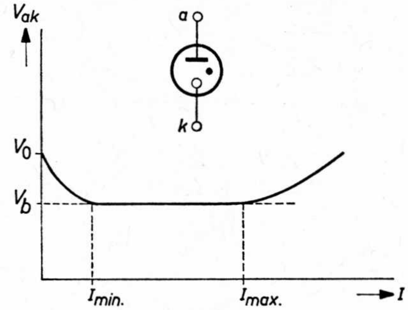

Fig. 28-17

voltage is to a very large extent independent of the current density over a wide range. Normally used gases are neon, helium and argon. Stabilizing valves for current ranges varying from a few to hundreds of milliamps are commercially available. For example, neon valve 85A2 has a burning or stabilizing voltage of approx. 85 volts at currents between 1 and 8 mA, whilst the slope of the curve in this region corresponds to a resistance of approx. 300 Ω.

When using stabilizing valves in amplifiers, we must first take into account that the quoted internal resistance of 300 Ω is only valid for d.c. The a.c. resistance is greater than the d.c. value. For example, the impedance of the 85A2 valve at 10^4 c/s is approx. 1000 Ω. Moreover, the burning voltage of stabilizing valves shows small spontaneous fluctuations which can interfere with sensitive amplifiers. Quantitative data will be given in Section 29.

In most cases the effect of these fluctuations can be reduced sufficiently by placing a capacitor across the valve. Since the valve behaves as a self-inductance at high frequencies, it is desirable to damp the circuit which occurs on connecting the capacitor by means of a series resistance of approx. 100 Ω.

Decoupling the grids of a difference amplifier by means of a stabilizing valve as indicated in Fig. 28-18 is not allowed, because the cathode resistance would be drastically reduced by the shunting effect of the supply resistor R. We can eliminate this disadvantage by the apparent increase of R for signal

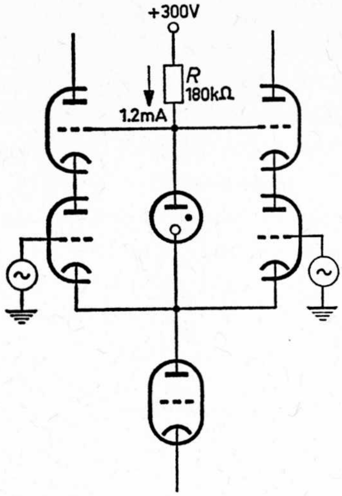

Fig. 28-18

currents, by inserting a cathode follower.

In Fig. 28-19, the upper terminal of R follows the voltage of the grids to

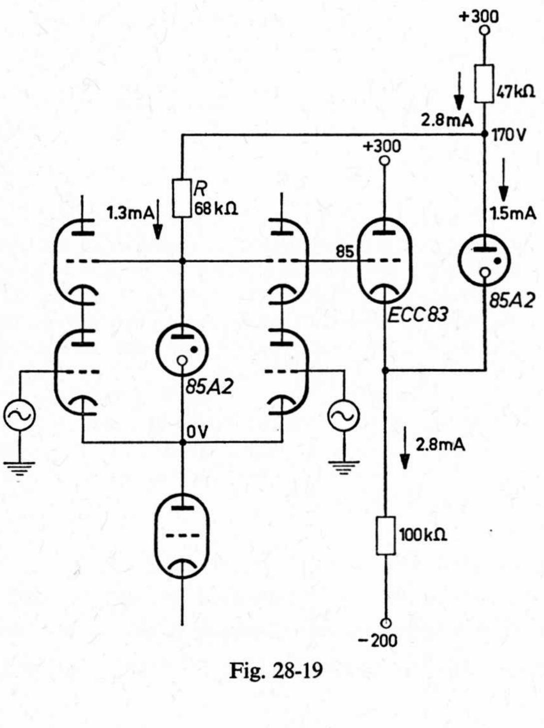

Fig. 28-19

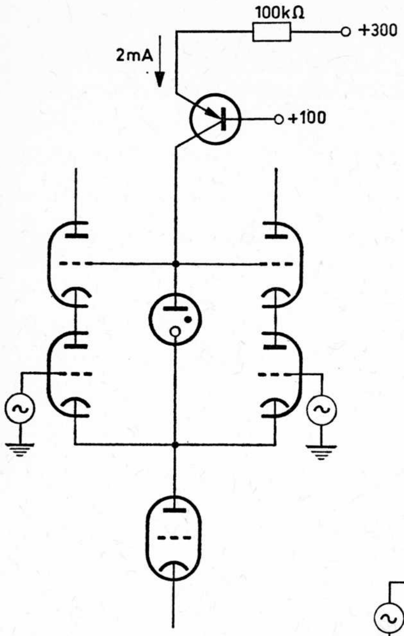

Fig. 28-20

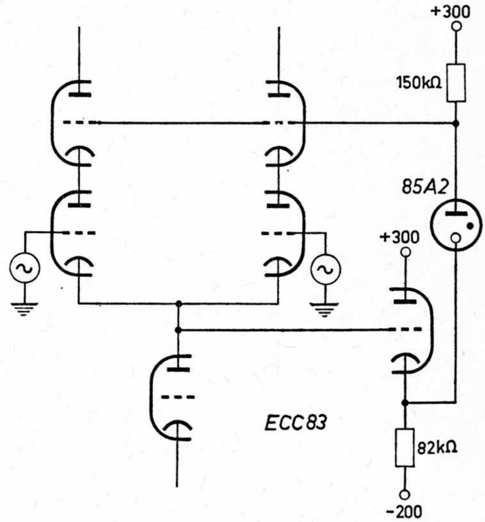

Fig. 28-21

a small fraction $1/k$, so that R is apparently increased by a factor k. When using valve ECC83 with $\mu = 100$ and $S \approx 1.5$ mA/V, k will be about 30 for the indicated values ($1/k = 1/SR_k + 1/\mu$; $R_k = 47$ kΩ//100 kΩ), so that R has an apparent value of 2 MΩ.

A simpler and more effective method for increasing the resistance to the positive supply rail consists of replacing R by a transistor current source, as discussed in Section 21. When using transistors of the alloy-diffusion type, it is easy to obtain resistances of a few meg-ohms (Fig. 28-20). On the other hand, when a cathode follower circuit is used as indicated in Fig. 28-21, the cathode resistance will not be reduced at all. The upper grids will follow the cathode to within a small fraction. The difference is approx. 2.5 per cent for the values given.

The equation for the rejection factor of a cascode difference amplifier can also be applied when the grids of the upper triodes do not completely follow the common cathode but within a fraction $1/k$, provided for the amplification factor of the cascode is taken:

$$\mu_{\text{casc}} = \frac{\mu_1 \mu_2}{1 + \dfrac{\mu_2}{k}}$$

where μ_1 and μ_2 are the amplification factors of the lower and upper triodes.

The circuits of Figs 28-19, -20 and -21 can also be used when d.c. or a.c. amplifiers with pentodes are applied. For an a.c. amplifier, the circuit of Fig. 28-22 is also possible, because the voltage divider $R_1 - R_2$ can be given a high value.

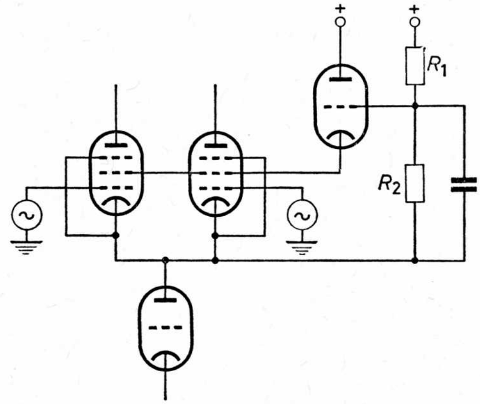

Fig. 28-22

Most applications of difference amplifiers with valves are based on the ability to compare accurately two voltages without loading the signal sources. The possible difference between the two must then be amplified, but usually it is not necessary for the amplification to be known accurately. The amplified difference signal serves as an indication or as a control voltage. For example, when measuring the temperature of a thermostatically controlled enclosure by means of a Wheatstone bridge, one of the resistors, which is strongly temperature-sensitive is placed in the enclosure; the bridge voltage will then serve, after amplification, as the control voltage for the power supply to the oven, so that the system tends to keep the bridge in the balanced condition. The "loop" amplification does not have to be fixed too accurately.

The same situation is also found in feedback amplifiers (as discussed in the previous sections), where a difference amplifier is used to compare the input voltage with a part of the output voltage. However, there are many measurements where the common signal only occurs as an interference signal and where the difference voltage is the actual measurement signal. For example, in encephalography, small changes in electrical potential between two points on the scalp are measured to record the electrical activity of the brain. These voltages are of the order of magnitude of microvolts and contain frequencies of approx. 0.1 to 100 c/s. However, the human body possesses a certain capacitance with respect to the mains supply. This capacitance can be reduced to a few picofarads by taking precautions in positioning the mains cables. Assuming a value of 5 pF, this gives a current of $220 \times 2\pi \times 50 \times 5 \cdot 10^{-12} \approx 0.3\ \mu$A at 50 c/s, 220 volts. The patient is usually earthed by means of an electrode with an area of a few cm^2, which is covered with a conducting jelly. Even when this electrode is pressed tightly against the patient's body, the contact resistance between body and electrode

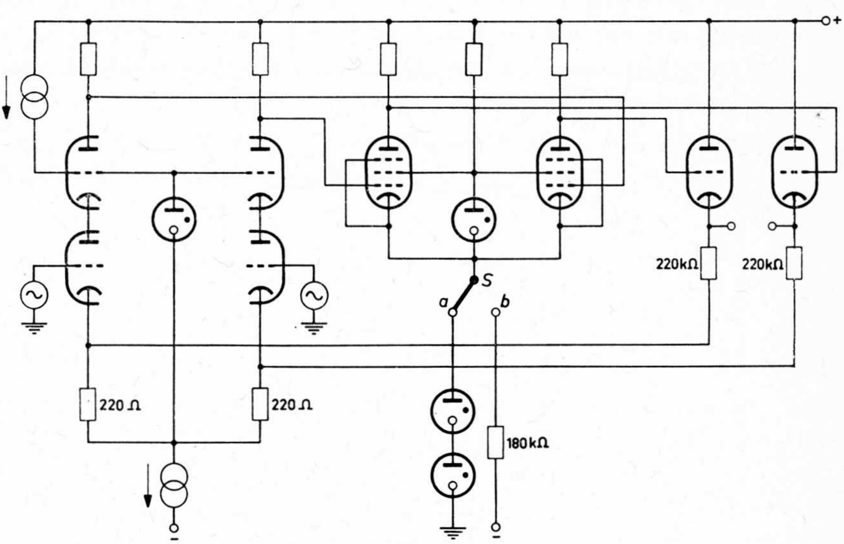

Fig. 28-23

can still amount to several kilo-ohms, so that the entire body can have a 50 c/s potential of a few millivolts to earth. In order to make this interference signal sufficiently small, the amplifier must have a rejection factor of 10^3–10^4.

In cases like this it is sometimes desirable, (as with the single-sided amplifier), for the amplification to be fixed accurately, which means that feedback must be used for the difference amplification. However, there is the danger that the rejection factor of the amplifier will be spoilt. This follows directly from the fact that the "open" amplification for the common signals is much smaller than for the difference signals, thus also the loop gain and hence the reduction factor and this increases proportionally the influence of the common signals. Whether this also leads to a degradation of the rejection factor, will be determined by the asymmetries. A complete analysis of the feedback difference amplifier, which should show the general conditions for permissible feedback, is not known. We can state, however, that feedback is always permissible if, after the first stage, no stages with a common element occur, and therefore have a discrimination factor of 1. Fig. 28-23 gives an example of a two-stage amplifier where this condition is satisfied with switch S in position a. The amplification will then be about 1000. Fig. 28-24 shows a comparison of the results for the rejection factor H_a of this amplifier having various valves in the first stage, with the rejection factor H_b of the amplifier, where the second stage has a common cathode resistance (S in position b). Although incidentally a higher value can be obtained in the latter case, the guaranteed minimum value is considerably larger in the first case.

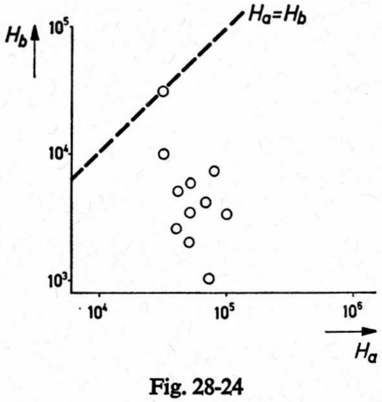

Fig. 28-24

We shall conclude our discussion of difference amplifiers with valves by mentioning some more advantages. One of the greatest is undoubtedly that an amplifier which consists of a number of difference stages, shows hardly any tendency to oscillate, despite large amplification. With unbalanced amplifiers this is usually not the case, as we can see if we examine a three-

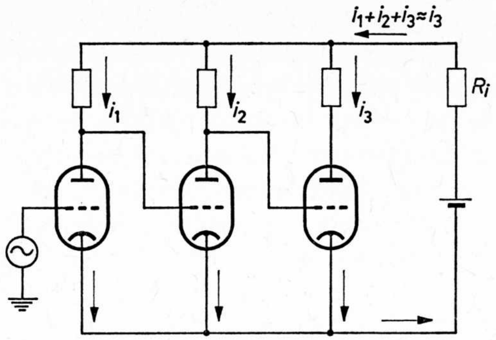

Fig. 28-25

stage amplifier as shown in Fig. 28-25, the signal current i_3 in the third stage will be relatively large due to the amplification in the first two stages, and cause a voltage across the internal resistance of the supply source.

This voltage will be directed to the grid of the second stage via the anode resistance of the first stage and then amplified, so that a feedback system exists via the supply, which can easily give rise to oscillation because of the large amplification. Complications such as these can only be avoided by a careful decoupling of the supply voltages for the various stages.

With a good difference amplifier stage (schematically represented in Fig. 28-26), the signal current through one half is equal, but opposite, to the current passing through the other half, i.e. if balance is complete, no signal

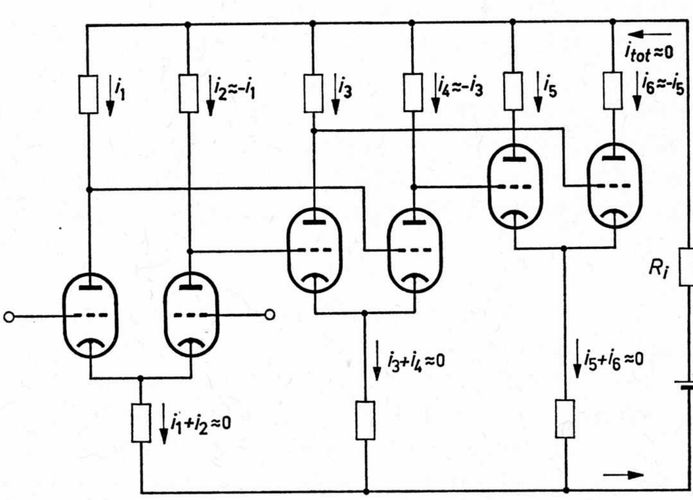

Fig. 28-26

current will pass through the supply source. If, because of inbalance, a small signal voltage is still generated across the internal resistance of the supply, it will be fed to the stages symmetrically and thus only cause a small signal current. In this respect the use of difference stages in an amplifier gives a quadratic improvement. This makes it possible to obtain amplifications of 10^7–10^8 with difference amplifiers without taking excessive precautions. This is why in many cases, where a single-sided amplifier would suffice for the actual amplification, a good difference amplifier is nevertheless used.

Regarding the influence of changes in supply voltages, it should be noted that these too are entirely determined by asymmetries of the stages. As far as the effect via anode and cathode resistances is concerned, an equal variation of the positive and negative supply voltages is equivalent to a common signal at the input terminal. Since the circuit is designed to enable the effect of this signal to be reduced even under the most unfavourable circumstance by the minimum rejection factor, this will also apply in the case of these changes in the supply voltage. When deriving other auxiliary voltages from these supply voltages (e.g. for feeding screen grids when pentodes are used, or the grid of the current source valve in the cathode circuit), one should ensure that changes introduced via these paths are sufficiently small. It is usually easy to meet this requirement.

It is also important for d.c. and low-frequency balanced amplifiers that the effect of heater voltage changes are smaller by one order of magnitude than is normal for single-sided amplifiers. Quantitative data will be given when discussing d.c. amplifiers in Section 35.

Let us finally note that because of the large cathode resistance, a good difference amplifier stage is more suitable for the processing of large signals than a simple long-tailed pair circuit, because of the large cathode resistance. As soon as one of the two halves is cut off, the current passing through the other half will remain almost constant. As we have mentioned, when discussing the long-tailed pair, this is a desirable property for certain limiting circuits.

With regard to the capabilities of transistorized difference amplifiers we should primarily state that the above considerations can be applied without alteration to circuits using field effect transistors. Because the characteristics S and μ of these elements have values of the same order of magnitude as those of a triode, the same precautions should be taken to obtain a good rejection factor, i.e. by increasing the resistance in the common cathode (source) circuit, and increasing the effective μ by using a cascode circuit.

When ordinary transistors are used, a complication occurs because of the base current. This not only introduces differences in the division between base and collector current, but also in the voltage drop across the possibly unequal resistances in the base circuits.

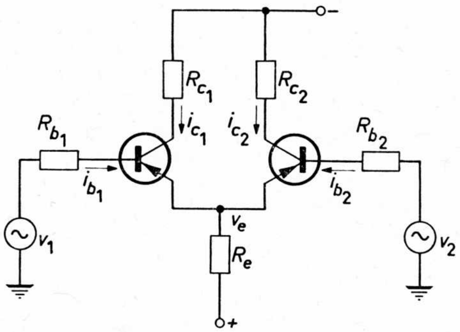

Fig. 28-27

We can illustrate what can be expected in this case by calculating the rejection factor for the circuit of Fig. 28-27. We shall simplify the calculation by neglecting the influence of the collector voltages and start with the equations $i_c = Sv_{be}$ and $i_b = i_c/\alpha'$. We thus obtain for the circuit:

$$i_{c1} = S_1(v_{b1} - v_e)\,,\; i_{c2} = S_2(v_{b2} - v_e)\,,$$

$$i_{b1} = \frac{i_{c1}}{\alpha_1'},\quad i_{b2} = \frac{i_{c2}}{\alpha_2'},\quad v_{b1} = v_1 - i_{b1}R_{b1}\,,\; v_{b2} = v_2 - i_{b2}R_{b2}$$

and

$$i_{c1}\left(1 + \frac{1}{\alpha_1'}\right) + i_{c2}\left(1 + \frac{1}{\alpha_2'}\right) = \frac{v_e}{R_e}.$$

After elimination of various quantities we obtain for i_{c1} and i_{c2}:

$$\begin{cases} i_{c1}\left\{\dfrac{1}{S_1} + \dfrac{R_{b1}}{\alpha_1'} + R_e\left(1 + \dfrac{1}{\alpha_1'}\right)\right\} + i_{c2}R_e\left(1 + \dfrac{1}{\alpha_2'}\right) = v_1 \\[2ex] i_{c1}R_e\left(1 + \dfrac{1}{\alpha_1'}\right) + i_{c2}\left\{\dfrac{1}{S_2} + \dfrac{R_{b2}}{\alpha_2'} + R_e\left(1 + \dfrac{1}{\alpha_2'}\right)\right\} = v_2 \end{cases}$$

We can now calculate $i_{c1} - i_{c2}$ for $v_1 = v_2 = e$ and for $v_1 = -v_2 = e$. Thus, for the rejection factor relating to the collector currents this gives:

$$H = \frac{4\ R_e}{\Delta\left(\frac{1}{S} + \frac{R_b}{a'}\right)}$$

if $R_e \gg 1/S$, R_b/a' and using the same notation as before for the average values of the corresponding parameters.

The rejection factor for the collector voltages is easily derived from the above expression by using the equation: $v_{c1} - v_{c2} = -(i_{c1}R_{c1} - i_{c2}R_{c2}) = i_c\Delta R_c + R_c\Delta i_c$. For a difference voltage $v_1 = -v_2 = e$ only the second term (which has already been calculated) is important, so that only with a common signal an extra term is obtained which takes into account the inequality of the collector resistors. This means that the denominator in the above expression for H will be increased by a factor 1–2.

This new expression for H shows that not only does the term $\Delta\ (1/S)$ appear but also $\Delta(R_b/a')$, which will already be of importance at relatively low values of the resistances R_b in the base circuits. Typically the current amplification factors can show a great spread, and this term will supply a relatively large contribution, even under conditions of matched base resistances. The minimum value of the above expression is obtained when one of the resistances R_b tends towards zero and the other is divided by the smallest possible a'. When neglecting $\Delta(1/S)$ this yields:

$$H_{\min} = \frac{4\ R_e\ a'_{\min}}{R_b}$$

Example: In the circuit of Fig. 28-27 we use transistors which have an a' value between 20 and 50. R_e is 0.5 MΩ. If one of the signal source resistances equals zero and the other 10 kΩ, this gives $H_{\min}=4{,}000$. At a current of 1 mA: $1/S \approx 30\Omega$; $\Delta(1/S)$ is therefore less than 5 Ω, which proves that this term can indeed be neglected. If both resistances in the base circuits are 10 kΩ, the term $\Delta(R_b/a')$ cannot be more than $10^4/20 - 10^4/50 = 300\ \Omega$; this gives $H = 2 \cdot 10^6/300 = 6.700$.

In reality, a collector feedback effect also occurs with the transistor. As for the valve circuit, its effect on the value of H will again be of the order of magnitude $\Delta\mu/\mu^2$ (see equation 28.5). For transistors of the alloy-diffusion type, μ is more than 1000, and $\Delta\mu/\mu^2$ will therefore be 10^{-4} or smaller.
It therefore follows that in order to obtain a rejection factor of the order of magnitude 10^4 with a transistorized difference amplifier, it will usually be sufficient to increase only the common emitter resistance artificially, thus producing a circuit as shown in Fig. 28-28.

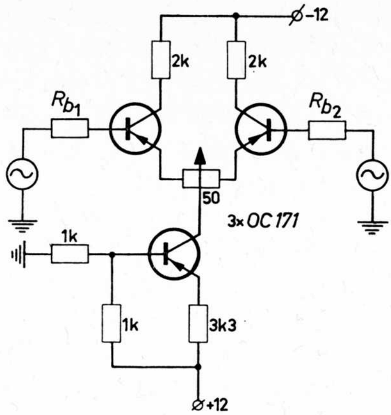

Fig. 28-28

With $R_{b1} = R_{b2} = 0$, the measured values for the rejection factor of this stage with different specimens of OC171 varied from $7 \cdot 10^4$ to $2 \cdot 10^5$. With $R_{b1} = 0$ and $R_{b2} = 10\ \text{k}\Omega$, a value of approx. $5 \cdot 10^3$ was measured in all cases.

The rejection factor will remain large even at high frequencies (tens of kc/s) because of the small parasitic capacitances occurring in transistor circuits. The simple transistorized difference amplifier therefore gives a considerably better performance than that with valves for the measurement of signal sources with relatively low internal resistances, provided transistors of the alloy-diffusion type are used.

Whilst the use of cascodes with transistors is almost never necessary because of the high value of the amplification factor, the need for an increase in α' is sometimes felt, particularly to reduce the base currents of the first stage and hence the load on the signal sources. Fig. 28-29 shows two ways of achieving this. In the left-hand figure we use two *p-n-p* transistors, and

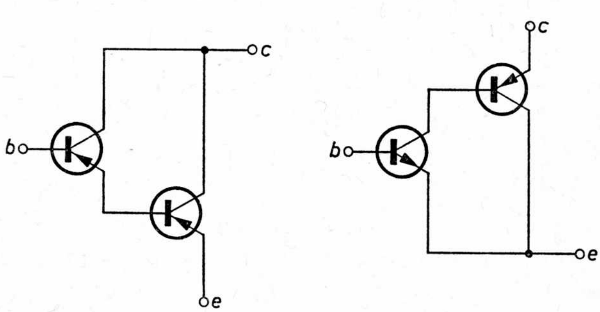

Fig. 28-29

in the right-hand figure a combination of an *n-p-n* and *p-n-p* transistor. The current amplification factor α_{eff} is in both cases approximately equal to the product of the current amplification factors of the individual transistors. The amplification factor of this combination is approximately equal to that of the transistor with the smallest value for μ. Such a combination of two transistors is known as a "Darlington pair".

Because of the small physical dimensions and particularly because of the absence of a heater supply, it is easier to realize the design of a "floating" amplifier with transistors than with valves. However, this kind of solution is still rather complicated and will only be applied in extreme cases. An additional facility given by transistors is the use of alternately *n-p-n* and *p-n-p* transistors in successive stages, which makes it possible to keep the supply voltages relatively low, even with direct-coupled stages. Fig. 28-30 shows such a circuit.

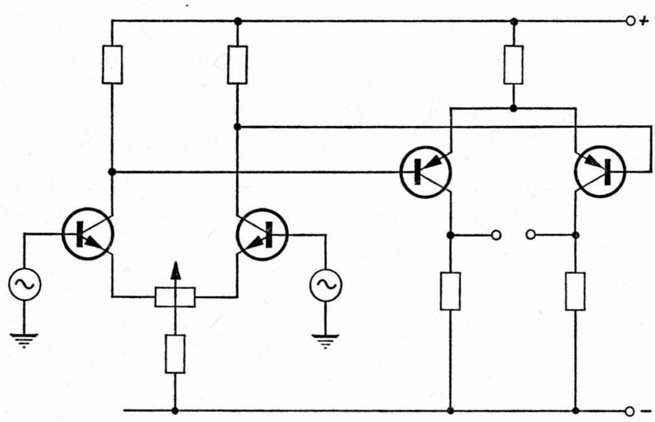

Fig. 28-30

With regard to applications, the effect of supply voltages and suchlike, everything we have said about valve circuits is also completely valid for transistorized amplifiers. Finally, we should note that also regarding the design of transistor difference amplifiers, many misconceptions seem to exist in the literature. In particular the concepts of discrimination and rejection factors are often confused. Many designers appear to seek a design solution by the careful adjustment of certain components. This is only justified when one can guarantee that possible changes in these components have no great effect on the characteristic parameters, at least not during the time of measurement. Often this is not the case, with the result that, for example, the guaranteed value of the rejection factor is one or several orders of magnitude smaller than the first measured value after initial adjustment of the components and which is often indicated as "typical".

29. Power supplies

So far we have thought of the d.c. supply for valves and transistors in terms of accumulators or batteries. It is, of course, possible to use these sources for equipment supplies. This practice is, however, objectionable for apparatus demanding higher power, if only because of the bulk and weight of these power sources, and the necessity for continuous maintenance and replacement. Moreover, even when we consider these sources from a purely electronic point of view, they have awkward deficiencies. For example, the supplied voltage is only reasonably constant for short periods, and may drop considerably over longer ones. Further the internal resistance is also often rather high. For these reasons wherever a mains supply is available it is to be preferred, but it is then necessary to take certain precautions. Nowadays the U.K. mains supply is nearly always an a.c. voltage of a frequency of 50 c/s and a rms value of 220/240 V. But rather considerable fluctuations (up to approx. 10 per cent) can occur in this amplitude because of load variations. By applying some of the principles discussed, it is, however, possible to obtain d.c. voltages of good constancy and with a very small internal resistance from the a.c. mains supply. The usual practice is first to rectify the a.c. mains voltage into a "raw" d.c. voltage, with valve or semiconductor diodes, and then to "smooth" this voltage with filters and electronic means until the desired stability is achieved. We shall now discuss these various processes.

The rectifying action of diodes is illustrated by the circuits of Figs 29-2--6, where the current-voltage characteristics of valve and semiconductor diodes are approximated by the idealized characteristic of Fig. 29-1. The series circuit of the diode and resistor R (Fig. 29-2) is connected to a d.c. voltage source V_o. The voltage is then divided between diode and resistor as shown on the right. If V_o is replaced by an a.c. voltage source V_b, the voltage V_R across the resistor will be as is shown in Fig. 29-3. With V_R we have obtained a voltage which has indeed a d.c. component, but also an unwanted large a.c. voltage component. By replacing resistor R by a capacitor C and connecting the circuit to a d.c. voltage source (Fig. 29-4), the full voltage V_o will at first be present across the diode, and a large current will flow which charges the capacitor. How rapidly this happens in practice is not only determined by the slope of the diode characteristic and the value of the capacitor, but also by the series resistance which is still present. When a diode of the wrong rating is selected, it may be damaged by a surge current which is too large. If this

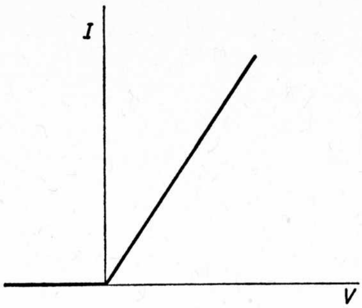

Fig. 29-1

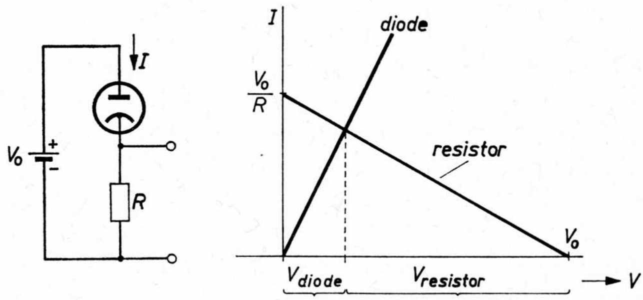

Fig. 29-2

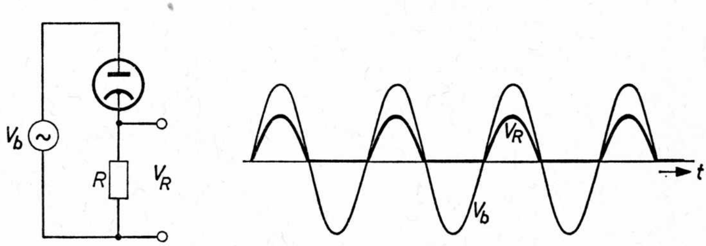

Fig. 29-3

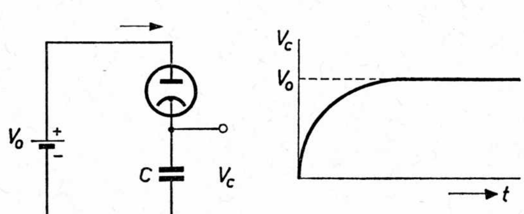

Fig. 29-4

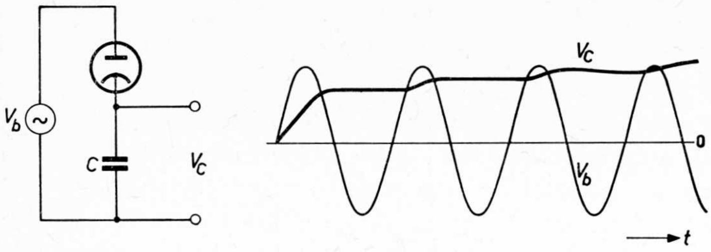

Fig. 29-5

is not so, the voltage across the diode, and therefore also the charging current will decrease with increasing voltage V_c. In the equilibrium state, in the absence of leakage currents, the total voltage will be present across the capacitor and the charging current will be zero. Replacing V_o by a.c. voltage V_b (Fig. 29-5) results in the presence of a charging current, but only during the periods when the voltage across the diode is positive. These periods become increasingly shorter and the charging current smaller, until in the equilibrium state (once again in the absence of leakage currents) the voltage over the capacitor will have become equal to the peak value of the a.c. voltage. We have thus obtained a d.c. voltage from an a.c. voltage.

We can now use the capacitor voltage as a supply for loads, so that current will be consumed. This can be taken into account by inserting a resistor in parallel with the capacitor (Fig. 29-6). The behaviour of the output

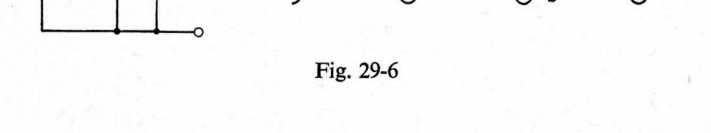

Fig. 29-6

voltage V_c will lie between those described in Figs 29-3 and 29-5, and whether it will correspond more to one of these than the other depends on the ratio of the resistance to the impedance of the capacitor. During the periods that the a.c. voltage is higher than the capacitor voltage, the capacitor will be charged; during the remainder of the time the capacitor will discharge through the load resistance. The d.c. voltage level of the capacitor will have such a value V_g that for each period the charge and discharge is equal. Both this value V_g and the amplitude V_r of the "ripple voltage", i.e. the a.c. voltage component of V_c, will depend on the slope of the diode characteristic, and the values of R, C, V_b and the frequency f of the latter. The value of V_r can in practice be closely approximated by $V_r \approx I/2fC$ where I is the d.c. current load. The charge taken from C during each period is approximately equal to I/f, thus producing a change in voltage of I/fC. The amplitude V_r is about half of this. This shows that the only value which can still be selected in a given situation, in order to reduce the value of V_r, is the value of the capacitor. This value is, however, restricted by the permissible peak charging current during turn-on.

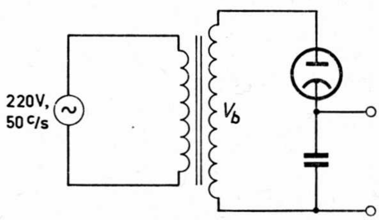

Fig. 29-7

The required value of V_g is achieved by deriving V_b from the mains by means of a transformer. An additional advantage is that the entire circuit is electrically isolated from the mains supply and therefore can be used at any desired d.c. voltage level. This circuit is shown in Fig. 29-7.

The fact that the current-voltage characteristics of valve and semiconductor diodes deviate slightly from Fig. 29-1 does not make much difference to the above considerations, but there are important quantitative differences between the various types. The vacuum diode has a rather high internal resistance, which gives relatively great voltage losses at high currents. For this reason, one often used in the past valves filled with mercury vapour, which exhibit a similar effect to that obtained with voltage stabilizer valves (p. 177). The voltage across the valve is practically independent of the current beyond a certain value, so that the internal resistance of the valve is small. Because of the use of mercury vapour, the voltage drop is very low (approximately 8 volts), consequently the voltage loss at high currents will be less than that with vacuum diodes.

From its inception the semiconductor diode has superseded the vacuum diode for the rectification of low voltages because of its small internal resistance and high permissible currents. This is now also the case for high voltages. One of the great additional advantages of this type of diode is the absence of a filament or heater which often necessitated a separate winding on the supply transformers. As a rectifier element the vacuum diode nowadays only has advantages for voltages exceeding 1 kV. The maximum permissible currents vary from a few milli-amps to several tens of amps with the usual types of semiconductor diodes.

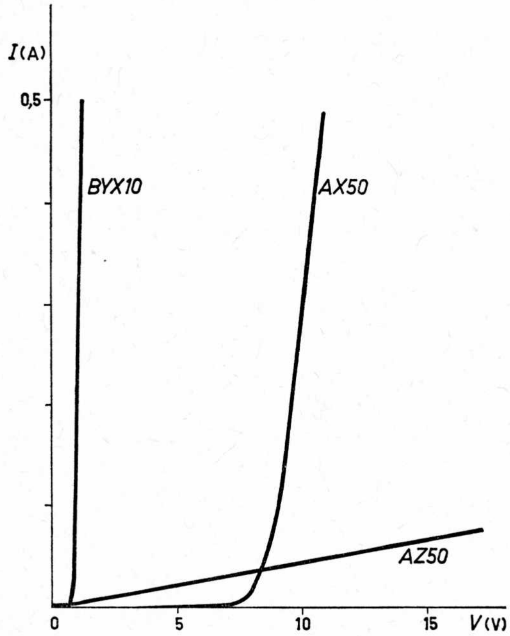

Fig. 29-8a

We can illustrate the differences between valve and semiconductor diodes by showing (Fig. 29-8a) the characteristics of a vacuum diode (AZ 50), a valve diode with mercury vapour (AX 50) and a semiconductor diode (BYX 10); all three are capable of supplying a current of approximately 200 mA at 300 volts. The length and diameter of both the AZ 50 and the AX 50 are 130 and 50 mm respectively, those of the BYX 10,7 and 3.5 mm. We can use a capacitor of 100 μF with the BYX 10 with a total series resistance of only 15 Ω; for valves this value is approximately 25 μF and 100 Ω. Fig. 29-8b indicates the ratio $V_g : V_{b,pk}$ against the load current for the BYX 10 and the AZ 50 at these values, certain proof of the great advantage obtained by the use of semiconductor diodes.

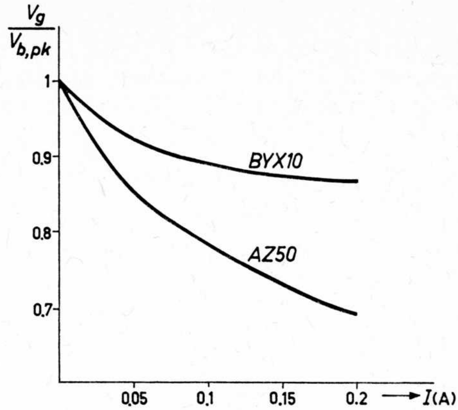

Fig. 29-8b

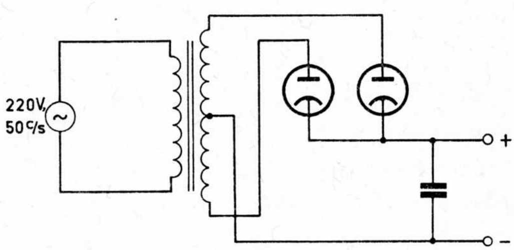

Fig. 29-9

The amplitude of the ripple produced by the half-wave rectifier circuit of Fig. 29-7 is often undesirably large. However, by using full-wave rectifiers (Fig. 29-9), where the capacitor is charged twice in each cycle, the amplitude of the ripple is practically halved, and the ripple frequency is twice that of the mains supply. This circuit has the additional advantage that each diode now only passes half the load current.

Apart from these two basic circuits, there are various other ways of rectification. The upper part of Fig. 29-10 shows the single-phase full-wave bridge, which is a bridge arrangement of the circuit of Fig. 29-9. When valves are used, this circuit has the disadvantage that the four cathodes have three different d.c. voltage levels, which usually requires three separate filament windings. Semiconductor diodes do not have this drawback and the circuit has the great advantage that the maximum voltage across each diode in the reversed direction is about half of that in Fig. 29-9. The lower

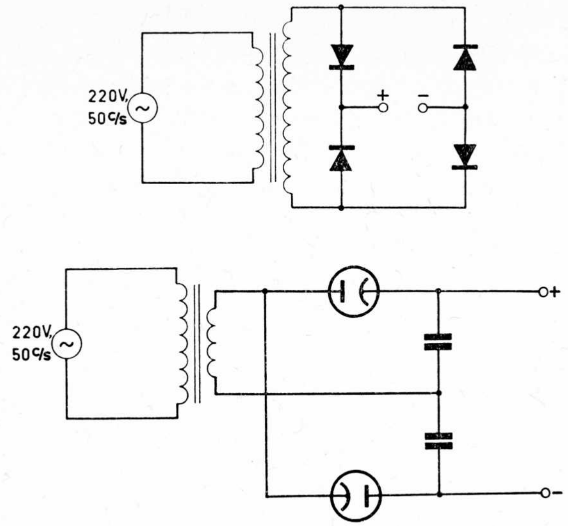

Fig. 29-10

part of Fig. 29-10 gives an example of a circuit for doubling the voltage; here the d.c. voltage is twice the amplitude of the a.c. voltage for the limiting case when no current is taken.

A considerable reduction in the value of the ripple voltage is obtained by using a filter consisting of a choke L and a second capacitor C_2 (Fig. 29-11). Whilst the d.c. voltage component of V_o is almost the same as that of V_c (because of the d.c. resistance of the choke L there is some loss in d.c. output voltage when loaded), the component of the ripple voltage with frequency ω is reduced by a factor of approximately $\omega^2 LC_2$. The value ω is approximately 300 and 600 for half-wave and full-wave rectification respectively, and factor $\omega^2 LC_2$ can therefore easily be made much larger than unity. Since this reduction increases with the square of the frequency, the components of higher frequencies will be reduced still further, so that the remaining ripple voltage on V_o will be almost sinusoidal.

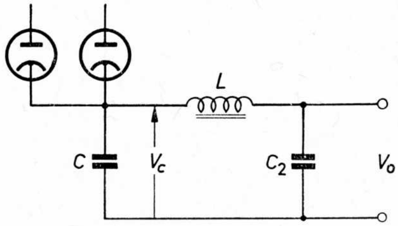

Fig. 29-11

With a vacuum diode circuit, values of $L = 6\ H$ and $C_2 = 25\ \mu F$ can be easily realized for 300 V at 150 mA full load. This gives $\omega^2 LC_2 = 60$ for full-wave rectification. For a circuit for supplying transistor circuits with 10 V at 2 A one can select $L = 0.1\ H$ and $C_2 = 1000\ \mu F$ sothat the corresponding value for $\omega^2 LC_2$ is then 40.

In theory it is possible to filter the output voltage once again. Attention should then be paid to the correct positioning of the second choke as otherwise the extraneous field of the transformer induces a "hum" voltage which may easily exceed the value of the residual ripple. It is therefore often more efficient as well as cheaper to apply one of the types of simple electronic stabilization which are to be discussed.

It is often inadmissible, particularly in measurement electronics, that any change in the mains voltage results in a correspondingly large change in the d.c. supply voltage, as was shown to be the case in the circuits discussed above. It is also often desirable for the internal resistance of the supply source to be considerably smaller than can be achieved with the circuits of Figs. 29-9, -10 and -11. At first we shall restrict ourselves to power supplies for valve circuits. The internal resistance of these circuits will be of the order of magnitude of 1 kilo-ohm. A great improvement in voltage stability as well as a small reduction in internal resistance can be obtained by using voltage stabilizing tubes. As we have seen on p. 177, the operating voltage of these valves is almost independent of the current over a given range. The differential resistance $\mathrm{d}V/\mathrm{d}I$ will ordinarily amount to several hundred ohms. When this type of valve is used in the circuit of Fig. 29-12, the changes in the "stabilized" output voltage can be 10 to 50 times smaller than those in the non-stabilized voltage, depending on the load and the value of R_1. This system is therefore efficient in simple cases for a small load, but is often insufficient because the internal resistance is still several hundred ohms, and the maximum allowed current variations will be rather small, for most valves in the neighbourhood of 10 mA.

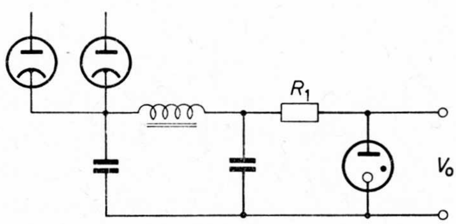

Fig. 29-12

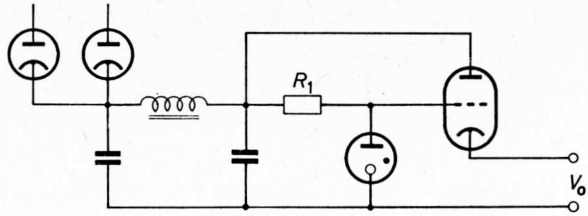

Fig. 29-13

Greater improvement is achieved by adding a series triode or pentode to the circuit of Fig. 29-12 to become that of Fig. 29-13. The allowed current changes are then no longer limited by the stabilizing valve and R_1, but by the triode, thus permitting a far greater freedom. Furthermore, since the load current no longer passes through R_1, the latter may be given a higher value, so that variations in the non-stabilized voltage will have less effect on the stabilizing valve voltage. On the other hand, these variations will now also affect the output voltage via the anode of the triode. The internal resistance of this voltage supply equals the cathode output impedance of the series valve. Because the impedance in the anode circuit is small, the output impedance will be almost equal to $1/S$, which means a value of less than 100 Ω when using a power valve with a slope of 10 mA/V or more. Here, in fact, we use the low output impedance of a cathode follower, which follows the voltage of the stabilizing valve, and this shows the way to a further reduction in the output impedance. If we use instead of the cathode follower an amplifier with feedback of its output voltage, the output impedance will be reduced by a factor equal to the loop gain. In a simple circuit design, this leads to the arrangement of Fig. 29-14, where a fraction k of the output

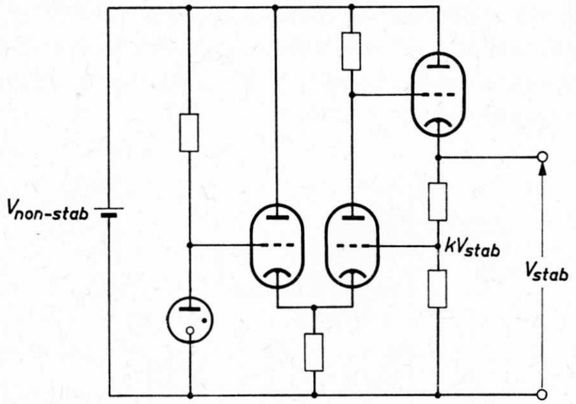

Fig. 29-14

voltage is compared to the reference voltage of the stabilizing valve. This is done here by means of a difference stage. When using a two-stage amplifier and a few obvious improvements, such as the use of the stabilized voltage to supply the stabilizing valve, highly stable d.c. voltage sources of several hundred volts with a very low internal resistance ($<0.1\ \Omega$) can easily be obtained.

Almost all power supplies of this type can be reduced to the basic scheme of Fig. 29-15. We find for the changes in the values, when neglecting the very small changes in the current through the voltage divider:

$$i_{load} = i_b = S(v_g - v_{stab}) + \frac{S}{\mu}(v_{non\text{-}stab} - v_{stab})$$

and

$$v_g = -Akv_{stab}$$

It follows for $Ak \gg 1$ and $\mu \gg 1$:

$$v_{stab} = \frac{v_{non\text{-}stab}}{(Ak+1)\mu} - \frac{i_{load}}{(Ak+1)S} \qquad (29.1)$$

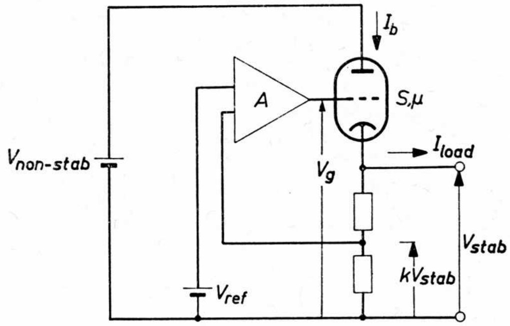

Fig. 29-15

Changes in the non-stabilized voltage are thus reduced by a factor $(Ak+1)\mu$, and the internal resistance R_i of the voltage source is $1/(Ak+1)S$. However, it must be remembered that the reduction factor $\mu(Ak+1)$ only applies to the effect of the raw non-stabilized voltage connected to the series valve's anode. For example, if the non-stabilized voltage is also used to supply the amplifier or the reference valve, changes can also affect the stabilized voltage via these paths. It is possible, however, to render their effect sufficiently small, as will be shown in the circuits of a few power supplies.

With power supplies we often distinguish between "stabilization" and "regulation". The first term indicates the degree of insensitivity of the output voltage against fluctuations in the voltage of the mains supply, whilst regulation indicates the effect of changes in load. In the above discussion, values $Ak\mu$ and AkS are therefore measures of the stabilization and regulation respectively.

By making the loop gain Ak sufficiently large, we can reduce the effect of changes in the non-stabilized voltage and load current to such an extent that the degree of constancy of the stabilized voltage will be mainly determined by changes in the reference voltage and the control amplifier. As we shall see when discussing d.c. amplifiers, only a little care is necessary to restrict the zero drift of a balanced valve amplifier to a few millivolts, referred to the input terminals. Special stabilizing valves have been designed which possess a good constant reference voltage; they are not primarily meant for stabilizing the voltage at a strongly varying current, but for producing at an almost constant current a potential which is little affected by ambient temperature and duration of operation. Examples are the 85A2 and 83A1 valves, both with reference potentials of about 85 volts, having a relative change with temperature of only approximately $3 \cdot 10^{-5}$, which can be partly compensated by means of a temperature-sensitive resistor. The differential resistance of these valves is 100–400 Ω. It is easy to keep the relative change in reference potential at not over-great changes in the ambient temperature to within 10^{-4} by supplying these valves from the stabilized output voltage (assumed to be larger). Since a good balanced amplifier contributes little to the variations, this is also the value at which a power supply of several hundred volts can be kept constant without much difficulty. Ageing of the reference valve may cause these variations to be exceeded slightly in the course of weeks or months.

Increasing the stability by a further order of magnitude demands considerable effort. But even the most sensitive measuring instruments, provided they are well designed, only rarely need such a high degree of power supply stability.

The stabilization and regulation obtained by the simple circuit of Fig. 29-14 will often be sufficiently large, and no second stage is required in the control amplifier. A still greater simplification is shown in the left-hand side of Fig.29-16. The current passes through both the amplifier triode and the stabilizing valve. However, because of the great effect the fluctuations in the supply voltages have in the case of a single-sided amplifier, this type of circuit only yields very moderate stabilization. Improvement can be achieved by replacing the supply resistor R connected to the non-stabilized voltage by a transistor current source connected to the stabilized voltage (right-hand side of Fig. 29-16). The grid-cathode bias of the series valve must here be at least 4 volts.

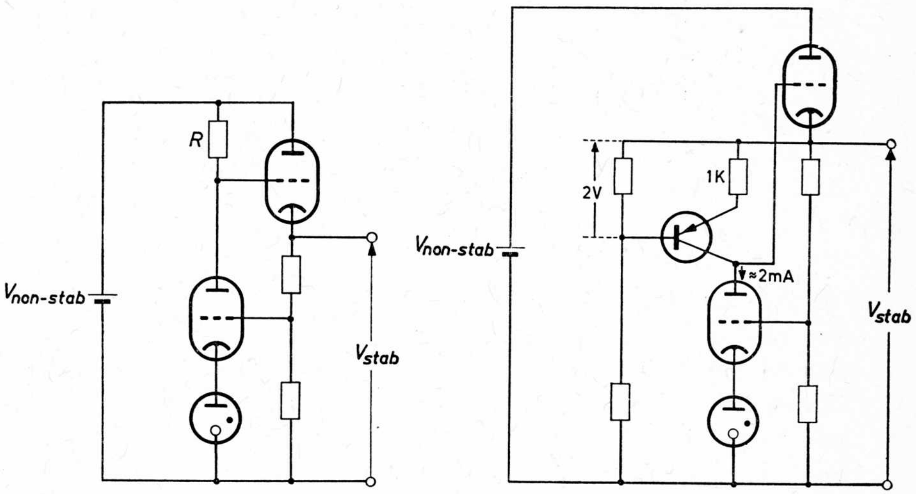

Fig. 29-16

When designing a stabilized power supply, we should pay attention to its transient behaviour, i.e. to the frequency characteristic of the system. Rapid changes will not occur in the non-stabilized voltage owing to the presence of the large smoothing capacitor and (possible) filter, but they may occur in the load current. In almost all cases it is possible to regard the load as the parallel combination of an variable resistor and a capacitor; and the effect of this follows from two different approaches.

Firstly, the supply source plus the load, is considered as a feedback amplifier (Fig. 29-17). The load $Z_o = C_o//R_o$ is then connected in series with the output impedance $1/S$ of the amplifier without feedback. This yields in the loop gain the factor

$$\frac{Z_o}{Z_o + \dfrac{1}{S}} = \frac{1}{1 + \dfrac{1}{SZ_o}} = \frac{1}{1 + \dfrac{1}{SR_o} + \dfrac{C_o p}{S}}$$

where the time constant $(C_o/S)/(1 + 1/SR_o)$ appears. $R_o \gg 1/S$ in all practical cases, so that this constant will be C_o/S. If the drop in amplification of the control amplifier A corresponds to a time constant τ_v, we have a feedback

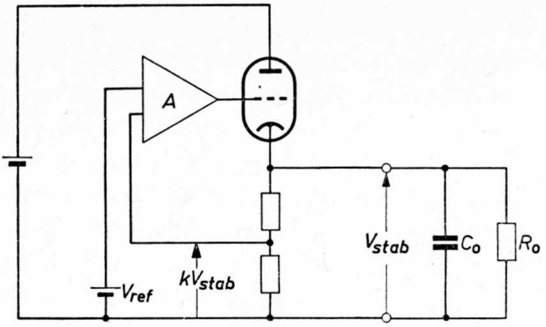

Fig. 29-17

system with two time constants as discussed in Section 22. Thus we find for a flat frequency characteristic: $2\,A_0k_0 \leqslant C_0/S\tau_v + S\tau_v/C_o$ and for a step function without overshoot A_0k_0 must be twice as small. Therefore, if we want to use a large loop gain, the ratio of C_o/S to τ_v must be large. It makes a very important difference to the dynamic properties of the supply which time constant is made the largest. If τ_v is made large (and therefore the amplifier slow) and C_o/S small a fast change in the load current will be quickly followed by the output voltage because no charge reserve is present. The slow amplifier will have a delayed reaction to this, so that a great change in voltage will occur in the first instance, which will only be compensated for a little later on (Fig. 29-18). The final accuracy is high, but the power supply requires a recovery time. On the other hand, when we give the greater value

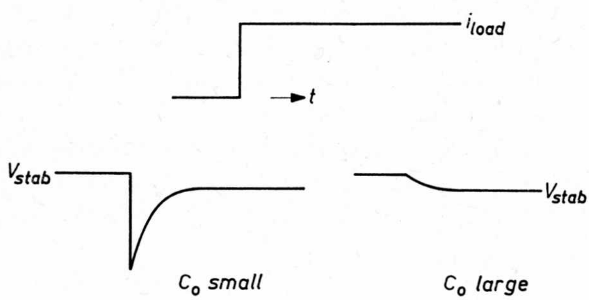

Fig. 29-18

to C_o by arranging additional output capacitance in the device, and keeping τ_v small, this additional capacitance will contain a great charge reserve. Any change in the load will moreover be compensated very quickly by the amplifier; no recovery time occurs in this case.

The other way of looking at it is as follows: the output impedance of the supply source is $1/(Ak+1)S$ and therefore consists of the parallel combination of $1/S$ and $1/AkS$. Because $Ak = A_0k_0/(1+j\omega\tau_v)$, this last impedance becomes $(1+j\omega\tau_v)/A_0k_0S$ and may therefore be considered as the series combination of a resistance $1/A_0k_0S$ and an inductance of the value τ_v/A_0k_0S. The supply can thus be represented by the circuit on the left-hand side of Fig. 29-19. When connecting a partially capacitive load, a tuned circuit will be produced which may exhibit strong resonances (see Section 33).

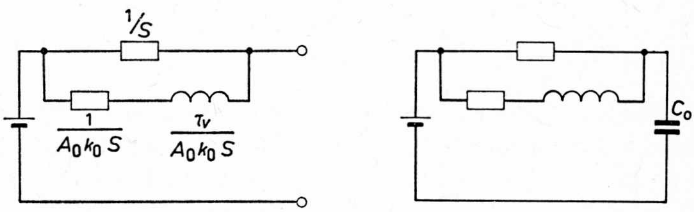

Fig. 29-19

This can be avoided without increasing the series resistance, by making C_o so great that the circuit is damped "critically". This approach to the problem has the advantage that it is easily seen that even at smaller C_o values sufficient damping can be obtained by inserting a small resistor in series. If series damping is not applied, we also achieve the result that the output voltage will be vibration-free for

$$C_o \geqslant 4A_0k_0S\tau_v$$

From this condition and the relation $R_i = 1/A_0k_0S$ for the internal resistance of the supply for slow changes, we derive the rule of thumb $C_o \geqslant 4\tau_v/R_i$. With a multi-stage control amplifier, the drop in amplification at high frequencies will usually be larger than that which corresponds to a single time constant. It is often sufficient in this case to make C_o a few times larger than would follow from the above condition, taking $\tau_v = 1/\omega_g$, where ω_g is the frequency for which amplification has dropped $\sqrt{2}$ times.

Example: a stabilized supply of 300 volts is designed with valve EL34 as the series valve. The internal resistance must be 0.1 Ω at the maximum load current. The 85A2 is selected as voltage reference valve. We wish to know the values required for A_o and C_o.

At a current of 100 mA, the slope of the EL34 is approx. 10 mA/V. $1/S$ is therefore 100 Ω, and a reduction of 1000 is necessary to arrive at 0.1 Ω. Therefore: $A_0k_0 = 1000$. With $k_0 = 85/300 = 0.28$, this gives $A_0 = 3{,}500$. We can achieve this with a two-stage amplifier consisting of pentode or cascode stages in a balanced form. The frequency characteristic of such an amplifier may have a flat behaviour up to 50 kc/s. τ_v is then smaller than $1/2\pi \cdot 50 \cdot 10^3 = 3\ \mu$s and for C_o we thus obtain:

$$C_o > \frac{4 \cdot 3 \cdot 10^{-6}}{0.1} = 120\ \mu\text{F}.$$

If we have the general case of a supply with a power capability of P watts at a voltage V, and it is required that the change from zero to full load gives a change in voltage not exceeding a per cent, we have $P/V \cdot R_i = 10^{-2}$ V and therefore, by means of the above-mentioned rule of thumb:

$$C_o > \frac{4 \cdot 10^2 P \tau_v}{a V^2}.$$

P has a value between 10 and 100 W for most supplies used in measurement electronics, and the required capacitance is inversely proportional to V^2 for a given power. The electrolytic capacitors necessary for reaching extremely high values of C_o at low voltages ($> 10^4 \mu$F) possess a rather large series resistance. At high frequencies, this resistance, as well as the inductance of the connecting cables, will determine the impedance of the supply. The equation also shows that it is advantageous to make the amplifier as fast as possible.

Example: Making $P = 25$ W, $a = 0.1$ and $\tau_v = 10\ \mu$s, gives:

$$C_o \geqslant 1/V^2$$

This means for a 6 V, 4 A device: $C_o \geqslant 25{,}000\ \mu$F; for a 250 V, 100 mA device: $C_o \geqslant 16\ \mu$F, and for a 2,000 V, 12 mA device: $C_o \geqslant 0.25\ \mu$F.

A fixed relation exists between the stabilization factor $Ak\mu$ and the regulation factor AkS, namely μ/S, i.e. the internal resistance of the series element. The latter's value is usually so great that the regulation determines the loop gain for supplies of up to several hundred volts. If it is required that the change from zero to full load affects the output voltage to the same extent as the fluctuations in the non-stabilized voltage, putting these at 25 per cent of the stabilized voltage, we find:

$$\frac{0{,}25\, V}{Ak\mu} = \frac{P}{V} \cdot \frac{1}{AkS} \quad \text{or} \quad \frac{\mu}{S} = \frac{V^2}{4P}.$$

This gives for $P = 25$ W and $V = 300$ V: $\mu/S = 900\ \Omega$. With h.t. supplies, however, or supplies with a smaller load value, or smaller load variations, the stabilization factor may well determine the required loop gain. If in such a case it is desirable not to make the loop gain any larger than is necessary for good regulation, in connection with the value of the output capacitance, it is necessary to increase the internal resistance of the series element. This can be achieved either by incorporating a series pentode or cascode, or by pre-stabilization.

Example: A scintillation counter required a supply of 400–2,000 V, 5 mA. The requirements were such that $Ak\mu$ must be greater than 60,000 and AkS approx. 0.06 A/V, and therefore μ/S approx. 1 MΩ. By using

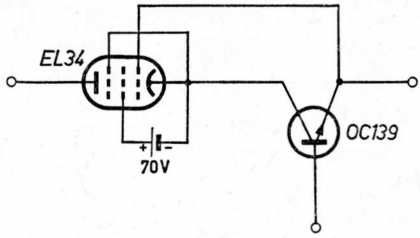

Fig. 29-20

a cascode consisting of a pentode and a transistor (Fig. 29-20), not only was this high internal resistance reached, but also very high values for S and μ:

$$S_{\text{casc}} = S_{\text{trans}} \quad \text{and} \quad \mu_{\text{casc}} = \mu_{\text{pent}} \cdot \mu_{\text{trans}}$$

S is already 0.03 A/V at a current of 1 mA, whilst the amplification factor exceeds 10^5 with valve EL34 and transistor OC139. Therefore, it is not necessary in this case to make the loop gain much larger than 1. With this cascode circuit, the grid cathode bias voltage is also the collector-emitter voltage of the transistor.

When considering the improvements brought about in the feedback amplifier with two time constants by introducing frequency-dependent feedback, we may wonder if a similar procedure would here be successful. However, it is easy to see that the gain will not be very great, because usually an important fraction of the output voltage is already fed back (10–25 per cent). Nevertheless, the relative resistor of the voltage divider is often parallelled by a capacitor in order to increase the a.c. loop gain and thus obtain a better elimination of the ripple or hum voltage. The value of this capacitor is therefore much greater than was proved necessary for use with amplifiers.

Fig. 29-21 shows an example of a stabilized supply for 350 V, 120 mA, where the above considerations have been taken into account. In order to make the control amplifier fast, the side of the second stage from which the signal is taken is designed as a cascode, thus reducing the Miller effect of this stage. Since such a supply will be designed to eliminate the effect of

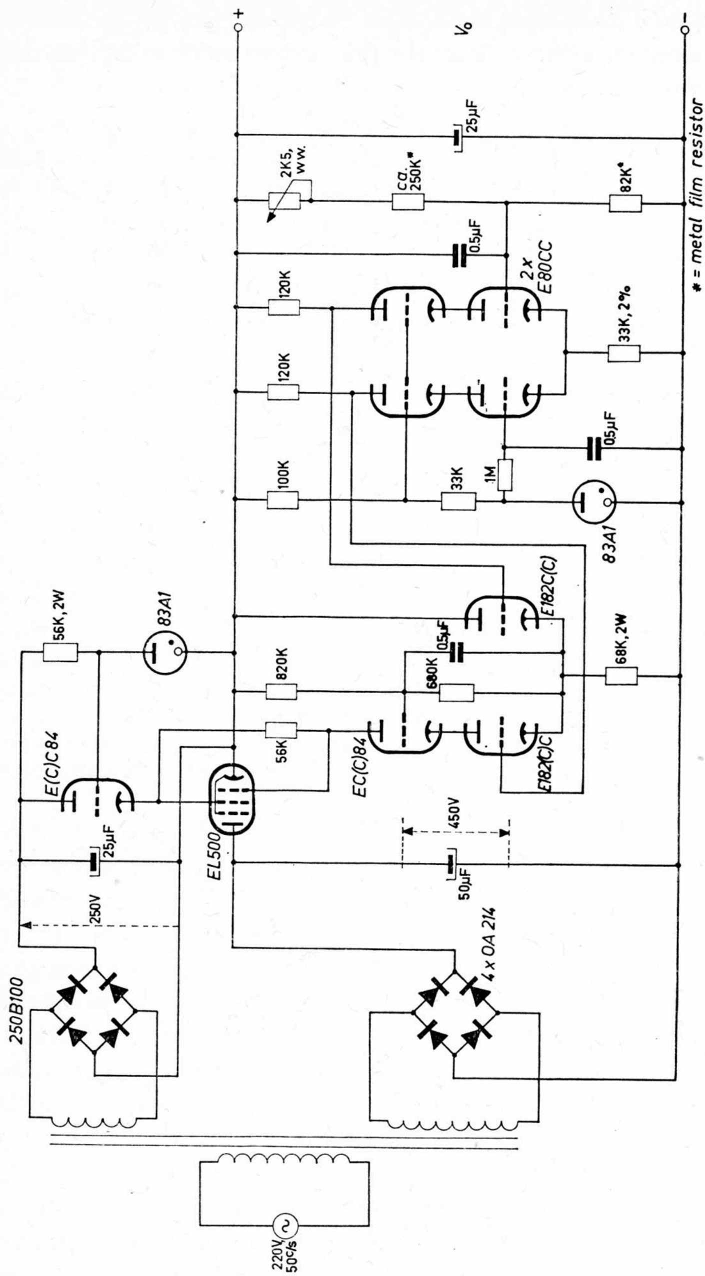
+
−
V_o
25µF
2K5, ww.
ca. 250K*
82K*
0.5µF
2 x E80CC
120K
120K
33K, 2%
0.5µF
100K
33K
1M
83A1
E182C(C)
56K, 2W
83A1
820K
0.5µF
680K
68K, 2W
56K
E(C)C84
EC(C)84
E182(C)C
EL500
25µF
450V
50µF
250V
250B100
4 x OA 214
220V 50c/s
* = metal film resistor

Fig. 29-21

large changes in the non-stabilized voltage, it is not necessary to filter this voltage first. The effect of fluctuations in the mains supply on the screen grid of series valve EL500 can be made smaller by separately stabilizing the screen grid voltage.

The operating voltage of gas discharge valves shows spontaneous fluctuations ("noise") in which all frequencies occur. For example, this noise has a maximum value of approx. 1 mV for valve 83Al in the frequency range between 30 c/s and 10 kc/s. The reference potential on the grid of valve E80CC will only show very slow variations if the noise is "smoothed" with an *RC*-filter with a time constant of approx. 1 second. Apart from temperature dependence, good reference valves then only show small random potential jumps (less than 1 mV with the 83A1), and a long-term voltage drift. The latter's value for the 83A1 is maximum 0.4 V or 0.5 per cent over the first 100 hours. During the next 1000 hours, the total excursion is less than 0.2 V.

It is in principle possible to earth one of the output terminals or to connect one of them to the terminal of a second power supply. However, we should remember that the secondary winding of the transformer, if the latter does not contain screening, may have a capacitance of several hundred picofarads to the normally one-sided earthed primary winding, so that a capacitive current of several tens of microamps will flow, which may produce a considerable hum voltage if an impedance to earth is present.

For the circuit of Fig. 29-21 the following results were measured:
Stabilization: When a non-stabilized heater supply is used for valve E80CC, a 10 per cent variation in the mains voltage will produce a change of ΔV_o in the output voltage, which will be less than 30 mV, i.e. < 0.01 per cent. With a sufficiently stabilized heater voltage, the minimum value of ΔV_o is determined by the effect of changes in the non-stabilized voltage of 450 volts. Here, a change of 10 per cent produces a relative change in V_o of less than $5 \cdot 10^{-6}$.
Regulation: For alternating currents, the internal resistance is approx. 0.03 Ω. For direct current and very slow changes, the internal resistance will be about 4 times larger, i.e. 0.12 Ω. Therefore, a change from zero to full load produces a change in the output voltage of approx. 15 mV, i.e. < 0.005 per cent.

Hum and ripple on the output terminals is less than 0.5 mV, peak to peak. The temperature dependence is determined by the reference element: $3 \cdot 10^{-5}$ per °C.

Sometimes a stabilized d.c. current source is required, e.g. for the supply of electromagnets. This can be obtained in the same way as a voltage source, the only difference being that in this case the voltage produced by the output current over a fixed resistor is compared to a reference voltage.

The principles described above can also be applied to the design of stabilized power supplies using transistors. However, some specific difficulties occur with supplies of relatively low voltages (not more than a few tens of volts) and large currents (up to tens of amps), which are much in demand.

Silicon diodes are normally used for rectifying these low voltages and high currents because they can operate at a high temperature (approx. 150 °C). This makes it possible to undertake the supply of a greater power within the same dimensions. Smoothing is accomplished by means of electrolytic capacitors of a very large total value, e.g. 10^4 μF. Fig. 29-22 shows the characteristic of the BYZ13 diode and a full-wave rectifier, which gives a ripple voltage of about 2 volts at full load for 10 volts and 5 amps.

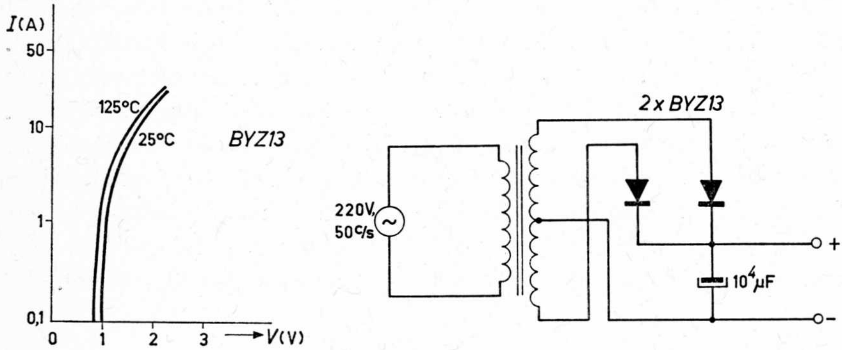

Fig. 29-22

A power transistor may be used for the series element of the stabilization circuit; Fig. 29-23 shows the characteristics of this type of transistor. It indicates that the base current will be approx. 0.17 A at an emitter current of 5 A. It is therefore not possible to make a direct connection between the base and the output of a conventional transistor amplifier. This difficulty can be overcome by the cascade connection of a few transistors as indicated in Fig. 29-24. The total current amplification factor of this combination is the product of those of the individual transistors. There is obviously a great spread possible in this value. The transconductance from the base voltage of the OC171 to the emitter current of the ASZ17 is about one third of that of the ASZ17 alone because the applied voltage is distributed almost equally over the three transistors. In practice resistors must be put in the emitter circuits of the AUY10 and OC171 as indicated in Fig. 29-26; otherwise the leakage currents of the power transistors may alter the working points. The properties of the combination are only slightly changed by this.

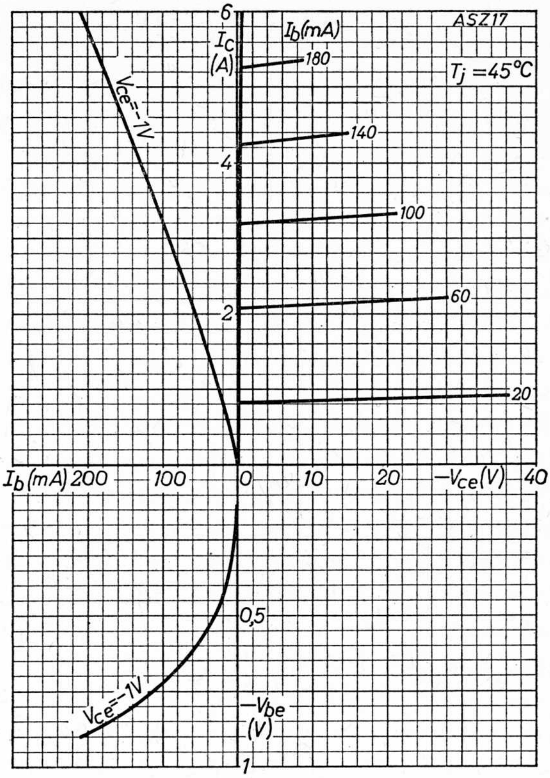

Fig. 29-23

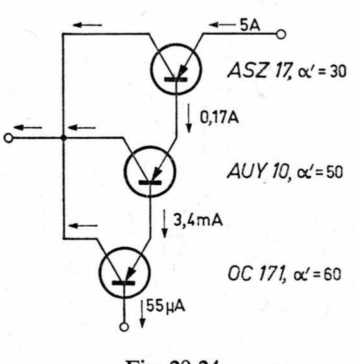

Fig. 29-24

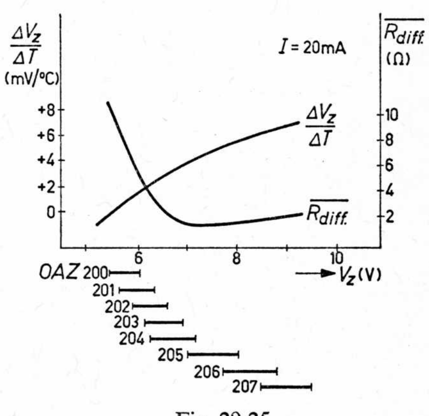

Fig. 29-25

The amplifier used will be preferably a one- or two-stage transistorized balanced amplifier, possibly followed by an emitter follower for the control of the series element. The zero-point drift of such a transistorized amplifier is normally less than 10 μV when referred back to the input and at constant temperature. With a change in the ambient temperature, the difference in the changes of the base-emitter voltages of the input transistors will generally be the determining factor. If we make the thermal conductances of these transistors as equal as possible by placing them close together in a good heat-conducting material, this zero drift can be limited to a few tenths of a millivolt. This method thus gives a relative drift of less than 0.01 per cent at a reference voltage of the order of 10 volts.

In transistorized power supplies a semiconductor element known as a Zener diode is normally used as a voltage reference element. Its operation is based on the fact that a semiconductor diode "breaks down" in the reverse direction at a certain voltage; the electrons which break away then have sufficient energy to liberate new current carriers, so that the current will increase strongly for a small increase in voltage. The differential resistance then becomes very small, particularly in the case of large currents, whilst the temperature coefficient of the Zener voltage can be made small by an adequate doping with impurities.

As this doping also affects the Zener voltage V_Z itself, a relation exists between V_Z and its temperature coefficient. Fig. 29-25 gives a few data on the Zener diodes of the OAZ 200–207 series. It shows that in the case of type OAZ201, a very small temperature coefficient is possible, whilst the differential resistance amounts to only a few ohms at currents of 10–20 mA. The stability of the voltage of this type of reference element can easily be

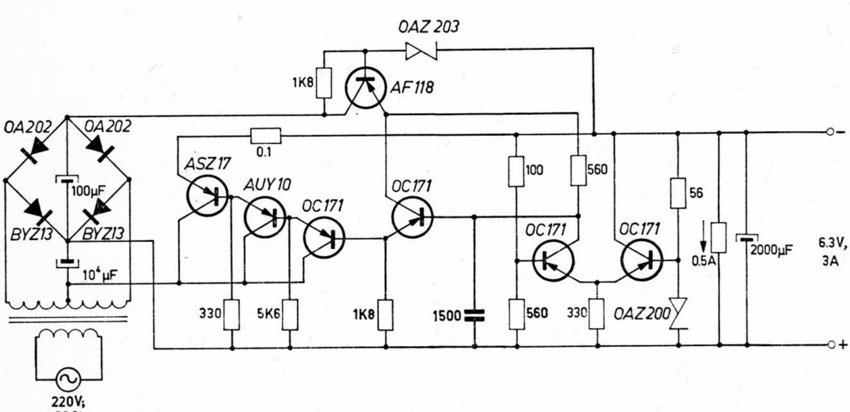

Fig. 29-26

arranged to be less than 10^{-3} and even 10^{-4} if the temperature is also stabilized to some extent.

It is obvious that the normal rules for the ratio of the time constants present in the control system also apply to transistorized power supplies. The relatively low cut-off frequency of power transistors necessitates the presence of capacitors of a very high value across the output when a small internal resistance of the supply is required. This is avoided in many commercially available designs by making the control amplifier slow and the output capacitance small. This solution should, however, be rejected for general use, because of the occurrence of a recovery time and the danger of oscillation with a capacitive load.

Fig. 29-26 shows the circuit of a design where the time constant of the output has been made large. A change of 10 per cent in the mains supply produces here a change of less than 1 mV in the output voltage of 6.3 volts. The internal resistance amounts to less than 15 mΩ at all frequencies. The combined hum and ripple voltage at the output is less than 0.5 $mV_{pk\text{-}pk}$ This shows that in most cases it is not necessary to accept the less elegant solution of exchanging the time constants.

High-value electrolytic capacitors often possess considerable series resistance. Because such a resistance strongly reduces the efficiency of the output capacitor, this capacitor should preferably consist of the parallel combination of a large and a small capacitor, the latter having very little or no series resistance.

30. Interference

We have now learnt about the most important passive and active elements and their behaviour in various basic circuits. We have also discussed amplification and feedback, among others as a preparation for the ability to make adequate power supplies for these circuits. The basis for electronic design appears to have been laid in the previous sections. It remains, however, to discuss an often neglected subject, which is still very important because it immediately confronts anyone who practises electronics, and that is the problem of interference. It is remarkable to observe how the value of many experiments is reduced as a result of avoidable interference. It is certainly worthwhile therefore to give the matter some close attention.

We shall, of course, not include under the general subject of interference the failure of equipment, for example, by short-circuits, open-circuits, failure in leads, or the catastrophic failure of components. Modern components are of such a quality that this kind of failures need only occur sporadically in a well-designed and well-constructed piece of equipment. Except in those cases where failure would have very serious consequences, as in amplifiers in undersea cables, reactors and some medical applications, it would thus be uneconomic to take special precautions against their occurrence. What we shall discuss here are "interfering factors and influences" but for brevity's sake we shall use the term "interference".

Interference can be roughly divided into two categories. The first comprises all interference caused by external sources. They do not represent an essential limitation because in theory unlimited counter-measures can be taken. The second category comprises interference of a fundamental nature, which is inherent in the conducting mechanism of the various components. This difference can best be illustrated by an example. We have seen that the amplification of feedback amplifiers cannot be increased in an unlimited fashion because oscillation would unavoidably occur. This restriction is not present in an amplifier without feedback. It is there theoretically possible to increase the amplification unlimited by adding further stages. This would lead one to think that any signal, however small, can also be measured in this way. Practice, however, reveals quite quickly that this assumption is not correct. Even apart from the fact that without precautions, unwanted feedback and hence oscillation may easily occur, we soon discover that output signals can exist which are completely unrelated to the input phenomenon under examination. In most cases such an output will prove to have a clear correlation with the voltage of the mains supply. We shall see later that this voltage affects the amplifier's behaviour capacitively or inductively. This is clearly a case of external interference and its effect can be reduced by taking appropriate steps. But even if we feed the amplifier from perfectly constant supplies and place it in an ideal environment with perfectly constant temperature, illumination and humidity, and completely free from external electric and magnetic fields and mechanical and acoustic vibration, we shall still see that output signals will nevertheless be present which have no connection with the required signal. These signals are caused by the thermal motion of electrons and ions, and by other phenomena inherent in matter, which have an effect in the components of the amplifier. They are unavoidable and have a random character. These spontaneous fluctuations are termed "noise". In this section we shall discuss the most frequently occurring external interferences and deal with the most important noise phenomena in the next section.

The most important external interferences are of a thermal, optical, mechanical, acoustic, electric and magnetic origin. If we mention the various modes in which they can make themselves felt and indicate the measures to counteract them, this naturally does not mean that all these precautions must be taken every time. Electronics would become an extremely involved business if they were. All the precautions mentioned will only be necessary in extremely sensitive apparatus and only then for those components which handle the smallest signals.

The ambient temperature normally changes very slowly. This results in interference particularly occurring in the case of amplifiers for slow signals, especially d.c. amplifiers. But high-frequency amplifiers can also be subject to this troublesome interference because either the adjustment of the active elements, or the elements which determine amplification, change with temperature. Air currents can be very troublesome at low frequencies. This can be countered by introducing thermal insulation, but should not be overdone because then the heat which is produced in the circuit will be poorly dissipated, so that the temperature within the thermal enclosure will continue to rise for a long time and may become quite high. Usually, an efficient remedy against changes in temperature is found in compensation. For example, anode resistors of a balanced stage are preferably placed together and if necessary thermally coupled by arranging the whole in a single unit. Since this can also increase the stray capacitances, we must often accept a compromise solution. The same considerations apply to valves; in a cascode differential stage the lower two triodes together are enclosed in a single envelope. Transistors in balanced stages are placed close together in the same small space like a brass block, or better still a single crystal. It is also obvious that one should select components with a low temperature coefficient for the sensitive part of the circuit. If necessary it is, of course, possible to incorporate thermostatics in the equipment.

Apart from the tropical regions, humidity usually plays a small part in equipment which uses valves, because its own heat generation has a sufficient drying action. It can become apparent that instruments for the amplification of signals at a high impedance level (current and charge measurements), which have not been used for a long time, will reach their full performance only after having been operational for some time. In the case of transistorized equipment, the heat generation is usually very much smaller, but then on the other hand, so is the impedance level, so that only the deterioration of components due to humidity remains as an unwanted effect.

A troublesome source of interference is "microphony", an effect which can be especially important with valves. Shocks and acoustically transferred vibrations can cause geometrical changes in the physical structure of a valve, for example the resonance of grids made of thin wire. When the interference signal has components with frequencies in the neighbourhood of the resonance frequency (100–1000 c/s) of the grids, the effect can become very disagreeable. Some types of valves are specially designed to prevent this effect by means of a more rugged construction. Shock or vibration can also cause permanent change, such as a change in the position of the heater in the cathode structure which may result in a different temperature distribution. This effect is extremely troublesome with d.c. amplifiers.

However, once again it is possible to take precautions against vibration. Mechanically produced vibration can be counteracted by spring suspension of the sensitive parts. The resonance frequency of this suspension must then be lower than the frequencies which are of interest or those of the expected vibrations. This can lead to extremely slack suspensions in the case of low-frequency amplifiers. A measure which can also be taken with existing equipment is to place the whole, possibly weighted, on a springy support such as foam rubber, foam plastic, tennis balls, air cushions, etc. Acoustic insulation is the obvious countermeasure for the penetration of vibration transferred acoustically. This insulation must then be so rigid that its own resonant frequency lies above that of the vibration. However, this may sometimes lead to heavy constructions and as an alternative the resonance must be heavily damped. For adequate damping of thin metal walls (1 mm), materials are available with a very high internal damping factor ("Silentblock", "Silentium", etc.) which can give excellent results.

On some occasions, interference can be the result of vibrating leads which have a high d.c. voltage with respect to their environment (capacitive effect). If these leads are used to transfer the output signal from a suspended part to the rigid part of the equipment, the obvious remedy is not possible. We must then ensure that the signal is transferred at a d.c. voltage level which corresponds better to that of the environment.

It should also be noted that transistors are much less vulnerable to acoustic interference than valves.

That light can also affect the behaviour of valves is not so well known. This effect is small, but nevertheless noticeable interference can occur as a result of strong changes in ambient light. The level of ambient light alters the ignition and burning voltages of gas-filled voltage reference valves. The remedy in this case is also obvious as transistors are very sensitive to light and so they are enclosed in an opaque envelope.

The greatest interferences, which can be troublesome even in less sensitive equipment is caused by magnetic and electric fields. The mains supply is their principal origin and in some cases such interference can even enters apparatus by conduction via the primary winding of the supply transformer. We shall first deal with the consequences of this.

With most mains supplies, one side is supposed to be earthed. Most electronic devices are also earthed, not only for safety reasons but also to prevent various capacitive interference effects. However, earth is far from being an ideal conductor. Therefore an impedance exists between different points and thus potential differences can occur. It should therefore not surprise us if we can measure a 50 c/s a.c. voltage of a few volts or even a d.c. voltage between the neutral side of the mains supply and the "earth" of, say, a laboratory. This is normally not very important because most apparatus are fed with transformers. It is, however, important that different earth leads can have different voltages against the earth-side of the mains supply and thus can have a potential against each other. Therefore it is a prime rule, when more than one earth lead is present, to use only one of them for the earthing of devices which are used in conjunction with each other.

It is also not always possible to neglect the resistance, and, at high frequencies, the inductance, of the cables used for earthing. Fig. 30-1 represents the situation that we find when an instrument is supplied from a mains transformer. The secondary winding is then connected to some "zero-point" of the circuit, which is connected to an existing earth terminal by means of a lead having an impedance Z. There will be capacitance between the windings of the transformer which usually lie one on top of the other. Although this

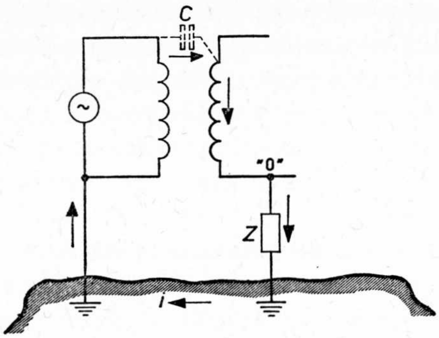

Fig. 30-1

capacitance is "distributed", it can nevertheless be taken into account as a single capacitance C which usually has a value of some hundred picofarads. The current passing through this capacitance follows the path indicated by the arrows and flows through the impedance Z of the earth lead. C is by far the largest impedance in the circuit and thus determines the value of the current. The latter will amount to about 20 μA at a voltage of 220 V, 50 c/s and $C=300$ pF. Since both voltages and capacitances in a transformer can be larger than we have assumed here, we must reckon with still greater values for this current. The "zero"-point of an instrument can thus have a certain potential with respect to the external point assumed to be earth. A second instrument will generally have a different potential with respect to this earth, so that there will be a potential difference between the zero-points of both instruments. The frequently recommended remedy of inserting an electrostatic screen between the transformer's windings gives little improvement. For when connecting this screen to the earth side of the mains supply (Fig. 30-2) the capacitive current i_1 of the primary winding to the

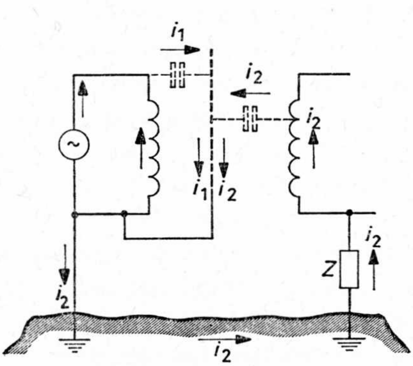

Fig. 30-2

screen will have been rendered innocuous, but current i_2 of the secondary winding still passes into the lead. Connecting the screen to the circuit's zero-point gives the reverse situation (Fig. 30-3). The correct solution is to

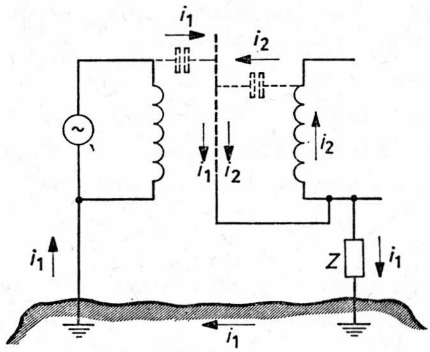

Fig. 30-3

insert two screens, one at the secondary winding connected to the zero-point, and the other to the earth side of the mains supply or to another earth point (Fig. 30-4). The screens are designed as "open" windings of copper foil, separated by a winding of insulating material. We shall refer later to the problem of the earthing of more than one instrument.

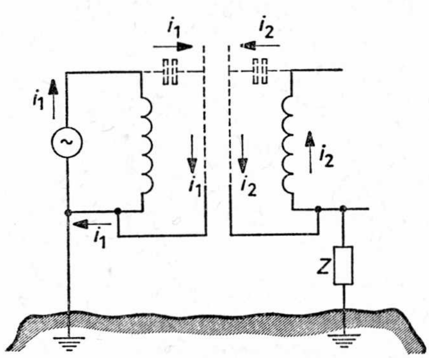

Fig. 30-4

Induced interference voltages in the circuit as a result of external electric fields can be avoided simply by placing the whole in a conducting case which is then earthed. Although, because of resistance and inductance, the screening effect of such a "Faraday cage" decreases at high frequencies, it will nevertheless remain very efficient. Iron of 1 mm thickness is normally used but copper may also be used for very high frequencies or rigorous requirements. Sometimes, the entire piece of equipment is placed in a Faraday cage, which is so large that there is room for the observer as well. These cages normally have double walls and those of good quality are very expensive. However, if the equipment is properly designed, it is nearly always possible to take appropriate precautions which render such a cage unnecessary.

It is also sometimes expedient to remove the voltages induced by internal electric fields by conducting screens, e.g. to remove parasitic coupling from output to input, which might result in oscillation or a change in the transfer characteristic.

The situation becomes more difficult, at least at low frequencies, when one must avoid interference voltages induced by magnetic fields. At high frequencies, conventional conducting cabinets offer a good screening against magnetic fields, because the Foucault currents induced in these cabinets produce a large enough inverse field to ensure that the internal space is almost entirely field-free. However, the Foucault currents decrease with the frequency so that at low frequencies we must use materials with better magnetic properties such as iron or (better but more expensive) μ-metal. The most efficient screening is given by alternating layers of μ-metal and copper. Such a rigorous measure will be necessary mainly with coils and transformers: a good reason for eliminating these when working with low frequencies. We shall, however, see that this is not always possible (Section 32).

In the case of equipment supplied from the mains, the principal origin of an interfering magnetic field is the supply transformer. The 50 c/s extraneous field increases very strongly with the permitted maximum magnetic field in the magnetic material. One often goes up to as far as 12,000 gauss to economize on the volume of the core material. A better way, however, is to use a little more material in the core and to limit the maximum magnetic field to 10,000 gauss (7,000 gauss with very sensitive instruments). This is more economical than trying to contain the extraneous field with the means previously mentioned. It is obvious however, that the supply transformer should be positioned at the greatest possible distance from the sensitive parts of the circuit, and particularly from coils and transformers used in the amplification of low-level signals. In addition, we should note that valves are also sensitive to magnetic fields, because the path of the electrons is changed. Since magnetically induced interference voltages are proportional to the enclosed flux, they can also be combated (apart from reducing the flux) by avoiding enclosure, i.e. by avoiding large loops in the assembly.

Further the loops present should preferably not contain iron. For example, a wire-wound resistor in the first stage of a sensitive amplifier must not be mounted with a steel bolt.

Leads to heaters of valves supplied with a.c. current can cause interference voltages along magnetic and electrical paths. The leads should be twisted together to avoid the first occurrence. Electrical interference is combatted by compensation; the centre tap of the heater winding is then earthed (top of Fig. 30-5). If no centre tap is available, an artificial centre can be earthed: the centre is formed by two low-value resistors (centre of Fig. 30-5). A hum compensation is obtained by making these resistors adjustable ("humdinger" potentiometer; bottom of Fig. 30-5). Earthing of the heater circuits can also prevent "cross coupling" from output to input. This simple precaution is therefore always recommended. If all these possibilities for reducing hum voltages are not sufficient, a d.c. heater supply should be used.

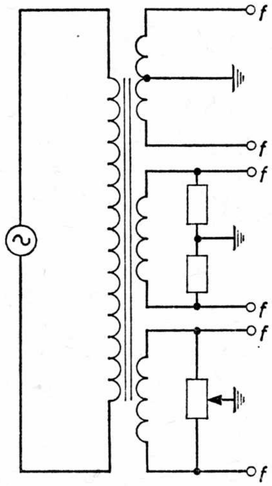

Fig. 30-5

As we have mentioned before, a large number of these precautions can be omitted in the majority of cases. On the other hand, some of them constitute an inexpensive routine, which should always be followed. For example, in many cases a good metal case offers adequate protection against draughts, light, acoustic vibration, as well as electric and magnetic fields. Experimental arrangements excepted, electronic equipment is therefore usually housed in such cases.

On the other hand, the precautions mentioned are not entirely adequate

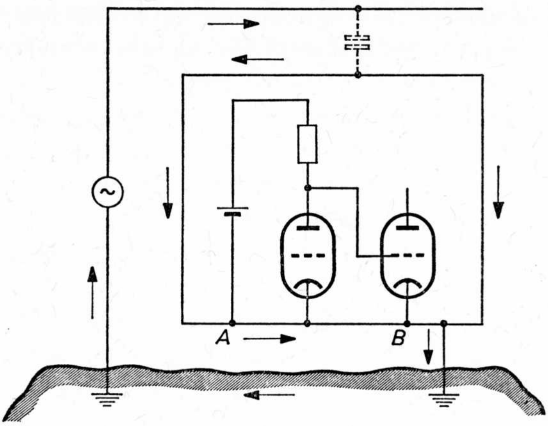

Fig. 30-6

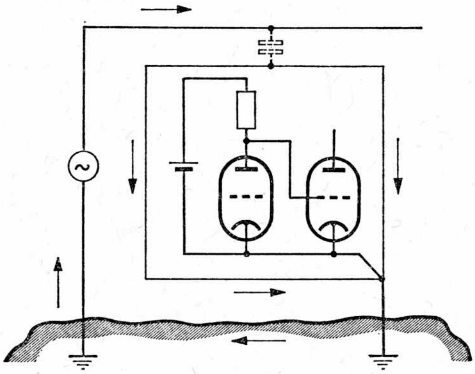

Fig. 30-7

for preventing all interference, however carefully applied. This can be seen from Fig. 30-6, where once again the mains supply acts as example of an interference source. Because the mains can easily have a capacitance of a few picofarads to the earthed case, a 50 c/s a.c. current will pass through the case, which gives a potential difference between points *A* and *B*. The input voltage of the second valve in this case is not only determined by the signal voltage of the first valve, but also contains the potential difference between A and B, so that the latter is also amplified. Thus the case must only be used for screening and not as part of the circuit. For the latter purpose, a separate signal earth bus-bar should be used, which is connected to the earthing point of the case (Fig. 30-7) or is entirely isolated from it.

As we have seen above, a common earth connection must be used whenever possible in order to avoid potential differences when using more than one case. The cases may be connected by means of a coaxial cable (Fig. 30-8). If the right-hand case is earthed as indicated in the

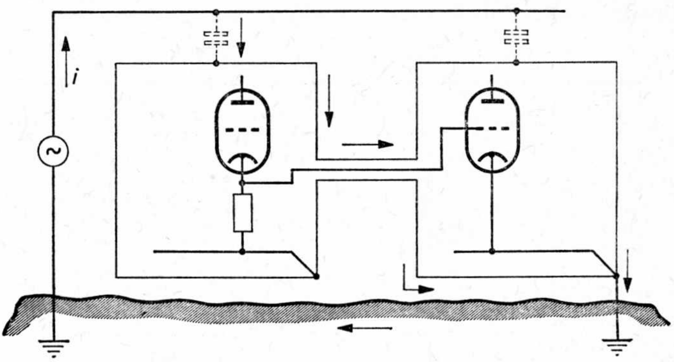

Fig. 30-8

figure, the 50 c/s a.c. current of the mains supply to the left-hand case must pass through the cables' outer screen in order to reach earth. The current then produces a potential difference across the impedance of the cable's outer screen which arrives at the input of the right-hand case in series with the required output voltage of the left-hand case. If only the left-hand case is earthed, the capacitive current of the right-hand case will pass through the cable's outer screen, with a corresponding result. It would be wrong to deduce from this that both cases must be earthed. This is explained by the fact that, apart from the above-mentioned

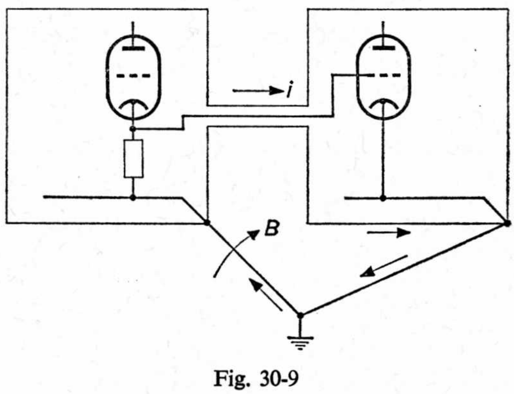

Fig. 30-9

effect of the occurrence of potential differences, a loop will be produced, which will pass a loop current as a result of the extraneous magnetic field of the mains supply (Fig. 30-9). The basically correct solution for coupling cases is to use a twin-lead cable and a difference stage input for the second case (Fig. 30-10). It is unfortunate that electronic instruments are only occassionally designed in this correct way.

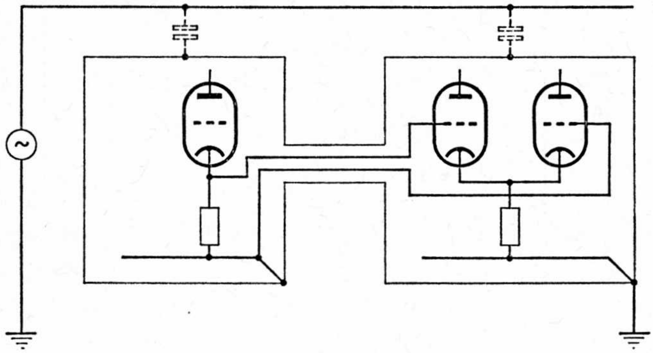

Fig. 30-10

Although we have seen that interferences can enter equipment in many different ways, they are basically simple and their total can be surveyed comparatively easily. We can compare this with a maze: when surveying the maze from a vantage point, it is far easier to see the solution than when one is walking in it. Therefore, if interference signals are encountered in practice, these often enter the instrument by a few of the paths mentioned. They can even counter-balance each other to a certain extent, so that the total interference may become greater by taking individual measures which are quite correct in themselves. This is not to say that these precautions are therefore wrong and should not be taken, but that there is a further path which needs elimination, which has to be traced and remedied.

That good measurement enclosures reduce external fields is shown by the following data: copper foil of 0.1 mm thickness reduces magnetic and electric fields beyond 100 kc/s about 10^4 times, at 30 kc/s the reduction of magnetic fields is only 10^2. A massive iron plate of 0.5 mm thickness gave a reduction in magnetic field at 30 kc/s by 2,500 times. At values above 100 kc/s, the reduction was approx. 10^5 times for both magnetic and electric fields.

A measurement of the magnetic field caused by the mains supply was carried out in the centre of a laboratory measuring 6 $\times$ 6 metres, when electrical equipment with a total drain of approx. 10 A was placed on benches along the walls. The following values were found for the fields: 50 c/s: 0.75×10^{-9} V/cm^2, 150 c/s: 2.3×10^{-9} V/cm^2 (!), 800 c/s: 10^{-11} V/cm^2.

From this it is clear that any loop formation in the input circuit must be avoided with very sensitive measurements, for example with selective amplifiers (Section 33) which have to amplify frequencies which are a harmonic of the mains supply frequency.

31. Noise

Even after having succeeded in reducing the effect of external interference to the very minimum, it is not possible to increase an amplifier's sensitivity for ever because we then have to deal with natural limitations in the form of "microscopic" interference. This results from the fact that the transport of electrical charge takes place in discrete quantities, and that fluctuations occur in a current composed of these quantities. The resultant effect of these fluctuations is indicated by the generic term "noise", which clearly originates from acoustics. There are various forms of noise, and we can speak of noise sources in the same way as we do of signal sources. Which noise source plays the most important part in limiting the sensitivity depends on the circuit used and on the components it contains.

Good numerical comparison data on noise can be arrived at by considering the fluctuations in charge transport as random phenomena. It would lead too far afield to discuss this in depth, but it is desirable to understand some of the background by considering some results of the theory of probability.

Let p be the probability that a certain result occurs from a certain event, and let $P_m(n)$ be the probability that from m of these events (which are independent) this certain result will occur n times ($0 \leqslant n \leqslant m$). We can now work out the relations as follows. A first possible course of events is that the certain result is obtained in the first n events, but not with the (m–n) subsequent ones. The probability of the former happening is p^n, that of the latter $(1-p)^{m-n}$ and the total probability therefore $p^n(1-p)^{m-n}$. Then we have all those possibilities which only differ from the above by the sequence in which the n results occur. Here, too, the probability of occurrence of each possibility is $p^n(1-p)^{m-n}$. The number of possible sequences is $m!/n!(m-n)!$, so that the total probability will be:

$$P_m(n) = \frac{m!}{n!(m-n)!} p^n(1-p)^{m-n}$$

We then obtain for the mean value $\bar{n}$ of the number of positive results n which is achieved as the sequence of m events is repeated many times

$$\bar{n} = \sum_{n=0}^{\mathrm{m}} nP_m(n) = \sum_{n=1}^{\mathrm{m}} \frac{n \cdot m!}{n!(m-n)!} p^n(1-p)^{m-n} =$$

$$= mp \sum_{n=1}^{m} \frac{(m-1)!}{(n-1)!\,(m-n)!}\, p^{n-1}(1-p)^{m-n} = mp,$$

a result that could well be expected.

Example: The probability that when tossing a coin ten times, we throw "heads" six times and "tails" four times is:

$$\frac{10!}{6!\,4!}\,(\tfrac{1}{2})^{10} \approx 20\%$$

If this experiment is repeated many times, all values between 0 and 10 will occur for the number of times heads is thrown in each separate series. The mean of these values will be 5.

An important value for indicating the degree of spread of the individual values around the mean is the "mean deviation" σ, defined by

$$\sigma^2 = \overline{(n-\bar{n})^2}$$

Since $\bar{n}$ is a constant when averaging the square of the deviation, we can write:

$$\overline{(n-\bar{n})^2} = \overline{n^2 - 2\,n\bar{n} + \bar{n}^2} = \overline{n^2} - 2\,\overline{n\bar{n}} + \bar{n}^2 =$$
$$\overline{n^2} - 2\,\bar{n}^2 + \bar{n}^2 = \overline{n^2} - \bar{n}^2$$

$\overline{n^2}$ is similarly derived as $\bar{n}$ from:

$$\overline{n^2} = \sum_{n=0}^{m} n^2\, P_m(n) = \sum_{n=0}^{m} \{n(n-1)+n\}\, P_m(n) = \sum_{n=0}^{m} n(n-1) P_m(n) +$$

$$\sum_{n=0}^{m} n\, P_m(n) = m(m-1)p^2 \sum_{n=2}^{m} \frac{(m-2)!}{(n-2)!\,(m-n)!}\, p^{n-2}(1-p)^{m-n} + mp =$$

$$m(m-1)p^2 + mp = m^2p^2 + mp(1-p) = \bar{n}^2 + \bar{n}(1-p),$$

so that
$$\sigma^2 = \bar{n}(1-p).$$

The absolute value of the mean deviation thus increases proportionally to $\sqrt{m}$ or $\sqrt{\bar{n}}$. The ratio $\sigma/\bar{n}$, however, is inversely proportional to $\sqrt{m}$ or $\sqrt{\bar{n}}$. It follows from the first relation that when two gamblers play together, one must eventually go bankrupt. From the second relation not only can we trace the properties of noise sources, but also the fact that the accuracy of a measurement increases in proportion to the root of the number of observations, or to the root of the measuring time.

The above relations are very useful when m is rather small. However, they will become less easy to manipulate for high values of m and n, because of the factorials. Approximations have been developed which do not show this disadvantage. The most important are those by Poisson and Gauss.

The Poisson distribution applies when m is very large, but, at the same time p is so small that their product $\bar{n}$ can still be small, whilst n is of the same order of magnitude as $\bar{n}$. We can then easily find from the limit as $m \to \infty$:

$$P(n) = \frac{(\bar{n})^n}{n!} e^{-\bar{n}}$$

and $$\sigma^2 = \bar{n}$$

The Gaussian distribution goes still further and assumes that $\bar{n}$ is also rather large. By using Stirling's equation:

$$n! \approx n^n e^{-n} \sqrt{2\pi n}$$

and expanding $\ln P(n)/P(\bar{n})$ into a series, we find:

$$P(n) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(n-\bar{n})^2}{2\sigma^2}}$$

where once again $\sigma^2 = \bar{n}$.

With the current fluctuations that we shall consider, both the number of charge carriers present and the number of charge carriers contributing to the current, are very large and the Gaussian distribution can be used. This is shown in Fig. 31-1. The curve obtained from the latter by integration (Fig. 31-2) is of greater interest; this indicates the probability for the absolute value of deviation $|n-\bar{n}|$ to exceed value $k\sigma$. It follows that the probability that $|n-\bar{n}|$ will be more than four times the mean deviation is so small that,

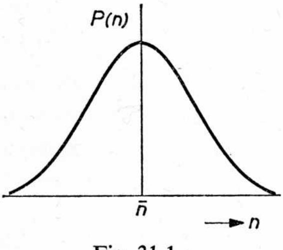

Fig. 31-1

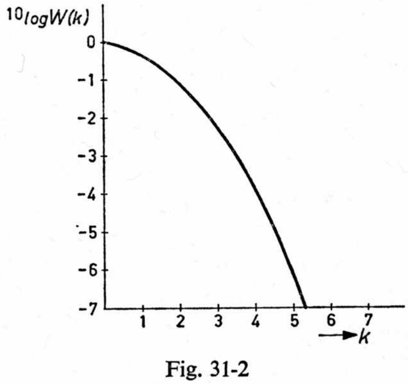

Fig. 31-2

apart from the rare cases when the results would be catastrophic, they do not need to be taken into account.

The above can be illustrated in a rather simple way by means of the noise current of a saturated anode. A diode is saturated when the anode voltage is so high that all electrons emitted by the cathode arrive at the anode. If n electrons are emitted by the cathode during a time interval τ, the mean current during this time interval will be nq/τ, where q = charge of the electron. Its mean will be the anode d.c. current, therefore $I_a = \bar{n}q/\tau$. The current deviation during the time interval τ is therefore:

$$i_\tau = \frac{(n - \bar{n})q}{\tau}$$

It follows from the above that:

$$\overline{i_\tau^2} = \frac{\overline{(n - \bar{n})^2} \cdot q^2}{\tau^2} = \frac{\bar{n}q^2}{\tau^2} = \frac{q}{\tau} I_a$$

We normally have only a moderate interest in this value because in practice the signal will pass through a system with limited bandwidth. It is therefore more important to know what frequencies can be expected in such a signal and what their amplitude will be. With pure random phenomena, as considered here, where the contribution of each electron gives an infinitely fast signal, there is no preference for any particular frequency; all frequencies will be equally represented. Even without additional bandwidth restriction, a limitation of the upper frequency spectrum should occur in practice because the mechanism prescribes a minimum time interval between emission and transfer of charge carriers (finite energy). In the case of the diode considered above and with most other noise sources these limit frequencies will be so high that the distribution over that part of the frequency spectrum which is important here, may actually be considered to be uniform.

By means of Fourier Analysis (Section 36) can then be derived the following value for the amplitude of the diode noise current in a frequency band having a width df:

$$\overline{i_a^2} = 2\,\tau\,\overline{i_\tau^2}\,\mathrm{d}f = 2\,q\,I_a\,\mathrm{d}f$$

The spontaneous occurrence of fluctuations in the current of valves is called shot noise (also known as Schottky effect).

Example: If I_a be 10 mA and the transmitted frequency band 10^4 c/s, we have: $\overline{i_a^2} = 2 \times 1.6{\cdot}10^{-19} \times 10^{-2} \times 10^4 = 3.2{\cdot}10^{-17}$. Therefore: $\sqrt{\overline{i_a^2}} = 5.6{\cdot}10^{-9}$ A. Thus we can reasonably expect here fluctuations up to approx. $2{\cdot}10^{-8}$ A in the diode current.

In case of a triode, the noise due to the shot effect under normal operation is considerably smaller than that which corresponds to the expression for the saturated diode. This can be explained as follows. At negative grid voltages there are a great number of slow electrons between cathode and grid, the so-called space charge; part of it moves towards the anode, whilst the remainder returns to the cathode. Increasing the number of emitted electrons will increase this space charge cloud. More electrons will return, so that this space charge has an equalizing effect on the anode current. Therefore, for the anode current fluctuations in the case of a triode we can write:

$$\overline{i_a^2} = 2\, q\, I_a\, \gamma\, \mathrm{d}f$$

where the suppression factor γ has in practice values of the order of 0.1–0.5.

The pentode shows more noise than the triode because the charge carriers passing through the control grid are distributed between the screen grid and the anode, and fluctuations occur in this distribution. The resulting additional noise is called partition noise, and can be taken into account by adding a factor >1 to the triode noise, which is dependent on the ratio I_a/I_{g2}. Partition noise does not occur of course with triode cascodes, which is another reason why this combination of triodes is used in many circuits instead of a pentode.

Fluctuations also occur in the grid current of valves. The relation is not necessarily the same as for the anode current because the grid current is composed of an electronic and an ionic component, which show mutually independent fluctuations. The current components will partly counterbalance each other, but the fluctuations are additive. The fluctuations are thus larger than corresponds to the relation $\overline{i_g^2} = 2qI_g\mathrm{d}f$, particularly in the neighbourhood of the point of complete compensation (where the grid current is zero). As, however, in those cases in measurement electronics where grid current fluctuations are important, the valves usually have such a large negative grid bias that the ionic part of the grid current is dominant, we can nevertheless derive a very good estimate from this relation. Good-quality valves have a low grid current, so that the absolute value of the fluctuations will be small. They can, however, not be neglected with high impedances in the grid circuit.

Example: The grid current of an electrometer valve is 10^{-14} A and practically consists of a single component. We thus have for a bandwidth of 10^3 c/s: $\sqrt{\overline{i_g^2}} = 1.8 \cdot 10^{-15}$ A, so that we can expect fluctuations of the same order as the mean grid current.

With transistors, fluctuations will occur in the diffusion of electrons and holes and in the formation and recombination of electron hole pairs. When we take a *p-n-p* transistor as an example, the noise in the collector current will be determined almost entirely by the number of holes that diffuse from the emitter, through the base to the collector. The shot effect equation therefore applies:

$$\overline{i_c^2} = 2\, q\, I_c\, \mathrm{d}f$$

The noise in the hole current to the base is mainly determined by the fluctuations in recombination occurring in the base. We can add the fluctuations in the much smaller electron component from base to emitter to this, whilst at high frequencies additional noise will be produced by the fluctuations in the diffusion time from emitter to collector. At low frequencies, we can also apply the equation of the shot effect to the noise in the base current so that in good approximation:

$$\overline{i_b^2} \approx 2\, q\, I_b\, \mathrm{d}f$$

Since the most important contributions in both these noise currents are the results of different processes, the mutual correlation will be small, particularly at low frequencies.

Fluctuations of an entirely different origin are found with good conductors. They are not caused by changes in the number of electrons which contribute to the conductance (in this case all electrons are free and their number is therefore constant) but the charge transport in the circuit can deviate from the mean value for short time intervals because of the thermal movement of the electrons. The value of this "thermal noise" follows from thermodynamic considerations of such a conductor in thermal equilibrium with its environment; it is given by the Nyquist formula: at a temperature T and in a frequency band df, a resistance has a mean noise power of $4kT\mathrm{d}f$ where k is Boltzmann's constant $= 1.38 \cdot 10^{-23}$ Wsec/°C. We can therefore describe the behaviour of a resistor R (Fig. 31-3) regarding thermal noise

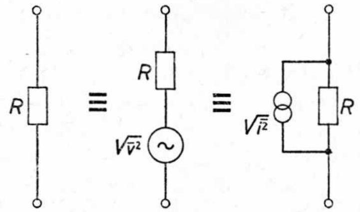

Fig. 31-3

as that of a noise-free resistor of the same value, connected in series with a purely fluctuating voltage generator v, given by:

$$\overline{v^2} = 4kTR\,\mathrm{d}f$$

or, connected in parallel to a purely fluctuating current generator:

$$\overline{i^2} = \frac{4kT\,\mathrm{d}f}{R}$$

$4kT$ has a value of approx. $1.6 \cdot 10^{-20}$ Wattsec at room temperature, so that the effective noise voltage $\sqrt{\overline{v^2}}$ is approx. 0.13 nanovolts per $\sqrt{\Omega\cdot\text{c/s}}$, and the effective noise current approx. 0.13 nanoamps per $\sqrt{\Omega^{-1}\cdot\text{c/s}}$. The limit frequency of thermal noise is of the order of magnitude of the natural frequency of the electrons, approx. 10^{13} c/s, so that this noise can be assumed to be independent of frequency in all practical cases ("white noise").

Indeed, metal wire resistors show this theoretically expected thermal noise; but with other types, such as carbon and metal film resistors, an additional noise occurs, termed "current noise", which has an effective value almost proportional to the current through the resistor, and therefore to the voltage across the resistor as well. The value of this noise also depends on the frequency; the noise energy being approximately inversely proportional to the frequency. It therefore becomes important at low frequencies. Measurements have shown that this frequency dependence remains valid for very low frequencies (0.01 c/s or even lower). Current noise is therefore determined by the ratio of the limit frequencies of the passed frequency band, so that the noise energy can here be given per octave or decade. The proportionality to the current indicates a fluctuating resistance value; for example in carbon resistors, because of the changing contact between the carbon particles. The method of manufacture is therefore important for the value of current noise. With good-quality carbon resistors, the resistance fluctuation is of the order of magnitude of 1 micro-ohm per ohm of resistance value in a frequency range of one decade. The fluctuation can be more than 100 $\mu\Omega/\Omega$ for poor-quality carbon resistors.

Metal film resistors also show this phenomenon, but good-quality specimens will still be one order of magnitude better than the best carbon resistors.

Example: The resistance fluctuation of a given carbon resistor of 100 kΩ is 2 $\mu\Omega/\Omega$ decade. With a d.c. current of 0.5 mA, the effective value of the current noise voltage in the frequency range of 1–10 c/s will be 100 μV. For the thermal noise in this range, this voltage will be $0.13 \cdot 10^{-9} \sqrt{10^5 \cdot 9} = 0.13$ μV. This indicates a strong preponderance of current noise. In the frequency range of 10^5–10^6 c/s, the current noise voltage is again 100 μV, and the thermal noise voltage $0.13 \cdot 10^{-9} \sqrt{10^5 \cdot 9 \cdot 10^5} \approx 40\ \mu$V.

Since thermal noise originates in the thermal motion of electrons as well as in the energy exchange between the conductor and the environment, the ideal capacitor and inductor will not show noise phenomena. These can only occur through imperfections in a capacitor's dielectric or in the core of an inductor. These conditions only occur very rarely. In a circuit consisting of passive elements, the total noise will be almost solely determined by the resistor contributions.

Both valves and transistors give an additional noise at low frequencies which shows a remarkable likeness to current noise produced by carbon resistors. With active elements one usually speaks of excess-, flicker- or $1/f$-noise. A great number of theories exist about the origin of this additional noise, but are of relatively little clarity. It is certain, however, that with the flicker noise as well, the frequency dependence remains valid for very low frequencies. With good-quality valves and transistors, the frequency at which both excess noise and shot noise have the same value, is of the order of 100 c/s. This can increase to more than 10^4 c/s for poor-quality types or individual bad specimens of the better types.

We shall see in the next section that the thermal noise of the base resistance R_{bb} of a transistor is also important.

The total noise at a point in a circuit can be calculated by adding together the contributions of the various noise sources at this point. As far as the noise sources are unrelated – and this is nearly always the case – it is necessary to add the mean squares because of the absence of a phase relation.

Example: A current of 3 mA passes through the valve of Fig. 31-4. The slope is 3 mA/V, the suppression factor of the noise due to the space charge is 0.1. Because of the excess noise, the noise power at 1 c/s is 1,000 times that in the "white-noise" region.

The anode resistor is of the metal film type and has a current noise of 0.25 $\mu\Omega/\Omega$ decade.

The various contributions to the noise voltage on the anode are calculated in the frequency range of 1–10^4 c/s, neglecting anode feedback.

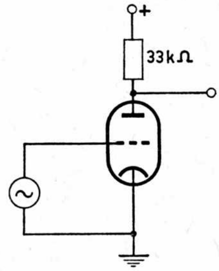

Fig. 31-4

Since the excess noise at 1 c/s is 1,000 times greater than the shot noise, the flicker effect makes a contribution in the 1–10^4 c/s band which corresponds to the shot noise in a frequency range of

$1{,}000 \int_1^{10^4} \mathrm{d}f/f = 9{,}200$ c/s. The total anode current noise is then:

$\overline{i_a^2} = 2 \times 1.6 \cdot 10^{-19} \times 10^{-1} \times 3 \cdot 10^{-3} \times 1.9 \cdot 10^4 \approx 1.8 \cdot 10^{-18}$ and therefore:

$$\overline{v_a^2} = 33^2 \cdot 10^6 \times 1.8 \cdot 10^{-18} \approx 2 \cdot 10^{-9}$$

The thermal noise of the anode resistor is:

$$\overline{v_a^2} = 1.6 \cdot 10^{-20} \times 33 \cdot 10^3 \times 10^4 = 5 \cdot 10^{-12}$$

The average voltage caused by current noise of the anode resistor is in four decades: $\sqrt{4} \times 0.25 \cdot 10^{-6} \times 3 \cdot 10^{-3} \times 33 \cdot 10^3 = 5 \cdot 10^{-5}$, therefore

$$\overline{v_a^2} = 2.5 \cdot 10^{-9}.$$

The mean anode noise voltage is $\sqrt{4{,}5 \cdot 10^{-9}} = 6.6 \cdot 10^{-5}$ V. With an amplification $SR_a = 100$, this gives a voltage of 0.66 μV when referred to the input.

32. Matching to the signal source

We shall now see how these rather complicated noise sources can be taken into account in practice, and what precautions are necessary for reducing the resulting interference signals to a minimum. When amplifying, or, more generally speaking, processing a signal, we shall always attempt to achieve the most favourable signal-to-noise ratio. A limit is obviously imposed by the ratio existing in the signal source under the same conditions (such as the same bandwidth). It is therefore necessary that the processing system reduces the signal-to-noise ratio as little as possible. Let us now consider a system with input and output terminals used for the linear processing

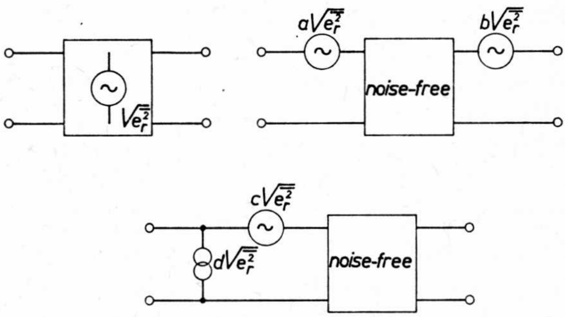

Fig. 32-1

of a signal (Fig. 32-1). If a noise voltage source $\sqrt{\overline{e_r^2}}$ is present somewhere in such a four-terminal network, its external effect can be taken into account by removing the noise source itself and replacing it by noise voltage sources $a\sqrt{\overline{e_r^2}}$ and $b\sqrt{\overline{e_r^2}}$ connected in series with the input and output terminals respectively, where a and b are determined by the circuit and the place of the original noise generator in the circuit. a and b will usually be complex numbers. Of course an internal noise current source can be replaced in a similar way. Four-terminal networks with a transfer from input to output which differs from zero, i.e. with all practical four-terminal networks, offer a different and more attractive possibility, which consists of replacing an internal noise source by a voltage generator in series with the input terminals and a current source parallel to the input terminals (right-hand side of Fig. 32-1). As for the first case, these two sources will be entirely correlated, unless one of them is zero. If the four-terminal network contains different, mutually independent, noise sources, these can all be transferred to the input. The current sources can then be replaced by a single current source I and the voltage sources to a single voltage source E (Fig. 32-2). These two sources will generally possess a certain degree of correlation.

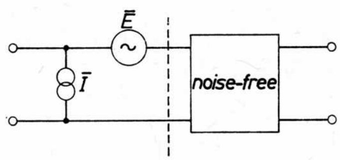

Fig. 32-2

The advantage of this method is that one can now easily compare the signal with all noise sources at the same signal level, namely the input of the network. It is sufficient to determine the values of the signal-to-noise ratio at the input terminals of the four-terminal network which has been made noise-free, for the cases that noise sources E and I are present or absent. Good measures for this signal-to-noise ratio are the ratio between signal energy and noise energy supplied to the four-terminal network, or the ratio of the mean effective signal voltage and noise voltage across the input terminals. Since the supplied signal energy is not dependent on the presence of either E or I, the effect of the four-terminal network on the signal-to-noise ratio is also given by the ratio of the supplied noise energies or of the mean effective noise voltages across the input, with and without E and I.

We thus arrive at the situation shown in Fig. 32-3, where R accounts for the load by the four-terminal network, and $v_{s,n}$ and R_s are the mean effective noise voltage and the internal resistance – which has been made

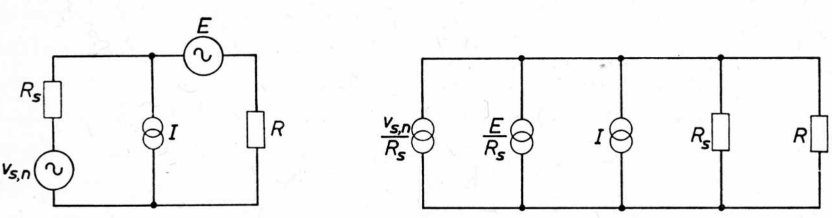

Fig. 32-3

noise-free – of the signal source respectively. A constant relation exists between the energies dissipated in R and R_s, so that the above-mentioned ratio of noise energies can be replaced by the ratio of the noise energies which the sources could supply to the internal resistance R_s of the signal source.

When calculating these energies, we must remember that sources E and I will generally be partly correlated. The energy which E and I can supply together to R_s is: $\overline{(E + I\,R_s)^2}\big/R_s = \overline{E^2}\big/R_s + 2\,\overline{EI} + \overline{I^2}R_s$.

The value of the second term, lying between 0 and $\pm\, 2\sqrt{\overline{E^2}\,\overline{I^2}}$, is determined by the degree of correlation between E and I.

The situation with a valve is very easily treated. The noise in the anode current can be accounted for by considering the valve noise-free and having the mean anode noise current $\sqrt{\overline{i_a^2}}$ caused by a grid noise voltage $\sqrt{\overline{i_a^2}}/S$. This noise voltage may in turn be visualized as caused by a resistance in the grid circuit. The value of this "equivalent noise resistance" R_{eq} follows from:

$$\sqrt{4kTR_{eq}df} = \frac{\sqrt{\overline{i_a^2}}}{S}$$

In the frequency range where excess noise can be neglected, the value of R_{eq} is between 300 and 3,000 Ω for the usual valve types. At lower frequencies, R_{eq} will increase because of the excess noise, and will become inversely proportional to the frequency at very low frequencies.

With special-quality valves, R_{eq} is 10 kΩ at 1 c/s, but with valves of poorer quality it is more than a hundred times as much. The value of R_{eq} which can be expected for a valve is usually indicated in the valve data for a number of frequencies. When no frequency is mentioned, the value refers to the "white-noise" range, i.e. the range where no excess noise is present.

Grid current noise $\sqrt{\overline{i_g^2}}$ can be accounted for by inserting a noise current source in the grid circuit. This may also be visualized as caused by a resistance R_{cur}, where

$$\frac{4kTdf}{R_{cur}} = \overline{i_g^2}$$

Since the grid current itself is strongly dependent on the quality of the vacuum in the valve and on the working point, so is the grid current noise. Hence the expected value for R_{cur} is usually not given but can be estimated from the grid current which has to be measured. In this case,

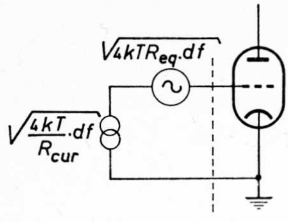

Fig. 32-4

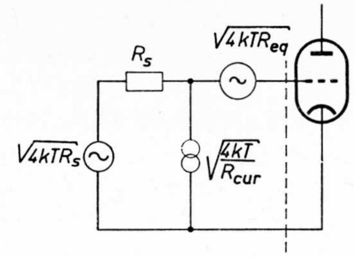

Fig. 32-5

there is no correlation between the two noise sources (Fig. 32-4, where d*f* is the width of transmitted frequency band).

That each of the noise currents can be replaced by one noise source at the input is because of the high value of I_a/I_g ($> 10^4$). It is in so far incorrect to replace the anode current noise by the voltage source that this also affects the grid current noise. This could be compensated for by the parallel connection at the input of a current source of value $\sqrt{\overline{I^2}} = S_g/S \cdot \sqrt{\overline{i_a{}^2}}$, where $S_g = \Delta i_g/\Delta v_g$.

Because $S_g/S \approx I_g/I_a$, it means that the energy of this current source is about I_a/I_g times smaller than the current source which must be inserted for the grid current noise. Similarly, the voltage source corresponding to the grid current noise is as many times smaller than that for the anode current noise. It is obvious that such a simplification is not normally permissible for a transistor where the ratio I_c/I_b is so much smaller.

Example: The data of a triode are: $I_a = 1$ mA, $I_g = 10^{-9}$ A, where $(I_g)_{\mathrm{el}} = -2 \cdot 10^{-9}$ A and $(I_g)_{\mathrm{ion}} = 3 \cdot 10^{-9}$ A. The transconductance is 2 mA/V, and the suppression factor γ is 0.2. We find for each cycle per second of bandwidth, outside the excess noise area:

$$\frac{\overline{i_a{}^2}}{S^2} = \frac{2qI_a\gamma}{S^2} = 1.6 \cdot 10^{-17}\ \mathrm{V}^2,$$

$$\overline{i_g{}^2} = 2\,q\{(I_g)_{\mathrm{el}} + (I_g)_{\mathrm{ion}}\} = 1.6 \cdot 10^{-27}\ \mathrm{A}^2$$

This yields, with $4\,kT = 1.6 \cdot 10^{-20}$ Wsec.

$$R_{\mathrm{eq}} = 10^3\ \Omega \quad \text{and} \quad R_{\mathrm{cur}} = 10^7\ \Omega$$

We therefore have the empirical rule that the ratio $R_{\mathrm{cur}}/R_{\mathrm{eq}}$ for a valve is of the order of magnitude 0.01 I_a/I_g in the white-noise area, i.e. 10^3–10^5 depending on the particular type. This ratio is reduced due to excess noise at lower frequencies.

When connecting the grid of the valve to a signal source v_s with an internal resistance R_s, and assuming that the noise from the signal source is only due to the thermal noise of R_s, we have the situation as shown in Fig. 32-5;

the noise signals are indicated here per cycle per second of bandwidth.

The grid noise voltage therefore consists of three components, $\sqrt{4kTR_s}$, originating from the source resistance, $\sqrt{4kTR_{\mathrm{eq}}}$, originating from the anode current noise, and $R_s\sqrt{4kT/R_{\mathrm{cur}}}$, originating from the grid current noise.

As these sources are not correlated, the mean value of the grid noise voltage will be:

$$\sqrt{4kT\left(R_s + R_{\mathrm{eq}} + \frac{R_s^2}{R_{\mathrm{cur}}}\right)}$$

In the absence of internal noise sources of the valve, this would have been:

$$\sqrt{4kTR_s}$$

which gives a relative increase of:

$$\sqrt{1 + \frac{R_{\mathrm{eq}}}{R_s} + \frac{R_s}{R_{\mathrm{cur}}}} \qquad (32.1)$$

This expression has, as function of R_s, a very flat minimum when R_s equals $R_s{}^* = \sqrt{R_{\mathrm{eq}} \cdot R_{\mathrm{cur}}}$. The value of this minimum is

$$\sqrt{1 + 2\sqrt{R_{\mathrm{eq}}/R_{\mathrm{cur}}}}$$

The ratio $R_{\mathrm{eq}}/R_{\mathrm{cur}}$ is usually very small in the case of valves (<0.001). This means that a signal originating from a source with an internal resistance of the order of $R_s{}^*$ can be amplified by means of a valve, adding practically no noise. Because the ratio is so small the noise factor only increases slightly when R_s deviates from $R_s{}^*$. The increase for $R_s = R_{\mathrm{eq}}$ and $R_s = R_{\mathrm{cur}}$ is only 40 per cent.

It is theoretically possible to change a source with a given internal resistance into a source with a different internal resistance by means of a transformer, without altering the signal-to-noise ratio. We could conclude from this that any signal voltage could be amplified by means of a valve without adding significant noise. However, we shall see that there are some practical limitations.

Since this is one of the rare cases where the use of a transformer as a signal-amplifying element cannot be avoided, we shall derive a few relevant equations.

A transformer (Fig. 32-6) with a primary inductance L_1, a secondary inductance L_2, and a mutual inductance M gives:

$$v_1 = j\omega L_1 i_1 + j\omega M i_2$$
$$v_2 = j\omega M i_1 + j\omega L_2 i_2$$

We can write $M = k\sqrt{L_1 L_2}$, with the coupling factor $k < 1$, but the deviation $1-k$ will be small for a good transformer: less than 0.1. The resistance of the windings can be visualized as being in series with windings without resistance. If this transformer is connected to a voltage source v_s and if a current i is taken from the secondary side (Fig. 32-7), we find, if

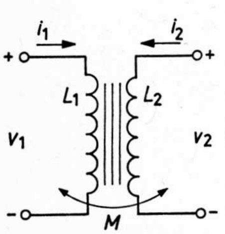

Fig. 32-6

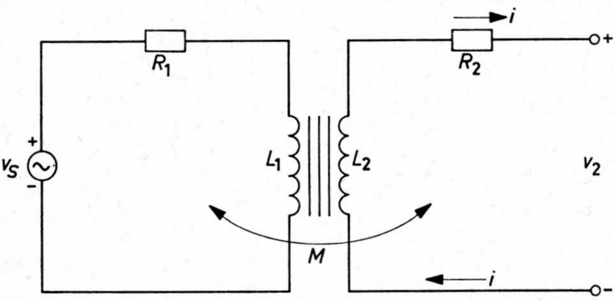

Fig. 32-7

R_1 is the sum of source resistance and the resistance of the primary winding:

$$v_s = (R_1 + j\omega L_1)i_1 - j\omega M i$$
$$v_2 = j\omega M i_1 - (R_2 + j\omega L_2)i$$

which gives:

$$v_2 = \frac{j\omega M}{R_1 + j\omega L_1} v_s - \left(\frac{\omega^2 M^2}{R_1 + j\omega L_1} + R_2 + j\omega L_2\right)i$$

The open circuit voltage is therefore:

$$\frac{j\omega M}{R_1 + j\omega L_1} v_s$$

and the real part of the output impedance

$$\frac{\omega^2 M^2}{R_1^2 + \omega^2 L_1^2} R_1 + R_2$$

Connecting the grid of a valve to the secondary terminals, the voltage here will be

$$v_g = \frac{j\omega M}{R_1 + j\omega L_1} v_s$$

or, the absolute value:

$$\left|\frac{v_g}{v_s}\right| = \frac{\omega M}{\sqrt{R_1^2 + \omega^2 L_1^2}}$$

So that the voltage transformation is not seriously affected by changes in the operating frequency, we choose L_1 such that for the frequencies of interest, and in particular for the lowest of these, applies: $\omega L_1 \gg R_1$, e.g. $\omega L_1 \approx 10R_1$. This gives for the voltage transformation:

$$\left|\frac{v_g}{v_s}\right| \approx k\sqrt{\frac{L_2}{L_1}} \approx k\frac{n_2}{n_1} \approx \frac{n_2}{n_1}$$

where n_1=number of primary turns, and n_2=number of secondary turns. Under the same conditions, we find for the apparent source resistance R_s' on the secondary side:

$$R_s' \approx \frac{n_2^2}{n_1^2} R_1 + R_2$$

By avoiding the use of over-fine wire for the primary windings, we can ensure that the source resistance R_s is predominant in the right-hand term and thus

$$R_s' \approx \frac{n_2^2}{n_1^2} R_s$$

Whether in practice a transformer will be used with a valve amplifier depends in the first instance on the resistance R_s of the voltage source. We can distinguish between the following cases:

1. R_s is smaller than R_{eq}. The term R_{eq}/R_s in the relative noise increase $\sqrt{1+R_{eq}/R_s+R_s/R_{cur}}$, which represents the share of the anode current noise, is then greater than 1. Noise increase will be at least 40 per cent. The complication of inserting a transformer may be worthwhile in this case.
2. If R_s becomes larger than R_{eq}, the profit in using a transformer will diminish gradually. If R_s is at least a few times larger than R_{eq} but smaller than R_{cur}, it will be preferable to connect the source directly without a transformer.

3. If R_s is of the same magnitude or larger than R_{cur}, inserting a transformer would once again becomes useful. However, with a valve circuit, R_{cur} is so great that it is no longer possible to design a transformer with the required properties. It is then better to select a different valve, designed for very low grid current (electrometer valve: $R_{cur} > 10^{11}\ \Omega$).
4. A special case presents itself when R_s is so great that the source can still be considered as a current source when connected directly to the valve (Fig. 32-8). The valve is then subjected to a grid current which equals the signal current, and no longer behaves as a linear amplifier but as one according to the relation between grid current and grid voltage.

   By making the leakage currents and ions current sufficiently small, it is possible to obtain an almost exponential relation between grid voltage and grid current over many decades and therefore also between anode voltage and grid current, so that the output signal is proportional to ln i_s. This circuit can only be used for the measurement of currents and demands many design precautions because of temperature dependence, etc.

   If under these circumstances it is required to obtain linear amplification instead of logarithmic amplification, an additional measurement resistor R is inserted (Fig. 32-9) across which the signal current gives a voltage

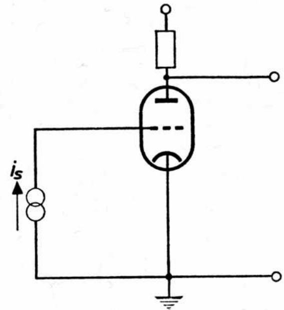

Fig. 32-8

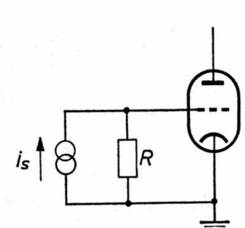

Fig. 32-9

which is great with respect to the noise voltage $\sqrt{4kTR_{eq}\,df}$. The important noise sources are now the thermal noise of resistor R and the grid current noise. The former contributes $\sqrt{4kTR df}$ to the grid voltage, the latter $R\sqrt{4kT df/R_{cur}}$ and together $\sqrt{4kT(R+R^2/R_{cur})df}$. The grid signal voltage is $i_s R$, so that the signal-to-noise ratio is given by:

$$\frac{i_s}{\sqrt{4kT\left(\frac{1}{R}+\frac{1}{R_{cur}}\right)df}}$$

Therefore, we once again select a valve whose value for R_{cur} is as high as possible (electrometer valve) and make R as great as can be justified by insulation, stability and linearity of the circuit. In practice, this value is usually 10^{11}–10^{12} Ω. The fact that such a large resistance with the parasitic capacitances makes the time constant of the input circuit of the order of seconds, is no impediment for following rather rapid phenomena. We mentioned in Section 22 how this resistance can be apparently reduced by means of feedback. It can be seen that this does not affect the signal-to-noise ratio. This situation is met in mass-spectrometers, ionization chambers, photomultipliers and similar instruments.

Let us now look at the case of transistors where, apart from fluctuations in the collector and base currents (which will be denoted by their mean effective values $\sqrt{\overline{i_c^2}}$ and $\sqrt{\overline{i_b^2}}$), we have also to take into account the noise voltage of the base resistance: $\sqrt{\overline{v_{bb}}^2}=\sqrt{4kTR_{bb}\mathrm{d}f}$.

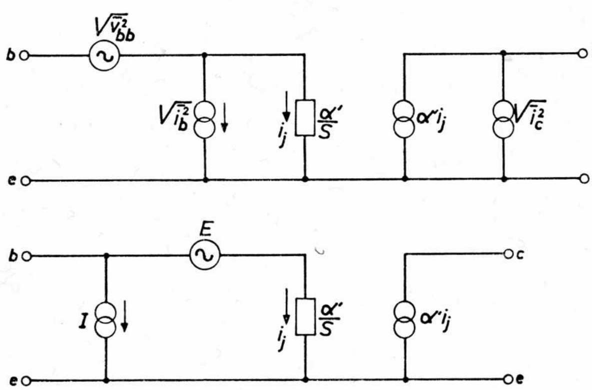

Fig. 32-10

Neglecting collector feedback we can write (upper part of Fig. 32-10):

$$i_b = i_j + \sqrt{\overline{i_b^2}} = \frac{S}{\alpha'}(v_{be} + \sqrt{\overline{v_{bb}^2}}) + \sqrt{\overline{i_b^2}}$$

$$i_c = \alpha' i_j + \sqrt{\overline{i_c^2}} = S\,(v_{be} + \sqrt{\overline{v_{bb}^2}}) + \sqrt{\overline{i_c^2}}$$

Rearranging this as Fig. 32-10, lower part:

$$i_b = \frac{S}{\alpha'}(v_{be} + \sqrt{\overline{E^2}}) + \sqrt{\overline{I^2}}$$

$$i_c = S\,(v_{be} + \sqrt{\overline{E^2}})$$

we find:

$$\sqrt{\overline{E^2}} = \sqrt{\overline{v_{bb}{}^2}} + \frac{\sqrt{\overline{i_c{}^2}}}{S}$$

and $$\sqrt{\overline{I^2}} = \sqrt{\overline{i_b{}^2}} - \frac{\sqrt{\overline{i^2{}_c}}}{\alpha'}$$

As we have seen in the previous section, the equation for the shot noise applies in very good approximation to both collector and base noise currents:

$$\overline{i_c{}^2} = 2\,q\,I_c\,\mathrm{d}f$$

$$\overline{i_b{}^2} = 2\,q\,I_b\,\mathrm{d}f$$

Neglecting the very small correlation between these two noise currents, there will only be non-correlated values in the right-hand sides of the equations for E and I so that

$$\overline{E^2} = \overline{v_{bb}{}^2} + \frac{\overline{i_c{}^2}}{S^2}$$

$$\overline{I^2} = \overline{i_b{}^2} - \frac{\overline{i_c{}^2}}{(\alpha')^2}$$

It follows from the equations for $\overline{i_c{}^2}$ and $\overline{i_b{}^2}$ that $\overline{i_b{}^2} \approx \overline{i_c{}^2}/\alpha'$, which gives for I in good approximation $\overline{I^2} = \overline{i_b{}^2}$, at sufficiently large values of α'. The noise of this current source corresponds to that of a resistance R_{cur}:

$$\frac{4kT}{R_{\mathrm{cur}}} = 2\,q\,I_b$$

or with $I_b = I_c/\alpha'$ and $R_0 = 1/S^* = kT/qI_c \approx 25/I_c(\mathrm{mA})$:

$$R_{\mathrm{cur}} = 2\,\alpha'\,R_0.$$

With $\overline{i_c^2}=2qI_c\mathrm{d}f$ and $S\approx qI_c/kT$, the quotient $\overline{i_c^2}/S^2$ can be written as $2kTR_0\mathrm{d}f$ so that the shot noise in the collector current is accounted for by the equivalent noise resistance $\frac{1}{2}R_0$. This resistance is connected in series with R_{bb} and is, except for very small currents, much smaller, so that the voltage source E will correspond to the thermal noise of resistor R_{bb} (Fig. 32-11).

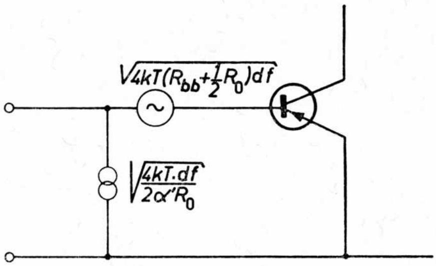

Fig. 32-11

Whilst the ratio R_{cur}/R_{eq} has a value of at least 10^3 with valves, and with certain valves even 10^5, this value is usually much smaller with transistors, namely $2\alpha' R_0/R_{bb}$. For $\alpha'=50$, $R_0=25\ \Omega$, at 1 mA and $R_{bb}=100\ \Omega$ this gives a value of 25. The maximum value, at very small currents, approaches $2\alpha' R_0/\frac{1}{2}R_0=4\alpha'$. Whilst the difference between R_{eq} of a transistor and R_{eq} of a valve is hardly one order of magnitude, the value of R_{cur} of a transistor is many orders of magnitude smaller than that of a valve (10^3–$10^5\ \Omega$ instead of 10^7–$10^{11}\ \Omega$).

This not only results in a much reduced optimum input resistance $R_s{}^*$ – of the order of several hundreds of ohms – but also in the fact that with optimum matching it is usually no longer allowed to neglect the increase in noise above the source noise.

With source resistances of 100 Ω or less, it is sometimes better not to attempt to achieve the best results by means of an impedance-changing transformer because its use may be spoilt over wide frequency ranges by natural resonances. We can then use transistor amplifiers for the amplification of smaller voltages than would be possible with valve amplifiers, because of the lower value of R_{eq}. On the other hand, the effect of the base noise current already becomes noticeable at slightly larger values of R_s (1000 Ω), and the valve is then to be preferred. Contrary to what is relevant in valve circuitry, transformation of a moderately higher source resistance to $R_s{}^*$ by means of a transformer ($n_1>n_2$) is here possible.

The numerical data given above refer to conditions where the emitter current is of the order of 1 mA. With modern Si-transistors, which have still a rather high current amplification factor at very small emitter currents (1 μA) the impedance level, and thus R_{eq} and R_{cur}, is raised considerably. The ratio between R_{eq} and R_{cur} remains unaltered. Although the base current is very small in these transistors, they remain less suitable for the measurement of small direct currents ($< 10^{-10}$ A). In this respect the further development of field-effect transistors is of great importance.

Finally we must take note that the increase in the noise at lower frequencies, due to the excess noise, can be taken into account by increasing R_{eq} and/or reducing R_{cur}. How much this increase has to be is determined by the nature of the noise; in transistors it is mainly an extra current noise that occurs, and this can be accounted for by a reduction of R_{cur}.

33. Resonant circuits

We have so far restricted ourselves to normal a.c. amplifiers which provide the amplification of signals in the frequency band between about 1 c/s and 10^6 c/s without much difficulty. In Section 36 we shall discuss the question of which frequency band is relevant for the best possible amplification of a signal with given time dependence. It will be obvious that in theory an almost infinite number of variations can occur. Rounding off the subject of amplification in measurement electronics it is helpful to pose the following questions:

1. How could one make the passed band narrower?
2. How could one widen this band?

We shall discuss the first question in this section, and the second one in Sections 34 and 35.

The frequency band transmitted by an amplifier can be made narrower at low or high frequencies by inserting a *CR*- or *RC*-circuit (Fig. 33-1). This

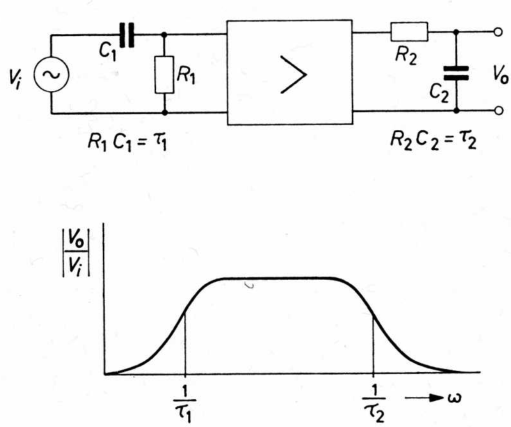

Fig. 33-1

has the advantage that experimental changes can be made easily but the sharpness of the boundaries is not very marked. In addition this method can only be applied when the border frequencies $1/\tau_1$ and $1/\tau_2$ are sufficiently spaced, as otherwise the flat region between these would be spoilt.

Better selectivity and general shaping of the passed band curves can be achieved by using passive and active circuits with a better frequency selectivity.

The simplest example of this is the *resonant circuit*, consisting of the series or parallel combination of a coil and a capacitor (Fig. 33-2).

However, because of the resistance of the coil and possible dielectric losses in the capacitor, we can never realize these ideal resonant circuits in practice. For frequencies in the vicinity of the "resonant frequency", i.e. the frequency of the natural resonance, we can convert the series resistance of coil and capacitor into a parallel resistance, and vice versa. In practice we can therefore calculate in the first instance the behaviour of resonant circuits at these frequencies with the networks shown in Fig. 33-3.

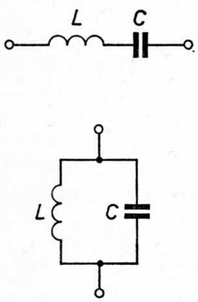

Fig. 33-2

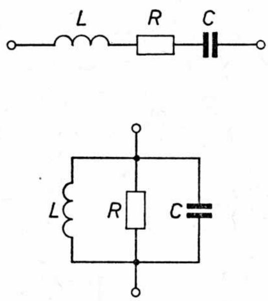

Fig. 33-3

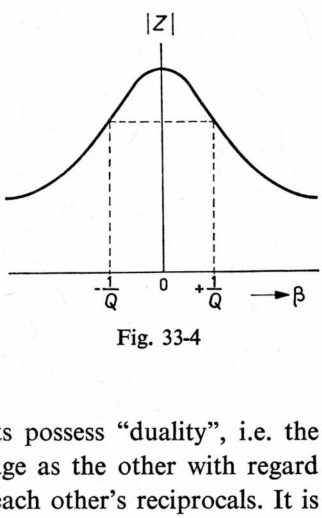

Fig. 33-4

Moreover, series and parallel resonant circuits possess "duality", i.e. the one behaves with regard to current and voltage as the other with regard to voltage and current: their impedances are each other's reciprocals. It is therefore sufficient only to consider one of them. We choose the parallel resonant circuit, because of its greater practical importance.

We obtain for the impedance Z of the parallel resonant circuit of Fig. 33-3:

$$\frac{1}{Z} = \frac{1}{\mathrm{j}\omega L} + \frac{1}{R} + \mathrm{j}\omega C = \frac{R(1-\omega^2 LC) + \mathrm{j}\omega L}{\mathrm{j}\omega LR}$$

hence:

$$Z = \frac{R}{1 + \mathrm{j}\dfrac{R}{\omega L}(\omega^2 LC - 1)}$$

Consequently Z is real and $|Z|$ has a maximum at frequency ω_0, which satisfies $\omega_0^2 LC = 1$ or $\omega_0 = 1/\sqrt{LC}$. For Z we can thus write

$$Z = \frac{R}{1 + \mathrm{j}\,\dfrac{R}{\omega_0 L}\left(\dfrac{\omega}{\omega_0} - \dfrac{\omega_0}{\omega}\right)} = \frac{R}{1 + \mathrm{j}Q\beta} \tag{33.1}$$

where $\beta = \omega/\omega_0 - \omega_0/\omega$, the "relative deviation" with respect to ω_0, and $Q = R/\omega_0 L = R/\sqrt{L/C}$, which is called the *quality factor* of the resonant circuit.

We find for small deviations $\Delta\omega = \omega - \omega_0$ that $\beta \approx 2\,\Delta\omega/\omega_0$, therefore twice what one would call the relative deviation. This latter term, however, is reserved for the more important quantity β.

The quality factor of an ideal resonant circuit ($R = \infty$) is also infinite. Values which can normally be obtained with practical LC circuits are of the order of 100. Physically large coils must be used for achieving this value at low frequencies.

The absolute value of Z follows from (33.1):

$$|Z| = \frac{R}{\sqrt{1 + Q^2\beta^2}} \tag{33.2}$$

Fig. 33-4 shows $|Z|$ against β. The width of this "resonance curve" is given by (33.2): for $\beta = \pm 1/Q$, or when Q is much greater than 1 for $\Delta\omega = \pm\,\omega_0/2Q$, $|Z|$ will have dropped to about 70 per cent of the maximum value, so that ω_0/Q indicates the order of magnitude of the transmitted band. In a practical resonant circuit it is easy to reduce the value of Q by shunting L and C with an additional resistor. Larger values and hence greater sharpness of tuning, can be obtained by positive feedback.

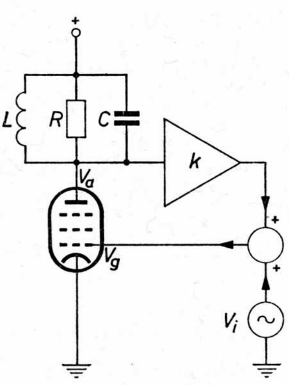

Fig. 33-5

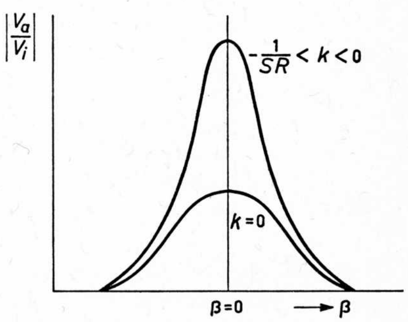

Fig. 33-6

If the following equation applies for the amplification of the pentode in the basic circuit of Fig. 33-5:

$$v_a = -SZ_a\, v_g = -\frac{SR}{1 + \mathrm{j}Q_0\beta} v_g$$

with $Q_0 = R/\omega_0 L$ and $\omega_0 = 1/\sqrt{LC}$, the amplification of the circuit with feedback will be:

$$v_a = -\,SZ_a(v_i + kv_a)$$

so that

$$v_a(1 + \mathrm{j}Q_0\beta + kSR) = -\,SRv_i$$

or

$$\frac{v_a}{v_i} = -\frac{SR}{1 + kSR + \mathrm{j}Q_0\beta} = -\frac{SR}{1 + kSR} \cdot \frac{1}{1 + \mathrm{j}\dfrac{Q_0}{1 + kSR}\beta}$$

Comparing this expression with (33-1) shows that the amplification v_a/v_i exhibits the same frequency dependence as a parallel resonant circuit with a quality factor $Q \doteq Q_0/(1+kSR)$. For $-1/SR < k < 0$, Q is greater than Q_0 and the resonance curve becomes sharper (Fig. 33-6). Amplification increases for all frequencies when $1+kSR$ becomes smaller, but this increase is larger when $|\beta|$ is smaller.

It follows that large amplifications will occur over a narrow frequency band when $(1+kSR)$ is small with respect to 1. However, the stability of the amplification will rapidly decrease in this case, for kSR will have a value of approx. -1, so that relatively small changes in this product will produce relatively large changes in $(1+kSR)$ and therefore in the amplification.

In the special case of $kSR = -1$, the amplification for $\beta = 0$, therefore $\omega = \omega_0$, is infinitely large. This appears to mean that any signal of frequency ω_0, including the internally generated noise signal, will produce an infinitely large output signal. But in reality the output signal obviously cannot exceed a certain value. In order to reach this value caused by the natural noise, kSR must not equal -1, but differ from it by a small amount (in practice of the order of 10^{-8}–10^{-12}), which obviously cannot be obtained by adjustment. When discussing oscillation in Section 37 we shall see that, because of phenomena yet unconsidered, which occur when $(1 + kSR)$ is negative, a stable situation can nevertheless be obtained. The output signal then is a sinusoidal voltage with a frequency $1/\sqrt{LC}$ and almost constant amplitude, though no input signal is present.

It is thus possible by means of resonant circuits, or active circuits containing resonant circuits, to amplify exclusively a narrow band around a central frequency. However, as we have mentioned before, the use of coils will be avoided as much as possible, particularly in measurement circuits for low frequencies. This will become more urgent as the measurement frequency is lower because, in general, inductances of a much larger value become necessary. Particularly at low frequencies it is, however, relatively simple to design active circuits with the selectivity properties of resonant circuits without the use of coils.

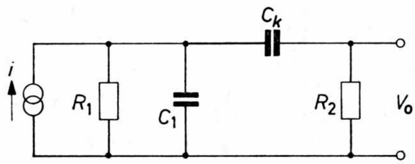

Fig 33-7

In the first place we can start with the above mentioned possibility to increase the quality factor of resonant circuits by means of positive feedback. In this respect there are networks which consist exclusively of resistors and capacitors, but whose frequency characteristics show great similarity to *LC* resonant circuits with a small quality factor. Fig. 33-7 shows an example of such a circuit. The transfer impedance v_o/i of this circuit is easily derived:

$$\frac{v_o}{i} = \frac{\mathrm{j}\omega\, R_1R_2C_k}{1 + \mathrm{j}\omega\{R_1(C_1 + C_k) + R_2C_k\} - \omega^2 R_1R_2C_1C_k}$$

This expression is real at a frequency ω_0, which satisfies

$$\omega_0^2\, R_1R_2\, C_1C_k = 1$$

Introducing this frequency

$$\frac{v_o}{i} = \frac{\mathrm{j}\omega\, R_1R_2C_k}{1 + \mathrm{j}\omega\{R_1(C_1 + C_k) + R_2C_k\} - \dfrac{\omega^2}{\omega_0^2}}$$

$$= \frac{R_1R_2C_k}{R_1(C_1 + C_k) + R_2C_k} \cdot \frac{1}{1 + \dfrac{\mathrm{j}\beta}{\omega_0\{R_1(C_1 + C_k) + R_2C_k\}}} = \frac{R}{1 + \mathrm{j}Q_0\beta}$$

where $$R = \frac{R_1R_2C_k}{R_1(C_1 + C_k) + R_2C_k}$$

and $$Q_0 = \frac{1}{\omega_0\{R_1(C_1 + C_k) + R_2C_k\}} = \frac{\sqrt{R_1R_2C_1C_k}}{R_1(C_1 + C_k) + R_2C_k}$$

The expression $R/(1+\mathrm{j}Q_0\beta)$ for the transfer impedance is the same as that for an LC resonant circuit. However:

$$Q_0^2 + \frac{R^2}{R_1R_2} = \frac{R_1(C_1 + C_k)\cdot R_2C_k}{\{R_1(C_1 + C_k) + R_2C_k\}^2} = \frac{\dfrac{R_1(C_1 + C_k)}{R_2C_k}}{\left\{\dfrac{R_1(C_1 + C_k)}{R_2C_k} + 1\right\}^2}$$

This latter form has a maximum value of $\frac{1}{4}$ for $R_1(C_1+C_k)=R_2C_k$

so that
$$Q_0^2 + \frac{R^2}{R_1R_2} \leqslant \tfrac{1}{4}$$

The maximum attainable value of Q_0 is therefore 0.5, though in practice, we have to be content with a smaller value (≈ 0.4) because the value of the "shunt resistor" R becomes very small for $Q_0 \approx 0.5$.

Inserting the network in a circuit with feedback, will again result in a transfer characteristic of the same form but with a greater Q. Fig. 33-8 shows an example of this. Assuming for the sake of simplicity that the cathodes completely follow the grids, we can write:

$$v_o = \frac{-Ri_1}{1+\mathrm{j}Q_0\beta} \quad \text{and} \quad i_1 = \frac{v_i}{R_k} + \frac{v_i - v_o}{R_0}$$

therefore:
$$\frac{v_o}{v_i} = -\frac{R(R_k + R_0)}{R_k(R_0 - R)}\cdot\frac{1}{1+\mathrm{j}\dfrac{R_0}{R_0 - R}Q_0\beta}$$

which gives for the quality factor of the system:

$$Q = \frac{R_0}{R_0 - R}Q_0$$

A large value for Q can be obtained by making R_0 only slightly larger than R. However, the difference $R_0 - R$ becomes more apparent for variations in the values of R_0 and R as it becomes relatively smaller. The use of high-stability components allows the designer in practice to realize values for Q up to approx. 50. The amplification in this case is quite considerable. At the resonant frequency we find, when $R_0 \approx R$:

$$A_{\beta=0} = \frac{-Q}{Q_0}\left(1 + \frac{R}{R_k}\right)$$

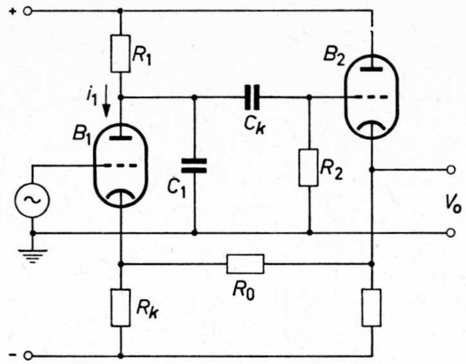

Fig. 33-8

The valve circuit of Fig. 33-8 is particularly suitable for low frequencies. The upper limit is determined by stray capacitances. For reasons yet to be mentioned, it is necessary to make the resistor values fairly large. We therefore do not use this type of filter at frequencies higher than a few kc/s in measurement electronics. The lower limit is determined by those values for which it is possible to obtain components of sufficient stability, and is below 1 c/s when valves are used. When transistors are used the lower impedance levels cause this frequency region to be shifted accordingly upwards.

In reality, however, valves are not ideal cathode followers. The output impedances of the cathodes ($\approx 1/S$) are connected in series with R_0, so that possible variations in the slopes can affect the amount of feedback and therefore the quality factor and the amplification. This effect can be minimized by choosing the largest possible values for the resistors for a given working point. This is also desirable at low frequencies in order to be able to use relatively small capacitances (precision mica types). When large resistors are used, we must take into account several limiting factors. Firstly it is very difficult to manufacture stable resistors of very high values. In order not to make the requirements unnecessarily high, it is recommended that one starts with a circuit for which Q_0 approximates 0.5 rather closely.

By choosing R_1, R_2, C_1 and C_k so that the condition $R_1(C_1 + C_k) = R_2C_k$ is satisfied, $Q_0 = \frac{1}{2}\sqrt{C_1/(C_1 + C_k)}$ and $R = \frac{1}{2}R_1$. The signal transfer in this case is 50 per cent of the theoretical maximum. With $C_1 = C_k$, and therefore $R_2 = 2R_1$, Q_0 will be about 0.35. By making $C_1 = 4C_k$ and thus $R_2 = 5R_1$, Q_0 will be about 0.45.

Particularly at very low frequencies, we sometimes feel the need to select larger values for R_1 and R_2 than possible or allowable in the circuit. However, it is often possible to solve this difficulty by connecting the anode and/or the grid to tappings of R_1 and R_2 respectively. This is indicated in Fig. 33-9 for the anode.

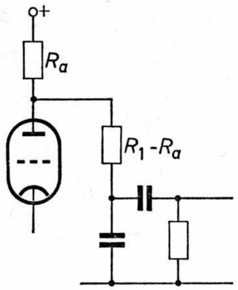

Fig. 33-9

Finally, we must take into account that the input capacitance of the right-hand valve B_2 is connected in parallel to R_2. When R_2 is large, or the resonant frequency of the filter relatively high, there is a possibility that the impedance of this input capacitance can no longer be neglected. Since such a capacitance

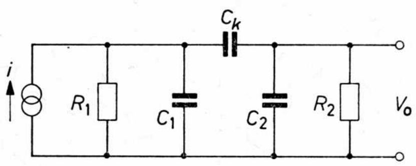

Fig. 33-10

has little stability, inadmissible fluctuations can occur. Apart from reducing R_2, the effect of this capacitance can also be minimized by inserting a sufficiently large and stable capacitor parallel to R_2 (C_2 in Fig. 33-10).

In this case we have for the filter:

$$\omega_0^2 R_1 R_2 (C_1 C_2 + C_1 C_k + C_2 C_k) = 1 \ ,$$

$$R = \frac{R_1 R_2 C_k}{R_1(C_1 + C_k) + R_2(C_2 + C_k)} \ ,$$

$$Q_0 = \frac{1}{\omega_0 \{R_1(C_1 + C_k) + R_2(C_2 + C_k)\}} \ ,$$

so that once again:

$$Q_0^2 + \frac{R^2}{R_1 R_2} \leqslant \tfrac{1}{4} \ .$$

The equality sign is valid for $R_1(C_1 + C_k) = R_2(C_2 + C_k)$. If this condition is satisfied, we obtain:

$$R = \tfrac{1}{2} R_1 \cdot \frac{C_k}{C_2 + C_k}$$

and
$$Q_0 = \tfrac{1}{2} \cdot \sqrt{1 - \frac{C_k^2}{(C_1 + C_k)(C_2 + C_k)}}$$

In the "symmetrical" network, $R_1 = R_2$ and $C_1 = C_2$, we obtain a good compromise for $C_k/C_1 = 3/2$, $R = 0.3R_1$ and $Q_0 = 0.4$, whilst $\omega_0 = 1/2R_1C_1$.

The greatest impediment to obtaining sufficiently high stability at high values of Q is caused in the circuit of Fig. 33-8 by the effect of variations in the slope of the valves. Marked improvement can be achieved by considering circuits by which a different method of feedback is applied. Using a separate valve for the feedback (B_3 in Fig. 33-11) gives such an improvement.

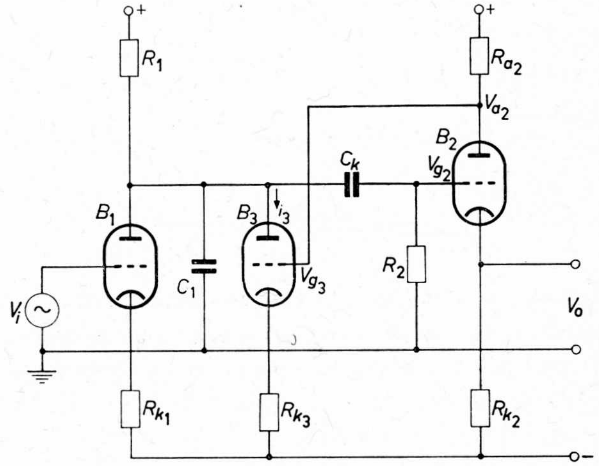

Fig. 33-11

The degree of feedback is then determined in the first instance by resistors R_{a2}, R_{k2} and R_{k3}:

$$i_3 = -\frac{R_{a2}}{R_{k2}R_{k3}} v_o$$

This approximation is better satisfied when the reciprocal values of the slopes of B_2 and B_3 are small with respect to R_{k2} and R_{k3}. Since there is no objection to making these resistors larger than R_0 in Fig. 33-8, the effect of variations in slope will be correspondingly smaller.

An even greater improvement is effected by inserting still another valve (B_4 in Fig. 33-12) and by applying the principle of "adding what is lacking". Assuming the valves B_2 and B_3 to act as ideal cathode followers, we have $v_{k3} = v_{g3} = v_{a2} = -(R_{a2}/R_{k2})v_o = -(R_{a2}/R_{k2})v_{g2}$. If resistors R_2 and R_3 are chosen such that $R_2 : R_3 = R_{k2} : R_{a2}$, then there will be no signal voltage at

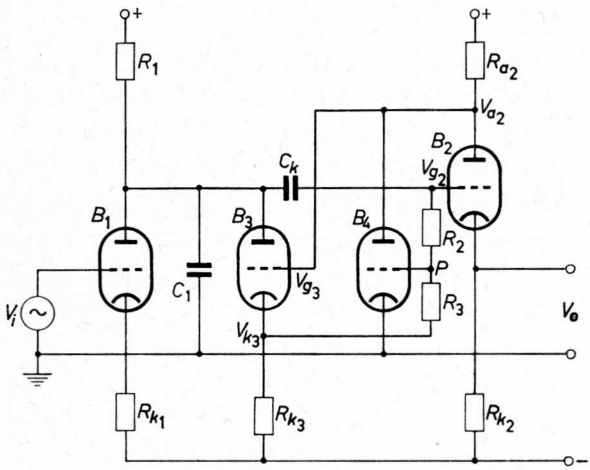

Fig. 33-12

point P with respect to earth, and therefore no signal current will pass through B_4. The operation of the circuit is then exactly the same as that of Fig. 33-11; the only difference is that in terms of signal current, the parallel network of R_{k3} and R_3 will now be present in the cathode circuit of B_3, instead of R_{k3} alone. A deviation from this situation, which may be caused by slope variations, produces a signal voltage at P, which is opposed however by the feedback through B_3. Voltages and currents will therefore always adjust themselves such that there is almost no signal voltage at P. The degree of feedback is then solely determined by resistors R_2, R_3 and R_{k3}:

$$i_3 = v_{k3}\left(\frac{1}{R_3} + \frac{1}{R_{k3}}\right) = -\,v_{g2}\left(\frac{1}{R_2} + \frac{R_3}{R_2 R_{k3}}\right)$$

Active resonant circuits based on the principle of Fig. 33-8 offer the great advantages of simplicity and versatility. Disadvantages are that the value of

Q is obtained by feedback, so that the danger of oscillation often becomes intolerably large above a certain value (≈ 50), and also, for many applications, that the amplification at the central frequency changes proportionally with the adjustment of Q. These disadvantages are absent in an other way of synthetizing resonant circuits without coils. This is achieved by inserting selective circuits in the feedback loop of feedback amplifiers. Such a circuit must give little feedback at the central frequency, i.e. zero transfer at this central frequency, but passing any other frequency. The danger of oscillation can easily be prevented here, even if the loop gain is made so high that very high quality factors are reached (≈ 1000). A disadvantage is that the adjustment of circuits which have a zero transfer for a certain frequency is rather critical, while at the same time the requirements for the stability of the components are very stringent.

The best-known circuit which has a zero transfer at a certain frequency is the "twin-T" network (Fig. 33-13). Its properties are usually calculated by applying the star-delta transformation.

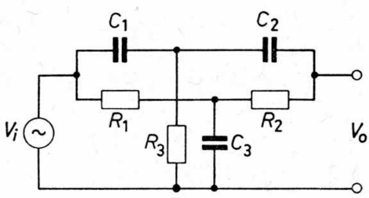

Fig. 33-13

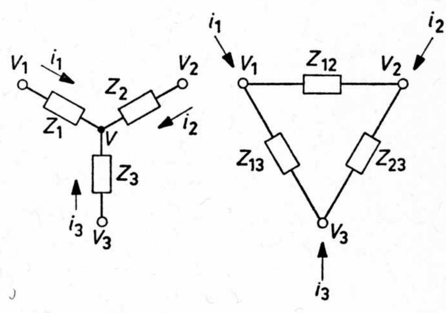

Fig. 33-14

We have for the star circuit (left-hand side of Fig. 33-14):

$$i_1 = \frac{v_1 - v}{Z_1}, \quad i_2 = \frac{v_2 - v}{Z_2}, \quad i_3 = \frac{v_3 - v}{Z_3}, \quad i_1 + i_2 + i_3 = 0$$

so that $$v\left(\frac{1}{Z_1} + \frac{1}{Z_2} + \frac{1}{Z_3}\right) = \frac{v_1}{Z_1} + \frac{v_2}{Z_2} + \frac{v_3}{Z_3}$$

or $$v\,(Z_1Z_2 + Z_1Z_3 + Z_2Z_3) = Z_2Z_3\,v_1 + Z_1Z_3\,v_2 + Z_1Z_2\,v_3$$

Provided $$Z_1Z_2 + Z_1Z_3 + Z_2Z_3 \neq 0:$$

$$v = \frac{Z_2Z_3\,v_1 + Z_1Z_3\,v_2 + Z_1Z_2\,v_3}{Z_1Z_2 + Z_1Z_3 + Z_2Z_3}.$$

Substituting this value for v in $i_1 = (v_1 - v)/Z_1$:

$$i_1 = \frac{(v_1 - v_2)Z_3}{Z_1Z_2 + Z_1Z_3 + Z_2Z_3} + \frac{(v_1 - v_3)Z_2}{Z_1Z_2 + Z_1Z_3 + Z_2Z_3}$$

Since expressions with the same terms are obtained for i_2 and i_3, it follows that the external behaviour of the star circuit of the left-hand side of Fig. 33-14 is identical to that of the delta circuit of the right-hand side of Fig.33-14, provided

$$Z_{12} = \frac{Z_1Z_2 + Z_1Z_3 + Z_2Z_3}{Z_3} = Z_1 + Z_2 + \frac{Z_1Z_2}{Z_3}$$

etc.

It is interesting to note that Z_{12}, Z_{13} and Z_{23} do not have to be impedances which can be realized themselves. In that case the transformation is a purely mathematical operation, which can however greatly facilitate the calculations involved for some circuits.

We can also transform the delta circuit into a star circuit, provided $Z_{12} + Z_{13} + Z_{23} \neq 0$. We then find:

$$Z_1 = \frac{Z_{12}Z_{13}}{Z_{12} + Z_{13} + Z_{23}}$$

etc.

In this case also, the resultant impedances themselves do not have to be realizable.

A few words about the additional conditions. If one of the terms $Z_1Z_2 + Z_1Z_3 + Z_2Z_3$ and $Z_{12} + Z_{13} + Z_{23}$ proves to be zero, the corresponding transformation gives a result which should not be interpreted in too strict a fashion, because the simultaneous condition of impedances becoming zero or infinite must not be considered as so many short- or open-circuits. For example, the following equations apply to an ideal transformer:

$$v_1 = \mathrm{j}\omega L_1 i_1 + \mathrm{j}\omega M i_2 = \mathrm{j}\omega(L_1 - M)i_1 + \mathrm{j}\omega M(i_1 + i_2)$$

$$v_2 = \mathrm{j}\omega M i_1 + \mathrm{j}\omega L_2 i_2 = \mathrm{j}\omega M(i_1 + i_2) + \mathrm{j}\omega(L_2 - M)i_2$$

where $M = \sqrt{L_1L_2}$, so that the transformer on the left-hand side of Fig. 33-15 is identical with the star circuit in the centre of Fig. 33-15. When transforming these into a delta circuit according to the above relations as shown on the right-hand side of Fig. 33-15, we find for all elements the value zero. It is obvious that this should not be interpreted as three short-circuits.

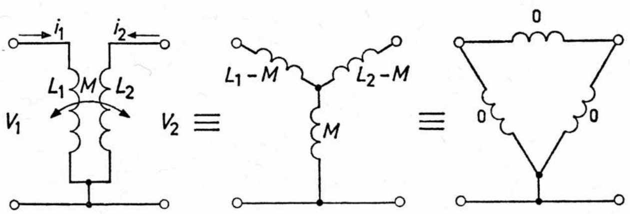

Fig. 33-15

The twin-T filter of Fig. 33-13 can now be considered as the parallel circuit of two star circuits. By applying the star-delta transformation to each of these, we obtain the result shown in Fig. 33-16. The impedance between points 1 and 2 consists of a parallel connection of $Z_1=R_1+R_2+j\omega R_1R_2C_3$ and $Z_2 = 1/j\omega C_1+1/j\omega C_2-1/\omega^2R_3C_1C_2$ and therefore has the value $Z_1Z_2/(Z_1+Z_2)$, where we can write for the denominator:

$$Z_1 + Z_2 = R_1 + R_2 - \frac{1}{\omega^2 R_3 C_1 C_2} + j\left(\omega R_1 R_2 C_3 - \frac{C_1 + C_2}{\omega C_1 C_2}\right).$$

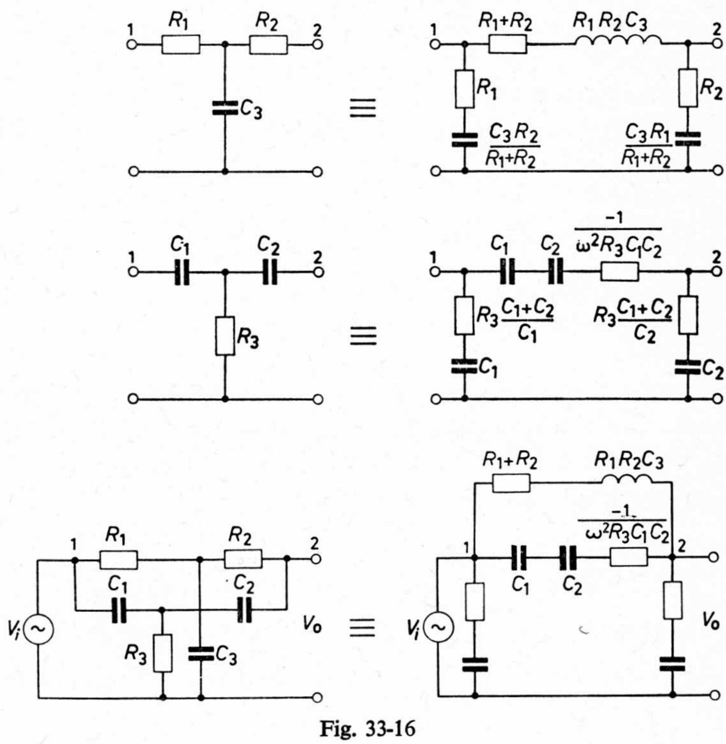

Fig. 33-16

By selecting the components such that for frequency ω_0

$$R_1 + R_2 = \frac{1}{\omega_0{}^2 R_3 C_1 C_2} \quad \text{and} \quad \omega_0 R_1 R_2 C_3 = \frac{C_1 + C_2}{\omega_0 C_1 C_2}$$

Z_1+Z_2 will be zero for this frequency, and the impedance between points 1 and 2 will be infinitely large, i.e. $v_o/v_i=0$. The most common filter which satisfies these conditions is the symmetrical twin-T filter of Fig. 33-17. The following applies to its transfer characteristic:

$$\frac{v_o}{v_i} = \frac{\mathrm{j}\beta}{4 + \mathrm{j}\beta}$$

where $$\beta = \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega} \text{ and } \omega_0 = \frac{1}{RC}.$$

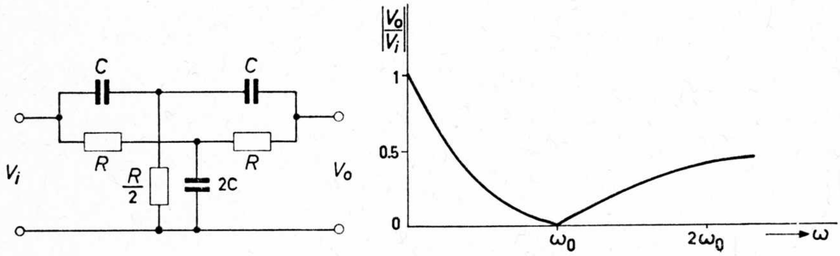

Fig. 33-17

This transfer corresponds to that of a voltage divider, consisting of an ideal resonant circuit and a resistor (Fig. 33-18) for which $R = \frac{1}{4}\sqrt{L/C}$. Using this filter as feedback network in an amplifier with an amplification A_0, the amplification with feedback will be:

$$A_{fb} = \frac{A_0}{1 + A_0 \dfrac{\mathrm{j}\beta}{4 + \mathrm{j}\beta}} = \frac{A_0(1 + \frac{1}{4}\mathrm{j}\beta)}{1 + \frac{1}{4}(A_0 + 1)\mathrm{j}\beta}$$

which for small values of β can be approximated to:

$$A_{fb} = \frac{A_0}{1 + \mathrm{j}\dfrac{A_0}{4}\beta}$$

Thus, the frequency response of this circuit behaves almost like a resonant circuit with a quality factor $\frac{1}{4} A_0$. In this case, the limit is determined by the stability of the filter components. Practical designs with $Q \approx 1000$ are quite feasible.

A very constant value of Q can be obtained by making A_0 extremely

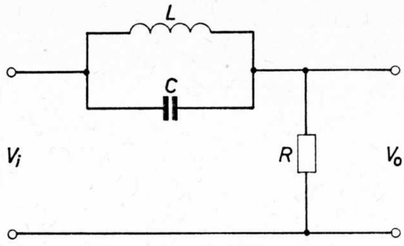

Fig. 33-18

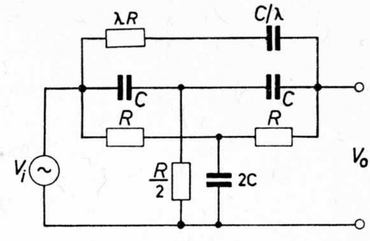

Fig. 33-19

constant by means of feedback. When using the voltage divider of Fig. 33-18, this would be achieved by connecting a resistor across the resonant circuit. With the twin-T filter, we must insert the series combination of a resistor and a capacitor between points 1 and 2 (Fig. 33-19). We obtain a value of $Q \ll \frac{1}{4}A_0$ with the symmetrical filter by making the resistor $2QR$ and the capacitor $C/2Q$. The amplification at the central frequency is then $4Q$.

Apart from the twin-T filter, there are many other RC-networks for which the transfer function is zero at a certain frequency. An interesting problem is to consider whether it is possible here to alter this central frequency by means of a single variable component.

This is possible with the twin-T filter by converting resistor $\frac{1}{2}R$ in Fig. 33-17 into a potentiometer. We then obtain the Andreyev circuit shown in Fig. 33-20, where

$$\omega_0 = \frac{1}{RC\sqrt{1 - \alpha^2}}$$

in which α is the fraction of the potentiometer track between the wiper and capacitor $2C$.

A different circuit with the same property is that by Hall (Fig. 33-21), where

$$\omega_0 = \frac{1}{RC\sqrt{1 + 2\lambda} \cdot \sqrt{\alpha(1 - \alpha)}}$$

when α is the fraction of the potentiometer track between the wiper and one of the terminals.

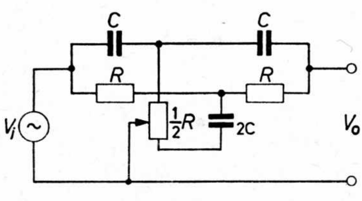

Fig. 33-20

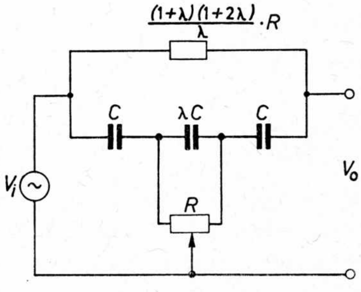

Fig. 33-21

It is often necessary to replace the bell-shaped transfer characteristic $|v_o/v_i| = A/(\sqrt{1+Q^2\beta^2})$ by a more rectangular one. One method of approximating to the latter is by means of the characteristic:

$$\left|\frac{v_o}{v_i}\right| = \frac{A}{\sqrt{1 + (Q\beta)^{2n}}} \tag{33.3}$$

This approximation will be best when the integer n is larger (Butterworth approximation). Such a characteristic can be obtained by means of n resonant circuits connected in cascade, with the quality factors and resonant frequencies having certain values. Hence the term "staggered tuning".

As we shall see below, this possibility generally exists for those cases where the form under the root sign contains terms in β with even exponents $\leqslant 2n$, and is positive for all values of ω, as otherwise the amplitude would not be real. This statement is in fact a special case of a more general statement regarding the possibility of practical realization of transfer functions of the form $\sqrt{P_1/P_2}$, where P_1 and P_2 are positive definite, even polynomials in ω.

Restricting ourselves to the case in hand, we can derive the quality factors and the central frequencies of each of the n resonant circuits as follows.

With the prescribed form of the modulus (33.3) goes a transfer of the type:

$$\frac{v_o}{v_i} = \frac{A}{1 + ax + bx^2 + \ldots + x^n} \tag{33.4}$$

where $x = \mathrm{j}Q\beta$.

We can write for the denominator:

$$1 + ax + bx^2 + \ldots + x^n = (x - x_1)(x - x_2)\ldots(x - x_n)$$

where $x_1, x_2 \ldots x_n$ are the zero's of $1 + ax + bx^2 + \ldots + x^n$. Since x is purely imaginary, we have $|x - x_k|^2 = -(x - x_k)(x + x_k^*)$, where x_k^* is the complex conjugate of x_k (left of Fig. 33-22). We can then write for the square of the modulus of the above form:

$$(x_1 - x)(x_2 - x)\ldots(x_n - x)(x + x_1^*)(x + x_2^*)\ldots(x + x_n^*).$$

It is necessary that this equals $1 + (Q\beta)^{2n} = 1 + (-\mathrm{j}x)^{2n}$ and hence the zero's of both forms must coincide. Those of the latter are equidistantly distributed over the unit circle in the complex plane (centre and right of Fig. 33-22). In order to obtain a system with damped natural modes (otherwise it would

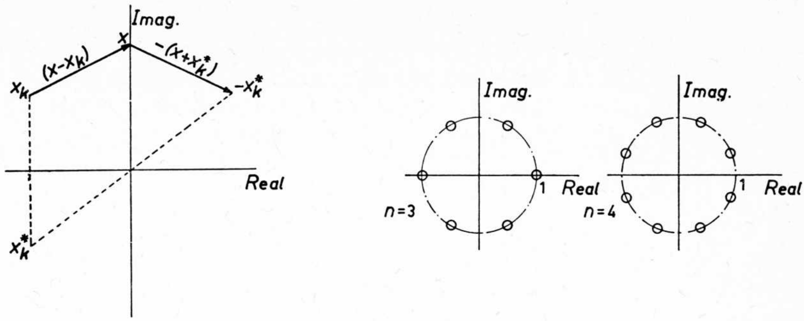

Fig. 33-22

oscillate) $x_1, x_2 \ldots x_n$ must have a negative real part, and therefore be those zero points of $1+(-\mathrm{j}x)^{2n}$ which lie in the left half-plane.

Apart from the positive factor Q, x equals $\mathrm{j}\beta = (\mathrm{j}\omega/\omega_0 + \omega_0/\mathrm{j}\omega)$. This quantity corresponds to $(p + 1/p)$, ignoring the positive factor ω_0, where $p = \mathrm{d}/\mathrm{d}t$. If the natural frequencies p_i have a negative real part (which is necessary for a stable system), then this will also apply to $p_i + 1/p_i$ and therefore to x_i.

We thus derive for the denominator of (33-4):

$$(x+1)\prod_{k=1}^{\frac{n-1}{2}}\left(x+\mathrm{e}^{\mathrm{j}\frac{k\pi}{n}}\right)\left(x+\mathrm{e}^{-\mathrm{j}\frac{k\pi}{n}}\right) \quad \text{for } n = \text{odd}$$

and

$$\prod_{k=1}^{\frac{n}{2}}\left(x+\mathrm{e}^{\mathrm{j}\frac{2k-1}{2n}\pi}\right)\left(x+\mathrm{e}^{-\mathrm{j}\frac{2k-1}{2n}\pi}\right) \quad \text{for } n = \text{even.}$$

The factor $x+1=1+\mathrm{j}Q\beta$ can be realized by means of a resonant circuit which is tuned to the central frequency ω_0 and which has a quality factor Q.

The factors of the form $x+\mathrm{e}^{j\phi}$ cannot be realized individually, but only the product

$$(x + \mathrm{e}^{\mathrm{j}\phi})(x + \mathrm{e}^{-\mathrm{j}\phi}) = (\mathrm{j}Q\beta + \mathrm{e}^{\mathrm{j}\phi})(\mathrm{j}Q\beta + \mathrm{e}^{-\mathrm{j}\phi}) = 1 + 2\mathrm{j}Q\beta\cos\varphi - Q^2\beta^2 \tag{33.5}$$

This appears to correspond to two resonant circuits with the same quality factors, which are tuned to frequencies which are a certain factor larger or smaller than the central frequency. The denominator of the frequency characteristic of two cascade-connected resonant circuits having the same quality factor Q_k and tuned to frequencies $\alpha\omega_0$ and ω_0/α respectively, will be:

$$(1 + jQ_k\beta_1)\,(1 + jQ_k\beta_2),$$

where $\beta_1 = \omega/a\omega_0 - a\omega_0/\omega$ and $\beta_2 = a\omega/\omega_0 - \omega_0/a\omega$.

Since $\beta_1 + \beta_2 = (a + 1/a)\,(\omega/\omega_0 - \omega_0/\omega) = (a + 1/a)\beta$

and $\beta_1\beta_2 = (\omega/a\omega_0 - a\omega_0/\omega)\,(a\omega/\omega_0 - \omega_0/a\omega) = \beta^2 - (a - 1/a)^2$ we find $(1 + jQ_k\beta_1)\,(1 + jQ_k\beta_2) = 1 + jQ_k(\beta_1 + \beta_2) - Q_k{}^2\beta_1\beta_2 = 1 + j(a + 1/a)\,Q_k\beta - Q_k{}^2\beta^2 + Q_k{}^2(a - 1/a)^2 =$

$$\left\{1 + Q_k{}^2\,(a - 1/a)^2\right\}\left\{1 + j\,\frac{(a + 1/a)Q_k\beta}{1 + Q_k{}^2\,(a - 1/a)^2} - \frac{Q_k{}^2\beta^2}{1 + Q_k{}^2\,(a - 1/a)^2}\right\}$$

By equating the last factor to (33.5) we find:

$$\frac{Q_k{}^2}{1 + Q_k{}^2\,(a - 1/a)^2} = Q^2 \quad\text{and}\quad \frac{(a + 1/a)Q_k}{1 + Q_k{}^2\,(a - 1/a)^2} = 2Q\cos\varphi$$

It follows from the first equality:

$$1 + Q_k{}^2\,(a - 1/a)^2 = \left\{1 - Q^2(a - 1/a)^2\right\}^{-1}$$

which, together with the second, yields:

$$(a + 1/a)^2\left\{1 - Q^2\,(a - 1/a)^2\right\} = 4\cos^2\varphi \qquad (33.6)$$

a can be calculated from this equation. (It can be seen easily that there is always a positive root a.) We then obtain Q_k from:

$$Q_k = (a + 1/a)\,\frac{Q}{2\cos\varphi} \qquad (33.7)$$

As $1 - Q^2(a - 1/a)^2$ must be positive according to (33.6), $(a - 1/a)$ will be small compared to 1, and therefore $a \approx 1$ and $a + 1/a$ will equal 2 in very good approximation in the case of relatively small filters, thus when Q is large. Equations (33.6) and (33.7) can then be approximated by

$$a - \frac{1}{a} = \frac{\sin\varphi}{Q} \quad\text{and}\quad Q_k = \frac{Q}{\cos\varphi} \qquad (33.8)$$

In most practical cases we take $n = 3$, whilst $n > 5$ is hardly ever applied because of the little gain as opposed to the complexity of adjustment.

Example: We require a response with a central frequency of 800 c/s involving three resonant circuits, such that the drop for a deviation of 20 c/s is 1 per cent.

For a deviation of 20 c/s, $\beta = 0.05$, and for this value of β we find therefore that the following must apply: $\sqrt{1 + (Q\beta)^6} = 1.01$. It follows that $Q \approx 10$, i.e. a single resonant circuit must be tuned to 800 c/s and possess a quality factor of 10. With $n = 3$, $\varphi = 60°$ so that, according to (33.8), $a - 1/a = \sqrt{3/20} = 0.086$, hence $a = 1.043$. The other circuits must therefore be tuned to $800 \times 1.043 = 835$ c/s and $800/1.043 = 765$ c/s. Their quality factor must be $10/\cos 60° = 20$. Fig. 33-23 gives the practical result of a design based on this calculation.

The frequency response of two staggered tuned circuits can also be obtained in very good approximation, when using resonant circuits incorporating coils, by tuning two circuits to the central frequency and adding inductive or capacitive coupling. This is one of the usual methods applied in telecommunication techniques, for example for the intermediate frequency

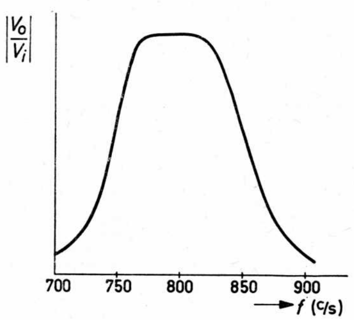

Fig. 33-23

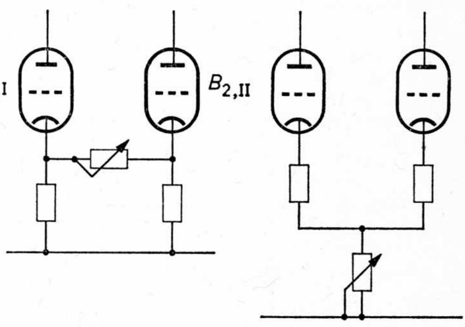

Fig. 33-24

band filters in radio receivers. A similar method can also be applied to active resonant circuits not using coils, and has the advantage of easier tuning because all resonant circuits can then be tuned to the same frequency.

It is then possible to write the form $1+2jQ\beta \cos\varphi - Q^2\beta^2$ from (33.5) as $\sin^2\varphi + (\cos\varphi + jQ\beta)^2 = \cos^2\varphi\left\{\tan^2\varphi + (1 + jQ\beta/\cos\varphi)^2\right\}$. Following the principle of Fig. 33-8 for making two filters with quality factors $Q/\cos\varphi$ and placing them in cascade, the transfer function will be:

$$A = \frac{A_0}{\left(1 + j\dfrac{Q}{\cos\varphi}\beta\right)^2}$$

If we insert in this system frequency-independent feedback from output to input with a feedback factor k_0, the new transfer function will be:

$$A_{fb} = \frac{A}{1 + Ak_0} = \frac{A_0}{\left(1 + j\,\dfrac{Q}{\cos\varphi}\,\beta\right)^2 + A_0 k_0}.$$

By adjusting k_0 so that $A_0 k_0 = \tan^2 \varphi$, the required frequency characteristic is obtained. Feedback can be arranged by inserting a resistor between the second cathode of the second resonant circuit, and the first cathode of the first resonant circuit (left-hand side of Fig. 33-24). It will normally be necessary in practice to give this resistor a very high value, resulting in difficult tuning; an easier solution is obtained by inserting a single adjustable common resistor in both cathode circuits. This is the star-delta transformation of both cathode resistors and the feedback resistor (right-hand side of Fig. 33-24). The twin-resonant circuit is shown in Fig. 33-25. Adjustments are very simple indeed. Both filters are first tuned to the cen-

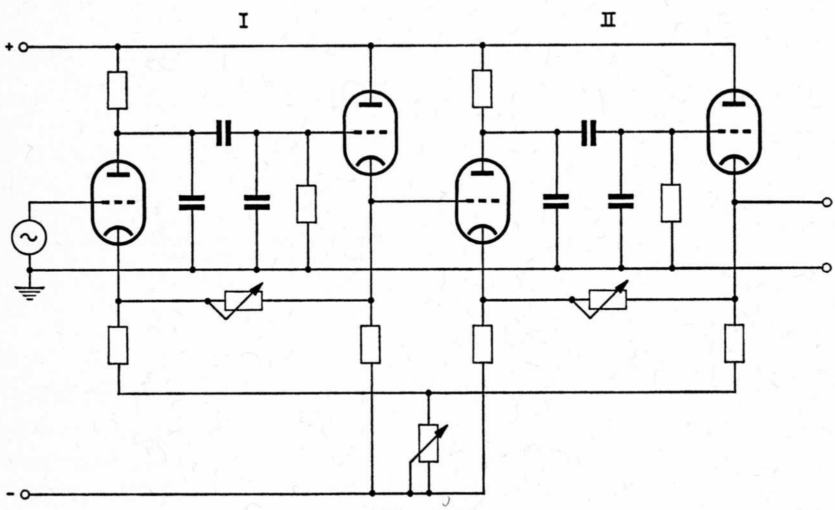

Fig. 33-25

tral frequency without feedback (i.e. without common cathode resistor). The most accurate way of doing this is by making the phase difference between input and output signal zero at this frequency, and by bringing their quality factor to the correct value by adjusting the feedback resistors. The latter adjustment can be checked by measuring the amplification. We then measure the total amplification A_0 of both filters and introduce so much overall feedback that the amplification is reduced to $A_0/(1 + \tan^2\varphi) = A_0 \cos^2\varphi$.

The filters discussed in this section are used in amplifiers when it is necessary to amplify a narrow band. This is the case when the measuring signal only consists of frequencies within such a narrow band and one wants to minimize the interference. Still greater bandwidth restriction than is obtainable with the "selective amplifiers" as described above, can be achieved by applying synchronous detection, which will be discussed later (Section 40).

34. Wide-band amplifiers

In the previous section we discussed some methods for restricting the transmitted bandwidth. We shall now see how we can extend the passed band to higher and lower frequencies. We shall begin this section by discussing the field involving extension to higher frequencies. Because of the limited application of these special techniques in measurement electronics, we shall restrict ourselves to indicating the difficulties and possible methodes. We shall also indicate in this section how limitation of the band to the lower frequencies can be obtained for a.c. amplifiers, whilst the problems met with d.c. amplifiers will be discussed in the next section.

Large amplification with a flat frequency characteristic over a frequency range of 1 c/s to 10^5–10^6 c/s is possible with ordinary a.c. amplifiers using valves, without excessive complication. Considerably higher frequencies can be amplified over a relatively narrow band by the application of resonant circuits. The use of staggered tuning can widen the frequency band over which the transmission characteristic is flat. For example, this solution allows good amplification of the band between 10 and 30 Mc/s. However, the ratio of the highest to the lowest transmitted frequency is then always relatively small, i.e. of the order of magnitude 1. With this method it is, however, not possible to amplify a relatively wide band starting at low frequencies, possibly zero frequency, and extending to several tens of Mc/s.

The very high cut-off frequency of modern h.f. transistors makes high amplification over wider bands more likely with transistors than with valves. This sometimes makes a relatively simple solution possible.

However, in some cases we shall have to use special techniques developed for the amplification of a relatively wide band. Since these are based on certain properties of cables and ladder networks, we shall derive these first of all.

A cable consists of two or more parallel conductors, separated by a dielectric. In the case of two conductors, one is often arranged concentrically around the other. We call this a coaxial cable. The outer conductor is called the sheath. Twin screened cables are also used rather frequently, which have two separate conductors inside the sheath.

Strictly speaking, calculations of the properties of cables should be based on Maxwell's theory. This leads to a complicated analysis, one of the results being that, as long as wavelength λ corresponding to a signal frequency f ($\lambda = c/f$, where $c =$ velocity of light) is large, compared to the diameters and mutual separation between the conductors, the behaviour of the cable can be accurately described by imagining the conductors to consist of uniformly distributed inductance and resistance, and uniformly distributed capacitance and conductance between the conductors. A small section dx of a cable with two conductors can therefore be represented by the circuit shown in Fig. 34-1. L_1 and L_2 account for the effect of a magnetic field in the conductors, and L_{12} for that of a magnetic field between the conductors. Due to

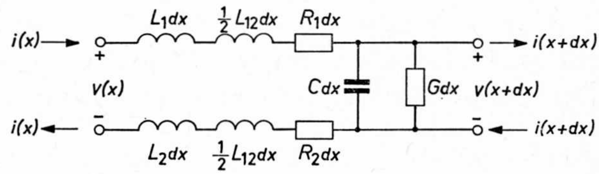

Fig. 34-1

the "skin effect", L_1 and L_2 decrease at high frequencies, whilst R_1 and R_2 increase. The losses of the dielectric which are accounted for in the conductance G, increase at high frequencies. In the following calculations we shall always consider the situation at a single fixed frequency, so that all parameters corresponding to that frequency will be constant. Analogous to its practical application, the cable will be considered as the connection between two circuits, so that, in each cross-section, the current passing through one conductor can be assumed to be equal and opposite to that passing through the other conductor. We shall now see how the amplitudes of the voltages and currents, which vary in a sinusoidal fashion with time, depend on the point of observation in the cable. If ω be the signal frequency, we have in that case the following equations for current $i(x)$ and the voltage between the conductors $v(x)$:

$$i(x + \mathrm{d}x) = i(x) - \mathrm{j}\omega C v(x + \mathrm{d}x)\mathrm{d}x - G v(x + \mathrm{d}x)\mathrm{d}x$$

and $$v(x + \mathrm{d}x) = v(x) - \mathrm{j}\omega(L_1 + L_2 + L_{12})i(x)\mathrm{d}x - (R_1 + R_2)i(x)\mathrm{d}x$$

With $L = L_1 + L_2 + L_{12}$, $R = R_1 + R_2$, $i(x + \mathrm{d}x) = i(x) + \dfrac{\mathrm{d}i}{\mathrm{d}x}\mathrm{d}x$

and $v(x + \mathrm{d}x) = v(x) + \dfrac{\mathrm{d}v}{\mathrm{d}x}\mathrm{d}x$, this becomes

$$\frac{\mathrm{d}i(x)}{\mathrm{d}x} = -\,(G + \mathrm{j}\omega C)\,v(x) \tag{34.1}$$

and
$$\frac{\mathrm{d}v(x)}{\mathrm{d}x} = -(R + \mathrm{j}\omega L)\, i(x) \tag{34.2}$$

Eliminating i from both equations:

$$\frac{\mathrm{d}^2 v(x)}{\mathrm{d}x^2} = (R + \mathrm{j}\omega L)\,(G + \mathrm{j}\omega C)\, v(x)$$

or
$$\frac{\mathrm{d}^2 v(x)}{\mathrm{d}x^2} = \gamma^2 v(x) \tag{34.3}$$

where γ = "propagation factor" $= \sqrt{(R+\mathrm{j}\omega L)(G+\mathrm{j}\omega C)}$. Because of the requirement that the dissipation in the system must be positive, the root with positive real part must be used here.

By solving $v(x)$ from (34.3) and $i(x)$ from (34.2), we find:

$$v(x) = A\mathrm{e}^{-\gamma x} + B\mathrm{e}^{+\gamma x} \tag{34.4}$$

and
$$i(x) = \frac{1}{Z_0}\,(A\mathrm{e}^{-\gamma x} - B\mathrm{e}^{+\gamma x}) \tag{34.5}$$

where $Z_0 = \sqrt{(R+\mathrm{j}\omega L)/(G+\mathrm{j}\omega C)}$, with positive real part; it is called the "characteristic impedance" of the cable.

The integration constants A and B are determined by the situation found at the beginning and the end of the cable. It will, of course, often happen that the beginning of the cable is connected to a voltage source v_s with internal impedance Z_s, whilst the end is connected to a circuit with input impedance Z (Fig. 34-2).

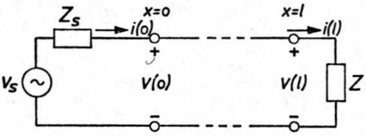

Fig. 34-2

At the beginning, where $x = 0$, we have:

$$v(0) = v_s - Z_s\, i(0)$$

and at the end, where $x = l$:

$$v(l) = Z\, i(l)$$

Substitution in these equations of the relations following from (34.4) and (34.5): $v(0)=A+B$; $i(0)=(A-B)/Z_0$, $v(l)=A\mathrm{e}^{-\gamma l}+B\mathrm{e}^{+\gamma l}$ and $i(l)=(A\mathrm{e}^{-\gamma l}-B\mathrm{e}^{+\gamma l})/Z_0$ yields two equations from which A and B can be solved. With these values (34.4) and (34.5) become:

$$v(x)=\frac{(Z+Z_0)\mathrm{e}^{-\gamma x}+(Z-Z_0)\mathrm{e}^{-\gamma(2l-x)}}{(Z+Z_0)(Z_0+Z_s)+(Z-Z_0)(Z_0-Z_s)\mathrm{e}^{-2\gamma l}}\cdot Z_0 v_s \qquad (34.6)$$

and
$$i(x)=\frac{(Z+Z_0)\mathrm{e}^{-\gamma x}-(Z-Z_0)\mathrm{e}^{-\gamma(2l-x)}}{(Z+Z_0)(Z_0+Z_s)+(Z-Z_0)(Z_0-Z_s)\mathrm{e}^{-2\gamma l}}\cdot v_s \qquad (34.7)$$

In a good cable losses R and G are small. This implies that the real part α of γ will be also small. The imaginary part of γ then equals $\mathrm{j}\omega\sqrt{LC}$ for sufficiently large values of ω. The results for the instantaneous values also follow from the above calculations, which are valid for the amplitudes, by replacing v_s by $v_s\mathrm{e}^{\mathrm{j}\omega t}$. The term $v_s\,\mathrm{e}^{-\gamma x}\,\mathrm{e}^{\mathrm{j}\omega t}$, occurring in (34.6) and (34.7), can then be approximated by $v_s\,\mathrm{e}^{-\alpha x}\,\mathrm{e}^{\mathrm{j}\omega(t-x\sqrt{LC})}$. This expression represents a signal which is propagated with decreasing amplitude and a velocity $1/\sqrt{LC}$ along the cable, in the positive x-direction. Analogously the term $v_s\mathrm{e}^{-\gamma(2l-x)}\mathrm{e}^{\mathrm{j}\omega t}$ represents a signal with decreasing amplitude, propagating with the same velocity, but this time in the negative direction. It appears from (34.6) and (34.7) that this last signal is not present when $Z=Z_0$; in all other cases we can say that at the end of the cable, a reflection occurs with a coefficient $(Z-Z_0)/(Z+Z_0)$. If $Z_s\neq Z_0$, reflection will also occur at the beginning of the cable for the returning wave, with a reflection coefficient $(Z_s-Z_0)/(Z_s+Z_0)$.

In general, when transferring signals through cables, one does not tolerate these mutually interfering signals because they affect the amplitude of the output signal and make this also strongly frequency-dependent. Reflections at the output of the cable can be avoided by making Z equal to the characteristic impedance Z_0. In order to keep the effect of reflections small in the case of a possible imperfect match at the output end, the source impedance Z_s should also be as closely equal to Z_0 as possible. If the cable is thus terminated at both ends with its characteristic impedance, (34.6) and (34.7) will become:

$$v(x)=\tfrac{1}{2}v_s\,\mathrm{e}^{-\gamma x}$$

and
$$i(x)=\frac{1}{2Z_0}\,v_s\,\mathrm{e}^{-\gamma x}.$$

From a source voltage $v_s e^{j\omega t}$ the output voltage will therefore be $\frac{1}{2}v_s e^{-al} \cdot e^{j\omega(t - l\sqrt{LC})}$; a damping e^{-al} occurs and the signal is retarded by the "delay time" $l\sqrt{LC}$.

According to the above, the input impedance $Z_i = v(0)/i(0)$ of the cable terminated with impedance Z, will be:

$$Z_i = \frac{Z + Z_0 + (Z - Z_0)e^{-2\gamma l}}{Z + Z_0 - (Z - Z_0)e^{-2\gamma l}} \cdot Z_0 \tag{34.8}$$

If $Z = Z_0$, Z_i will have the value Z_0, independent of the length of the cable. However, if Z is greatly different to Z_0, very high or very low input impedances may be obtained for certain lengths of cable. This effect is used in h.f. techniques (Lecher lines, etc.), whilst it must be taken into account when using cables for the connection of two h.f. circuits.

At sufficiently high frequencies and with a good cable, we may replace $Z_0 = \sqrt{(j\omega L + R)/(j\omega C + G)}$ by $Z_0 = \sqrt{L/C}$, so that the characteristic impedance is then a pure resistance. The values of L and C follow from the dimensions of the cable and the values of the relative permeability μ_r (≈ 1) and the dielectric constant ε_r of the medium present between the conductors. In order to minimize the dielectric losses, it is normal to use only as much dielectric as is necessary to support the conductors. For example, in the case of coaxial cables, beads, discs and spiral windings of dielectric are used for this purpose, and ε_r will therefore tend towards unity. We find for the propagation velocity $v = 1/\sqrt{LC}$ the value $C/\sqrt{\mu_r \varepsilon_r}$, where c is the velocity of light. $v = c/\sqrt{\varepsilon_r}$ is then valid in good approximation. We derive from this the characteristic impedance $Z_0 = \sqrt{\varepsilon_r}/cC$. Since C is proportional to ε_r, Z_0 will be inversely proportional to $\sqrt{\varepsilon_r}$, i.e. $Z_0 = 1/cC_0\sqrt{\varepsilon_r}$, where C_0 is the capacitance of the cable without dielectric.

With a coaxial cable, C_0 is inversely proportional to ln r_o/r_i, where r_o and r_i are the outer and inner diameters of the dielectric respectively. Numerically expressed:

$$Z_0 = 60\, \varepsilon_r^{-\frac{1}{2}} \ln \frac{r_o}{r_i} \text{ ohms.}$$

The value of Z_0 lies between 40 and 150 Ω for the usual practical dimensions.

We have for the characteristic impedance of two parallel conductors ("twin-lead"):

$$Z_0 = 120\, \varepsilon_r^{-\frac{1}{2}} \operatorname{arcosh}\left(\frac{l}{d}\right)$$

where l is the distance between the conductors and d the diameter of the conductors.
This can be approximated for $l \gg d$ by:

$$Z_0 = 120\, \varepsilon_r^{-\frac{1}{2}} \ln \left(\frac{l}{d}\right)$$

Normal values are found between approx. 100 and 400 Ω. Roughly speaking, the losses in the cables are proportional to $\sqrt{\omega}$. With practical cables, the value of α is between 0.1 and 2 per cent per metre at 100 Mc/s. Apart from the low cost, the great flexibility of twin-lead cable is an advantage. Its disadvantage is its much greater radiation and sensitivity to extraneous fields.

Ladder networks are similar to cables, particularly the type consisting of inductors and capacitors (Fig. 34-3). Also with the general ladder network of Fig. 34-4 one can speak of the characteristic impedance Z_0, by interpreting

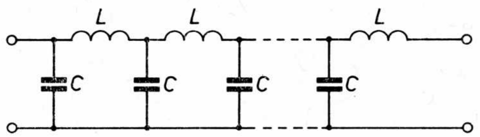

Fig. 34-3

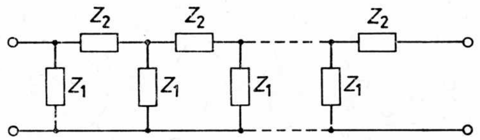

Fig. 34-4

this as being the input impedance of an infinitely long circuit, or the impedance necessary for terminating a finite number of sections resulting in an input impedance of the same value. Z_0 can be calculated quite simply by

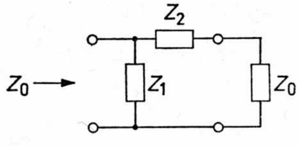

Fig. 34-5

considering that the input impedance of a single section which is terminated by Z_0, must be Z_0 (Fig. 34-5):

$$Z_0 = \frac{Z_1(Z_0 + Z_2)}{Z_1 + Z_2 + Z_0}$$

It follows:
$$Z_0 = -\frac{Z_2}{2} \pm \sqrt{\frac{Z_2^2}{4} + Z_1 Z_2}$$

We must select that sign for the root which makes the real part of Z_0 positive, once again realizing that a positive energy dissipation must exist. Only when the term inside the root sign is negative and real this rule is not decisive. This case occurs only when Z_2 and $Z_1 + \frac{1}{4}Z_2$ are both purely imaginary and have the same sign. Obviously the root then has the same sign. For example, we find for the circuit of Fig. 34-3:

$$Z_0 = -\frac{j\omega L}{2} \pm \sqrt{\frac{-\omega^2 L^2}{4} + \frac{L}{C}}$$

The root is real for $\omega^2 L^2/4 < L/C$, and the positive sign must be used. We therefore find in the frequency range $0 \leqslant \omega \leqslant \omega_0 = 2/\sqrt{LC}$ for the modulus of Z_0:

$$|Z_0| = \sqrt{\frac{\omega^2 L^2}{4} + \left(\frac{-\omega^2 L^2}{4} + \frac{L}{C}\right)} = \sqrt{\frac{L}{C}}$$

This is therefore constant. We find for $\omega > \omega_0$:

$$Z_0 = j\left(\frac{-\omega L}{2} + \sqrt{\frac{\omega^2 L^2}{4} - \frac{L}{C}}\right)$$

The curve of $|Z_0|$ against frequency is shown in Fig. 34-6.

Considering the network of Fig. 34-4, the attenuation factor per section is $Z_0/(Z_0 + Z_2)$. In the special case of Fig. 34-3 this becomes:

$$\frac{-\dfrac{j\omega L}{2} + \sqrt{\dfrac{-\omega^2 L^2}{4} + \dfrac{L}{C}}}{+\dfrac{j\omega L}{2} + \sqrt{\dfrac{-\omega^2 L^2}{4} + \dfrac{L}{C}}}$$

The moduli of numerator and denominator are the same for $\omega \leqslant \omega_0$, so that no attenuation, only a phase shift will occur. On the other hand, the absolute value of the ratio of numerator to denominator will rapidly decrease above the "cut-off frequency" ω_0. A "low-pass" filter is obtained by connecting a number of such sections in series. It will pass all frequencies below ω_0 without attenuation, while those above ω_0 will be strongly attenuated. In order to obtain a result in accordance with the above calculations, one must terminate such a filter with Z_0. However, it is not possible to do this exactly, because of the irrational form of Z_0. Termination can only be approximate. As a first approximation we can use a resistor equal to $\sqrt{L/C}$ for this. A better approximation is achieved by connecting this in parallel to a capacitor of the value $\frac{1}{2}C$.

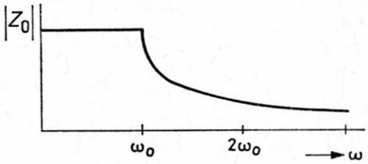

Fig. 34-6

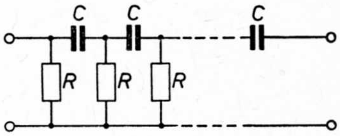

Fig. 34-7

Another simple ladder network is shown in Fig. 34-7. The characteristic impedance is here:

$$Z_0 = -\frac{1}{2j\omega C} + \sqrt{\frac{-1}{4\omega^2C^2} + \frac{R}{j\omega C}}$$

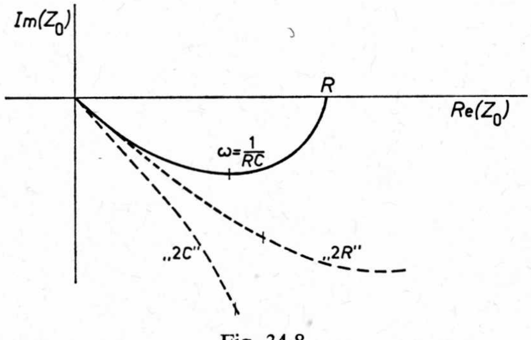

Fig. 34-8

For large values of ω, i.e. when $\omega \gg 1/RC$, we can write as good approximation:

$$Z_0 = \sqrt{\frac{R}{j\omega C}} = (1 - j)\sqrt{\frac{R}{2\omega C}}$$

This means that the modulus of Z_0 at these frequencies is inversely proportional to $\sqrt{\omega}$, whilst the argument is $-45°$, independent of ω (drawn line in Fig. 34-8).

One can often obtain a network with better properties in this respect by doubling the value of the first parallel branch, or by inserting half the series impedance in series with the network (Fig. 34-9). The corresponding input impedances are also indicated in Fig. 34-8.

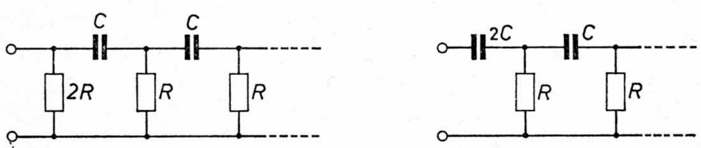

Fig. 34-9

It can be shown that with the more general ladder network of Fig. 34-10 impedances are obtained which, in a specific frequency range, possess to very close approximation any required argument φ between $-\pi/2$ and $+\pi/2$. The modulus of the impedance is in that region approximately proportional to $\omega^{2\omega/\pi}$. An application of this property is given in Section 42.

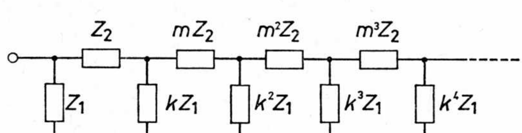

Fig. 34-10

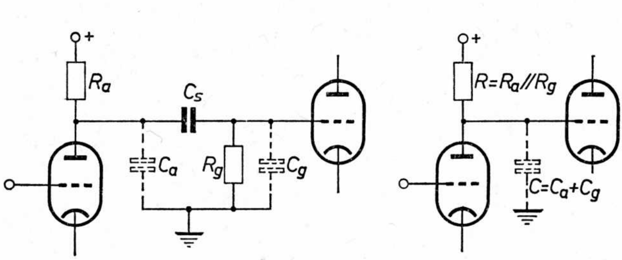

Fig. 34-11

With a given a.c. valve amplifier, the amplification behaviour at high frequencies is determined by the stray capacitances of the anode and grid circuit (left-hand side of Fig. 34-11, where C_a = capacitance of the anode with its wiring, and C_g = capacitance of the grid with its wiring).

For high frequencies it is usually possible to consider the coupling capacitance C_s as a short-circuit, so that the circuit can be reduced to that shown at the right-hand side of Fig. 34-11. Assuming μ to be large, the amplification of the first valve is $SR/(1+j\omega RC)$. The frequency ω_{max}, at which the amplification is smaller than amplification A_0 at low frequencies by a factor $\sqrt{2}$, is in this case $1/RC$, which may be written as S/A_0C, and thus: $A_0\omega_{max} = S/C$.

This is known as the gain-bandwidth product relation. This product is determined by the value S/C of the valve, where $C=$ the sum of anode and grid capacitances, if we assume that we are able to keep the wiring capacitance insignificantly low. This value S/C is called the "figure of merit" of the valve and is greatest for valves having a large slope: approx. 10^9 rad/sec. The above relation remains valid even when selecting another value than ω_{max} for the upper bandwidth limit, the only change being that of a proportionality constant. For example, if it is mandatory that the amplification only drops by 1 per cent at the limiting frequency, the product will be $S/7C$ because this drop occurs at $\omega \approx 1/7\ \omega_{max}$.

A perfectly flat amplitude characteristic over a bandwidth is theoretically possible by using a ladder network as the anode impedance and in which the stray capacitances appear as elements. The circuit of Fig. 34-3 is particularly suitable for this purpose and its application gives the configuration of Fig. 34-12. Since the impedance of this circuit has an absolute value $\sqrt{L/C}$ up to frequency $\omega_{max} = 2/\sqrt{LC}$, we have for the amplification $A_0 = S\sqrt{L/C}$, and for the product of amplification and bandwidth $A_0\omega_{max} = 2S/C$.

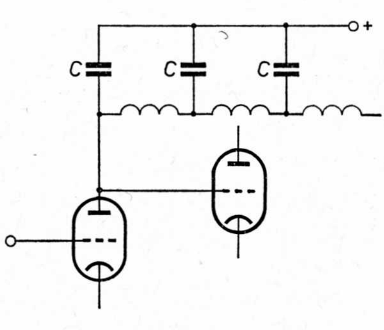

Fig. 34-12

In a practical design it will be necessary to terminate the ladder network after a number of sections, which causes a slight deviation from the theoretical properties. However, terminating even after only a few sections still gives a good result. Termination with a resistor immediately after the first inductor, is often applied (Fig. 34-13). The flattest amplitude characteristic is obtained when $L = R^2C(\sqrt{2}-1) \approx 0.4R^2C$. The amplification at the cut-off frequency is then reduced to approx. 0.6 of the theoretically obtainable value with an infinite circuit. A large gain is achieved by the additional insertion of a second capacitor (Fig. 34-14). Assuming that a reduction in amplification of 1 per cent is still acceptable, the increase in bandwidth is then approx. 15 times that when only an anode resistor is used.

Making use of the property that this ladder network does not attenuate below the cut-off frequency, an increase in bandwidth of approx. 2 times can be obtained by connecting the grid of the next valve not to the first, but to the second capacitor of the network (Fig. 34-15). C_a will then serve as first capacitor and C_g as the second. A correction may be necessary when C_a and C_g are not equal. Here, too, an approximate termination is possible, for example, as indicated in Fig. 34-16. The bandwidth, compared with that of the circuit with only an anode resistor, can be approximately of 25 times

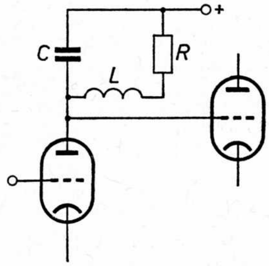

Fig. 34-13

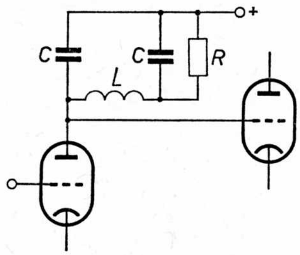

Fig. 34-14

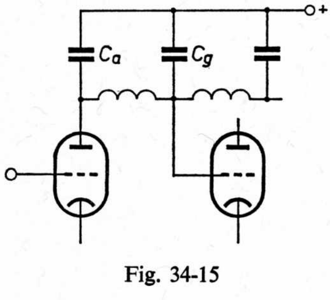

Fig. 34-15

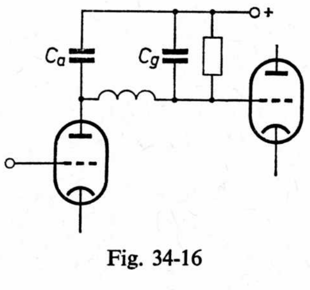

Fig. 34-16

greater. It should be noted that the amplification at high frequencies will now drop far more rapidly than with just the anode resistance, and this may cause overshoot with signals containing these frequencies (Gibbs's phenomenon, see Section 22). Amplifiers of the configurations of Figs 34-12--16, when designed for maximum bandwidth, are therefore solely to be used for the amplification of signals where all components lie within the transmitted frequency band. Depending on the requirements, it is also possible to adjust these circuits to other criteria, such as good step function response.

A still more intensive application of the properties of the ladder networks is made in the "distributed amplifier" (Fig. 34-17). Two ladder networks are used here, where the capacitances of the first network are the grid capaci-

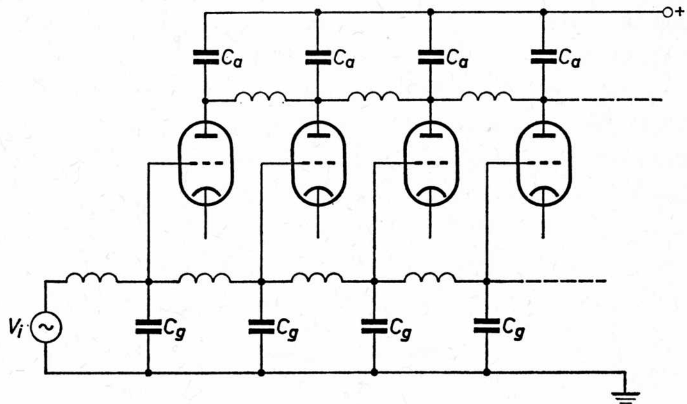

Fig. 34-17

tances of a great number of valves (in practice 10–20), corrected to equal values, and those of the second network the corresponding anode capacitances. If the propagation times are equal in both circuits, the signals amplified by the valves will arrive at the output in phase, and the resultant output signal will increase proportionally to the number of valves.

Another feasible solution is to design a multi-stage amplifier, where in each stage a certain number of valves are connected as in a distributed amplifier. It is obvious that the design and setting up of this configuration is far from easy, and that the gain is achieved only after much expenditure.

Whilst the upper frequency limit of the transmitted band of a.c. amplifiers has an obvious origin, namely the unavoidable presence of stray capacitance, the lower frequency limit is only of a practical nature. Considering the coupling of two stages as shown in Fig. 34-18, we have:

$$v_g = -\, i_a \frac{R_a R_g}{R_a + R_g + \dfrac{1}{\mathrm{j}\omega C}}$$

The frequency at which the transfer is attenuated to a given fraction is inversely proportional to C, and this limit can therefore be made lower by increasing the capacitance. The practical restriction lies in the first instance in the fact that suitable capacitors can only be obtained up to a certain value; electrolytic capacitors are not often used as coupling capacitors on account of their large leakage, and other types of capacitor are rather bulky above the 10 μF-value. Furthermore, capacitors with a large bulk will also introduce quite an amount of stray capacitance in the anode and grid circuits, which can mean a drastic reduction of the upper limit of the passed band for wide-band amplifiers. A further disadvantage of the use of very large time constants in the coupling of successive stages is in the long blocking times which can occur as a result of large interference signals. This can be illustrated by imagining a large negative pulse on g_1 in Fig. 34-19. No current will then pass through the first valve for a certain time. The voltage of anode a_1 will increase until the supply voltage is reached. It follows that the po-

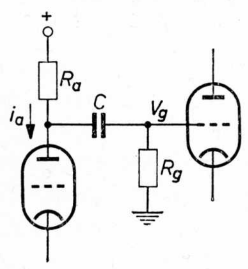

Fig. 34-18

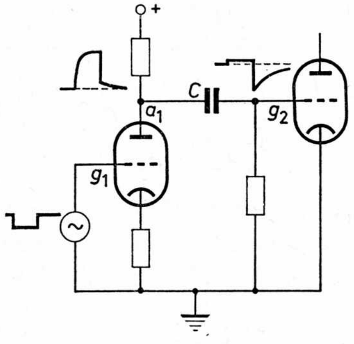

Fig. 34-19

tential of grid g_2 is pulled upwards because the voltage across capacitor C cannot change quickly enough. However, the moment this grid becomes positive with respect to the cathode, grid current flows, which injects charge into the capacitor. Since the grid-cathode diode has a low internal resistance (a few kΩ) and the anode resistance is usually much smaller than the grid resistor, the increase in voltage across the capacitor will occur much more quickly than that which corresponds to the normal time constant of the circuit. The first valve will again pass current after the negative pulse at g_1 has finished, and there is nothing to prevent a corresponding decrease in the voltage at anode a_1, thus taking grid g_2 with it. The latter becomes strongly negative, and since the leakage current is now determined by the large time constant of the coupling circuit, it may take many times this time constant before the voltage at g_2 has risen sufficiently close to the quiescent position. If the second valve is followed by a similar third stage, the same phenomenon will occur during the period when no current passes through the second valve. Because of this effect, the total effective duration in a 3–5 stage capacitor-coupled amplifier may amount to tens of times the "time constant" of the amplifier, even if the interference pulse only lasted a very short time. We have seen before that a great improvement can be achieved by the exclusive use of balanced stages, since much greater signals can be accepted by these without grid current occurring.

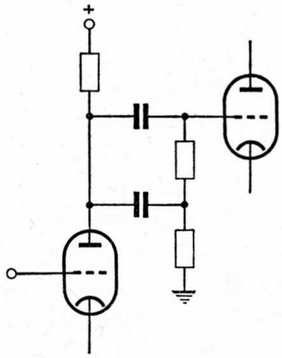

Fig. 34-20

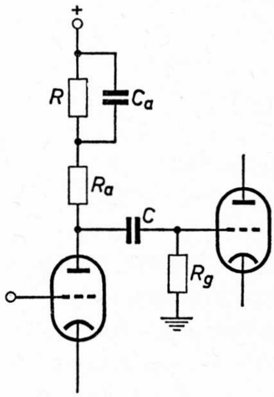

Fig. 34-21

There are some ways of amplifying lower frequencies than with the usual methods, in which only relatively small coupling capacitors are used. In the first place we can use the coupling circuit according to Fig. 34-20 which has been discussed in Section 27, if necessary extended with a third element. It is also possible to make the grid resistance apparently larger as was discussed in Section 26. A third method is indicated in Fig. 34-21. When the relation $R_aC_a = R_gC$ is satisfied, the distribution of the anode signal current over both branches would be independent of frequency in the absence of R; this would also apply to the transfer. This property is partly spoiled by resistance R (which is necessary for direct current), but to a lesser extent when R is made large with respect to R_a. This relation obviously necessitates a high value of C_a, but since no separation for direct current is required in this case, an electrolytic capacitor can be used.

The increase in bandwidth of all these circuits is not particularly great. With the last method the gain is achieved at the cost of the overall amplification since for a given value of the total anode resistance, R_a will have to become smaller. If one is prepared to sacrifice amplification, the target can also be attained by using feedback (Fig. 34-22).

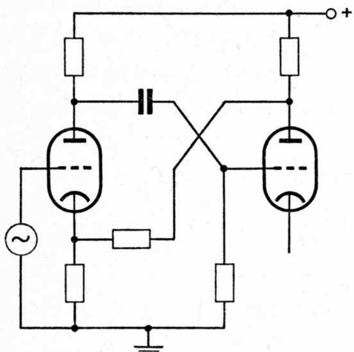

Fig. 34-22

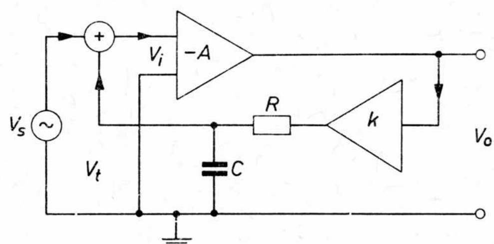

Fig. 34-23

Coupling capacitors are sometimes used, not to obtain an easy signal transmission with separation of the d.c. levels, but for the purpose of keeping the d.c. voltage amplification small as compared to that for a.c. signals. In these cases, the target can often be reached by applying selective feedback to a direct-coupled amplifier, as for example in Fig. 34-23. If a d.c. voltage amplifier is used which amplifies $-A$ times for $\omega < \omega_{\max}$, the following equations apply to the feedback system:

$$v_o = -A(v_s + v_t), \quad v_t = \frac{k}{1 + j\omega\tau} v_o$$

where $\tau = RC$ and therefore:

$$\frac{v_o}{v_s} = \frac{-A}{1 + \dfrac{Ak}{1 + j\omega\tau}}$$

Amplification is almost $-A$ for frequencies $\omega \gg Ak/\tau$. However, this drops to approximately $1/k$ for d.c. signals, and this will be much smaller than A if $Ak \gg 1$.

Example:
$A = 1000$, $k = 1/50$ and $\tau = 20$ sec. gives $Ak/\tau = 1$, so that the amplification is 1000 for frequencies $\omega \gg 1$, whilst a d.c. voltage will only be amplified 50 times.

We not only obtain a very stable d.c. voltage level with this method, but the circuit has also the advantage of not being blocked by short-duration overload pulses. This method is therefore highly suitable for amplifying step-function signals. It is also possible in this case to realize the required large time constant with relatively small coupling capacitors, as R can be made large with feedback to a grid. The d.c. voltage feedback permits a much higher value of R than would be possible with a fixed working point where there is a limitation due to grid emission dangers.

With transistor circuitry the effect of stray capacitances is much smaller than with valve circuits, because of the low impedance level. The high frequency limit is imposed by the cut-off frequency of the transistor and possibly by stray inductances. As stated before, high amplification extending to very high frequencies can be obtained with modern h.f. transistors, where the values of the cut-off frequencies are of the order of magnitude of 1000 Mc/s. For this reason, only the simpler correction methods as indicated in Figs. 34-13 and 34-14, are used with transistor amplifiers.

Transistor amplifiers for very low frequencies often require an artificial increase in the input impedance to avoid extremely large coupling capacitors. Such an increase can, for example, be obtained with an emitter follower circuit. This method is not particularly attractive however, since, as we shall see in the next section, direct coupling is usually much easier with transistor circuits than with valves.

35. D.C. amplifiers

It is necessary for some purposes that when a signal is being amplified, its d.c. voltage level is included, so that all changes in this level, however slow, will appear in an amplified form at the output and can be used. With a.c. amplifiers, there is no need to satisfy this requirement, particularly the "however slow" component. This makes it possible to include coupling capacitors between the stages. The d.c. voltage level at the output is then independent of that at the input and is therefore not a measure of it. When designing amplifiers where d.c. voltage changes from the quiescent position must also be amplified – and these are therefore called d.c. amplifiers – it is first of all not possible to use coupling and decoupling capacitors, and secondly we must take into account the fact that a possible "drift" of the quiescent voltages and currents will cause a signal at the output. We shall discuss the complications produced by these factors in this section.

First of all we shall see what methods there are for coupling the successive stages of d.c. amplifiers. In all these cases we speak of "direct coupling".

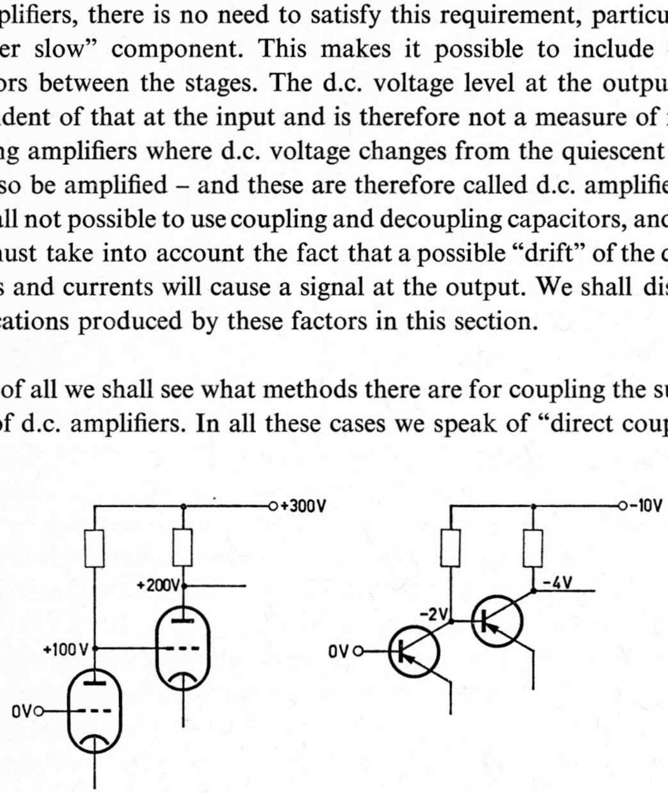

Fig. 35-1

The simplest example of such coupling is a circuit where the control electrode of a stage is directly connected to the output electrode of the previous stage (Fig. 35-1). However, in the case of multi-stage valve amplifiers the required supply voltages will become very high indeed, whilst the d.c. voltage level of the output, at zero signal, may be hundreds of volts higher than that at the input. Although this is not objectionable in some cases, we usually require that both levels are the same, normally earth potential. These considerations apply in a much lesser degree to d.c. transistor amplifiers because of the small collector-emitter voltages necessary, and because of the possibility of alternately using *n-p-n* and *p-n-p* stages.

Apart from the possibility of using accumulators or dry batteries (which are in any case not recommended), there exist several other methods of coupling stages which at the same time reduce the d.c. voltage level. They are all based on the use of voltage dividers. The simplest design is that of resistors connected to a suitable negative supply (left-hand side of Fig. 35-2).

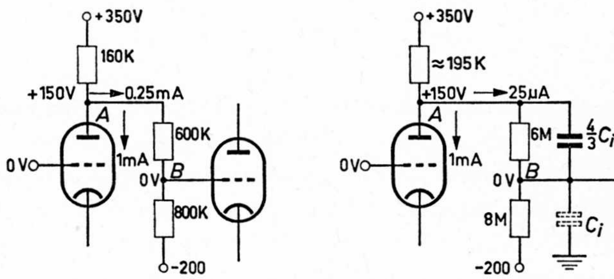

Fig. 35-2

The loss in amplification has two causes. One is that the anode resistor must be made smaller because of the additional current through the voltage divider and the anode load is further reduced because the voltage divider is arranged parallel to the anode resistance. Secondly, only a fraction of the anode signal voltage will emerge at the next grid (in the example $8/14 = 57\%$).

The first cause can be remedied by selecting high-value resistors for the voltage divider (right-hand side of Fig. 35-2). Because of input capacitance C_i of the second stage, the amplification will begin to drop at a relatively low frequency. This can be remedied by arranging parallel to the upper resistor a capacitor of such a value that the time constants of both branches become the same.

The second cause of signal loss, namely the fact that only part of the anode signal arrives at point B, becomes particularly troublesome when the signal at B must be large, as is the case for an oscilloscope deflection stage. In this case it is not so much the loss in amplification which is important, but the fact that anode A must be able to handle an unnecessarily large signal, which demands very conservative design of the left-hand stage. It is then often worth while to replace the upper resistor by a stabilizing valve (Fig. 35-3).

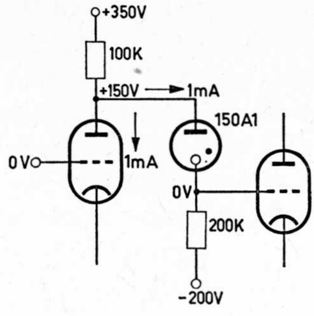

Fig. 35-3

As the internal resistance of this valve is small for changes when compared to the other resistors (approximately 400 Ω), the signal loss is now negligible, but the anode impedance must become much smaller because of the relatively large current this valve requires (in the example 100 kΩ parallel to 200 kΩ: approximately 67 kΩ). This too can be prevented, namely by inserting a cathode follower (left-hand side of Fig. 35-4). To avoid both the high noise of the stabilizer valve and the frequency dependence of the amplification because of the valve's increasing internal impedance at higher frequencies, a capacitor-resistor network is often placed across the valve (right-hand side of Fig. 35-4). The resistance is necessary because the impedance of the valve becomes inductive at high frequencies, and can, in combination with

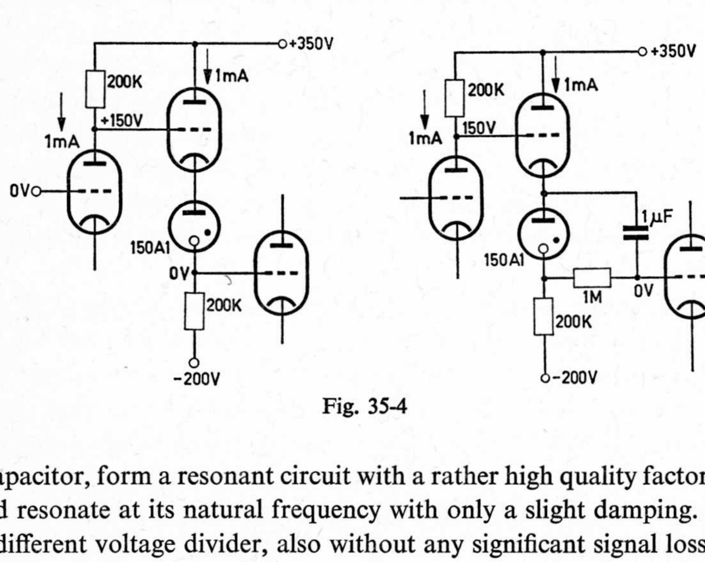

Fig. 35-4

the capacitor, form a resonant circuit with a rather high quality factor, which would resonate at its natural frequency with only a slight damping.

A different voltage divider, also without any significant signal loss can be designed by replacing the lower resistor by a circuit with a high differential resistance, for example, a triode with large cathode resistance (left-hand side of Fig. 35-5). Once again, a circuit arrangement with a cathode follower can be used to avoid a direct current load on the anode (right-hand side of Fig. 35-5). Similar facilities obviously exist for transistor amplifiers, but the need for them is much less frequent, as we have already remarked.

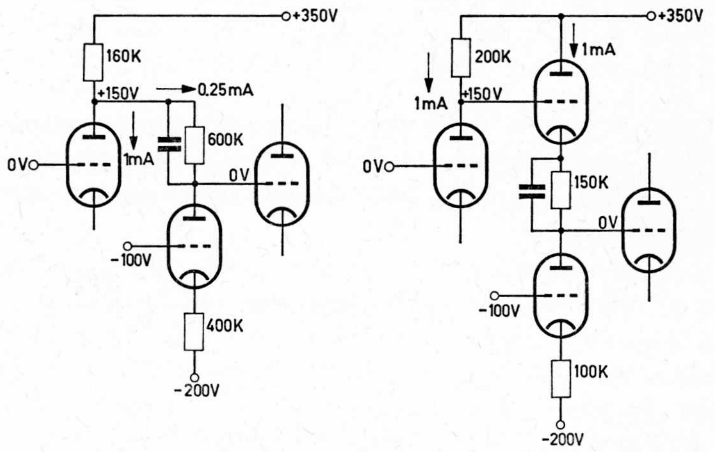

Fig. 35-5

The most troublesome phenomenon with d.c. amplifiers – in any case when a high degree of sensitivity is required – is the drift of the quiescent voltages and currents. Variation in supply voltages is an obvious cause, particularly when these are derived from the mains voltage. Another contributory factor is the change in the values of components and contact potentials because of temperature changes and ageing. It is theoretically possible to take precautions against all these phenomena, so that the ultimate limitation of the sensitivity is determined by "spontaneous" changes, mostly of a microscopic nature: recrystallization, diffusion of substances, adhesion to surfaces and suchlike. A measure of the total of these effects is usually taken to be the "equivalent input signal", i.e. the input signal which would give as large an output signal as produced by the phenomena. Since the effect of such phenomena is greater when there is less amplification between the input and the place where the disturbance occurs, it is essential that these phenomena are kept as small as possible especially in the first stages of an amplifier.

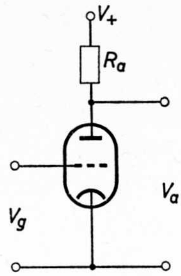

Fig. 35-6

The effect of changes in the supply voltage is very great with a single-sided stage. For example, in the circuit of Fig. 35-6 the equivalent input signal of a change ΔV_+ in the positive supply voltage equals $-\Delta V_+/SR_a$. For we have:

$$\Delta I_a = S\Delta V_g + \frac{S}{\mu}\Delta V_a \quad \text{and} \quad \Delta I_a = \frac{\Delta V_+ - \Delta V_a}{R_a}$$

so that

$$\left(\frac{1}{\mu} + \frac{1}{SR_a}\right)\Delta V_a = \frac{\Delta V_+}{SR_a} - \Delta V_g$$

A change in voltage ΔV_+ corresponds therefore to a change $\Delta V_g = -\Delta V_+/SR_a$ in the voltage on the control grid.

Even when SR_a has the relatively large value of a few hundred, a change of, say, 0.1 volt in the positive supply voltage nevertheless corresponds to an equivalent input signal of a few tenths of a millivolt. We must therefore conclude from this that it is necessary for the positive and negative supply voltages to be highly stabilized if a high sensitivity is asked of a single-sided amplifier.

The same applies to the heater voltage. Measurements have shown that a change in the heater voltage of 10 per cent causes a displacement of the $I_a - I_g$ characteristic, due to the consequent temperature change and hence the change in the emission of the cathode, which corresponds to a change in the grid voltage of 100–200 mV. It is therefore necessary that the heater voltage is also well stabilized.

When we discussed difference amplifiers in Section 28, we saw that the effect of supply voltage changes is considerably smaller with these symmetrical circuits; the better the two halves are matched the smaller the change. The rejection factor proved to be a measure of this, and calculations show that the equivalent input voltage for changes in positive and negative supply voltages is approximately as many times the rejection factor smaller than the changes themselves. Since the guaranteed rejection factor can easily be made larger than 10^4 when valves are used, the effect of these supply voltages is about 100 times less than with single-sided stages. In order to escape from the above-mentioned stringent requirement imposed on the stability of supply voltages for single-sided stages, it is almost mandatory to design the first stages of sensitive d.c. amplifiers as difference stages. Obviously the favourable result of a good first stage should not be spoiled by an insufficient successive stage; it is therefore recommended to use a difference stage for the second stage as well, with a reasonable degree of symmetry ($H \gg 100$). A simple triode difference stage is frequently sufficient. Because of the advantages of a balanced output and the small risk of oscillation, even at high amplifications, simple difference stages are often used for subsequent stages as well.

High demands are made on the stability of the anode resistors of the first stage, because they usually carry a d.c. voltage of 100–200 volts. A mutual difference of 10^{-4} between the relative changes of the anode resistors results in an anode difference signal of the order of 10 mV and an equivalent input signal of the order of 0.1 mV. This makes it necessary that the anode resistors are stable with a low temperature coefficient, and must be physically arranged so that their temperatures will be the same as far as possible.

If a voltage divider is present between the first and second stage, the situation becomes still more unfavourable because there are more resistors and therefore more sources of variation, whilst the effective amplification of the first stage is necessarily smaller. These disadvantages can be avoided by the direct connection of the first and second stages, i.e. without using voltage dividers, which are not inserted until after the second stage. We thus obtain the basic circuit of Fig. 35-7.

The advantage which can be obtained by difference stages in their insensitivity to variations in the positive and negative supply voltages, does not apply to the same extent to the effect of the heater voltage. The displacement of the $I_a - V_g$ characteristic as a result of a change in heater voltage is not the same for all samples of valves of the same type, which are theoretically identical in design. Although there will always be a certain degree of compensation because of their balanced usage, this does not exceed a factor 5–10 with entertainment valve types. Professional valves have smaller tolerances in their components and more care is taken in their manufacture, so that a better compensation can be expected. Valve E80CC (6085) is superior in this respect; the relative difference between the two halves rarely

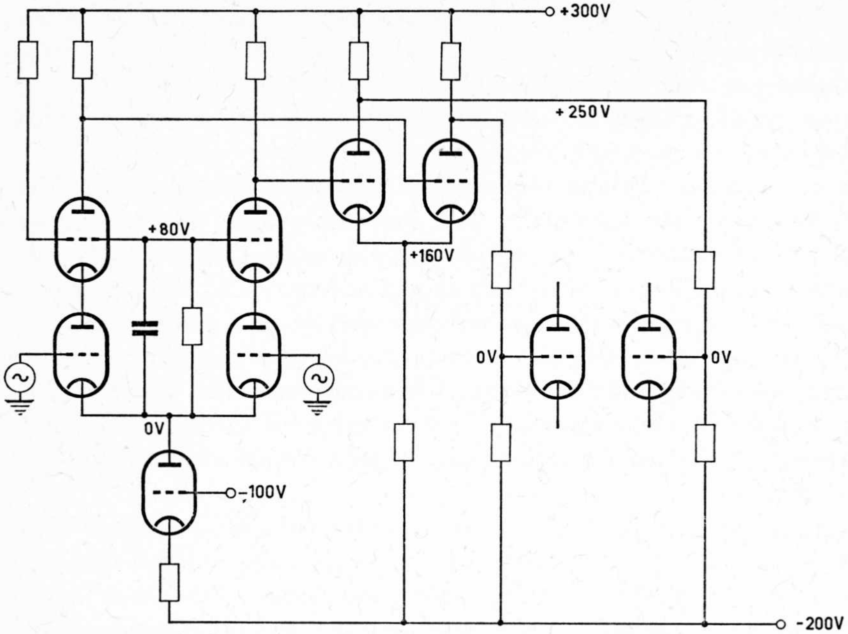

Fig. 35-7

amounts to more than 10 mV at a 10 per cent variation in heater voltage, and is normally only a few millivolts. However, even when using this valve in sensitive amplifiers, it will be necessary to stabilize the heater voltage of the first and possibly of the second stage to considerably better than 10 per cent.

Acoustic and mechanical vibration mentioned in Section 30 can also cause changes in the characteristics. These changes can sometimes be permanent and are caused by a mechanical displacement of the components in the valve, particularly of the heater. It is therefore necessary to apply the relevant precautions with d.c. amplifiers.

As far as applicable the above considerations are also valid for transistorized d.c. amplifiers. Here also, the use of difference stages gives a considerable improvement compared to single-sided stages. Whilst, with a suitable selection of the type of transistor we can obtain with a simple balanced design, rejection factors of the same value as with a cascode circuit using valves, the absolute values of changes in the supply voltages are considerably smaller for the same degree of stabilization. We can therefore state that with transistor amplifiers, more so than with valve amplifiers, variations in the supply voltages are not the limiting factor for sensitivity.

Although transistors have no trouble with changes in heater voltage, there is, however, the rather great effect of ambient temperature on the transistor characteristics. We have seen in Section 21 that for a transistor a temperature change of 1 °C can be taken into account by a change of the base-emitter potential of 2–2.5 mV and a change of approximately 8 per cent in the leakage current between base and collector. By using silicon transistors, the latter becomes of the order of magnitude of 10^{-9} A at room temperature, so that only the first effect is of importance for measurement of signals from sources with relatively low internal resistance.

In the case of a balanced amplifier, the effect of a common change of the base-emitter voltages is reduced by the rejection factor, and can thus remain sufficiently small. Only the difference between the changes of the base-emitter potential for the same change in temperature will remain. The difference in thermal conductance of the transistors can be the prime cause of this, as the transistor temperatures will not immediately follow at the same rate relatively rapid changes in the ambient temperature. This effect can be reduced by placing the transistors in a single copper or aluminium block, or, as is possible with modern techniques, by making the transistors on a single slice. A second cause is the poor match between the temperature coefficients of V_{be} for the two transistors. V_{be} equals the sum of the voltage across the junction and the voltage drop brought about by the base current flowing through the resistance of the base material. It is mainly the latter which causes the difference in the temperature coefficients. Favourable results can be expected here by selecting a transistor pair manufactured on a single slice of material.

Depending on the precautions taken, we must reckon with an equivalent input voltage of 1–50 μV per °C with the balanced design. The small bulk of transistorized d.c. amplifiers often makes it worthwhile to place them in a simple thermostatically controlled volume.

However, there will always remain a d.c. potential instability (or drift) at the output of both valve and transistor d.c. amplifiers, even with perfectly constant external conditions. This is mainly caused by changes in the surface of the cathode or other electrodes in the case of valves, or by those of the crystals in the case of transistors. The exact nature of these phenomena is still unknown, although experiments seem to indicate the effect of certain impurities and processes. Care in manufacture can usually keep these effects reasonably small. The professional valve E80CC (6085) is also very good in this respect. For about 90 per cent of all measured specimens, the equivalent input drift (after about 50 hours burning in) was less than 400 μV/hr, and for about 50 per cent less than 200 μV/hr. By selecting a value for the heater voltage approximately 5 per cent below the nominal value, it is possible to reduce these quantities a little more. Since the drift is not constant, we must allow for a greater value over shorter periods, with a maximum of approximately 1 μV/sec. Over still shorter time intervals the d.c. voltage drift is characterized by the flicker noise of the valve.

Next we consider a transistor amplifier in which the first stage consists of two selected Si-transistors (relative difference in the collector currents for equal voltages: $<10\%$) and which will operate satisfactorily with a collector current of 0.1 mA. Here, if the necessary precautions against disturbances are taken, the drift can be limited to much less than 10 μV/hr which is an order of magnitude smaller than that experienced with the best valves.

These values apply for a short-circuited input. For a valve amplifier the influence of the source resistance is small as long as this resistance is not greater than a few megohms. In transistor amplifiers with collector current of 0.1 mA in the first stage, one can expect fluctuations in the base currents of the order of 10^{-9} A so that such fluctuations become dominant with source resistances of 10 kΩ and more. Thus the transistor amplifier is particularly suitable for the measurement of voltages from a source with low resistance, like those presented by most thermo-couples.

The guaranteed zero-point drift of d.c. amplifiers with valves or transistors is still too large for many measurements. However, there are a number of ways and means which make a very sensitive measurement possible. These are all based on the same principle, although their implementation may vary greatly.

If we have a signal source which can be turned off periodically, it is possible to determine quite accurately the zero-point in so far as its drift is slow compared to the rate of switching. This method is used, for example, when measuring the intensity of a light beam by means of a thermo-couple. By interrupting the light beam periodically, the voltage of the thermo-couple with and without incident light can be measured. A disadvantage of this method is that part of the time no information is gained about the behaviour of the input signal, and part of the time no information about the zero-point. It is therefore not possible to determine rapid changes in this way.

A similar technique can be used for measuring the Hall effect (Fig. 35-8). The "Hall effect voltage" V_H between two points of a semiconductor is proportional to the product of the current I passing through the conductor and the magnetic field B perpendicular to the semiconductor. The proportionality constant depends not only on the dimensions, but also on the load and mobility of the charge carriers. This latter value is important for many applications, and can thus be derived from the value of the Hall effect voltage. Here, too, it is possible to switch-off the signal source by switching

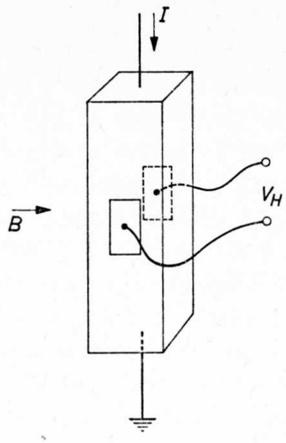

Fig. 35-8

off the current through the conductor. Twice the effect sought can be obtained by changing the direction of this current. This method can be considered as transposing a d.c. voltage signal into an a.c. voltage signal. Strictly speaking, however, we have then returned to an a.c. amplifier.

If it is undesirable or impossible to switch the signal source itself, it is still possible to supply an a.c. voltage to the amplifier that is proportional to the d.c. voltage, by inserting a "chopper" between source and amplifier so that the input is alternately connected to the two terminals of the source.

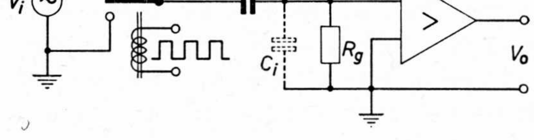

Fig. 35-9

In this case the amplifier may therefore also be an a.c. amplifier (Fig. 35-9). After sufficient amplification the a.c. voltage signal is then changed again into a proportional d.c. voltage, for which we can use synchronous detection. We shall refer to this in Section 40.

The above-mentioned chopper usually consists of a vibrating reed between two contacts, moved by an electromagnetic drive. Because of the rapid increase in the wear of the contacts at high frequencies, the switch frequency is usually below 100 c/s. As the signal is amplified at the same frequency as that of the drive, the screening between drive and amplifier must be very good indeed. The contact potentials can be kept surprisingly low by using suitable materials and taking advantage of the self-cleansing action of the vibrating reed with the contacts. The total residual voltage will not exceed 0.1–1 microvolt with ordinary designs, and the best types guarantee a drift voltage of not more than a few nanovolts per hour, provided the internal resistance of the source does not exceed a few tens of ohms. However, the difficulty here is that the connection to the signal source may possess contact potentials which will exceed this residual voltage by several orders of magnitudes.

A chopper amplifier can only be used for signals where the highest frequency component in the signal is not more than approximately one third of the chopper frequency. This means in practice that an amplifier with a mechanical chopper can only be used for l.f. and d.c. voltage signals. Since the input capacitance must be recharged at each cycle of switching, the amplifier will present an additional load, apart from the normal load imposed by the finite input resistance, e.g. the grid input resistance. The total absorbed charge per second is fC_iV at a switch frequency f, so that this additional load corresponds to a resistance of the value $1/fC_i$. With $f=100$ c/s and $C_i=25$ pF, for example, this works out to be 400 MΩ. In designs where the grid resistance is not apparently increased by feedback, the latter will thus usually determine the input impedance.

Much attention is paid to the development of electronic analogues of the chopper; in the first place because of the low switching frequency of the mechanical chopper, and also because of considerations of life and reliability. Fig. 35-10 shows the principle of the majority of solutions proposed for this problem. If the auxiliary voltage V_H is sufficiently positive, the diodes will not conduct and there will be no connection between A and B. If V_H is sufficiently negative, the diodes will conduct and have a small differential

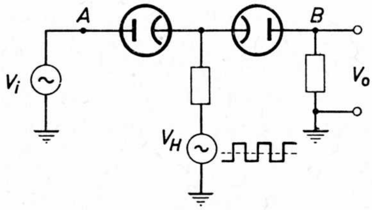

Fig. 35-10

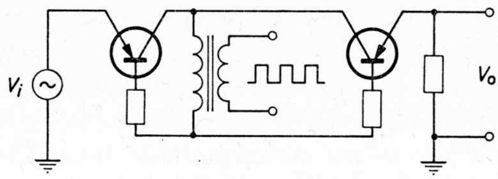

Fig. 35-11

resistance, so that the signal will be transmitted from *A* to *B*. Practical difficulties arise from the potentials which remain across the diodes when they are in the conducting state, and from the switching currents which pass through the input and output circuits. When vacuum diodes are used, the application is mainly restricted because of the difference in the rather high voltages across the diodes. The compensation obtained by the proposed use of two diodes is far from complete, so that the out-of-balance voltages are usually tens of millivolts, combined with poor stability. The use of semiconductor diodes makes the residual voltages considerably smaller, but by far the best results are achieved by using transistors as switching elements. As follows from the basic circuit of Fig. 35-11, a better isolation between the measurement and switching circuits is obtained by this method. Silicon transistors are best for this purpose because they have much smaller leakage currents than germanium transistors in the off-position. Practice shows (as can also be derived from theory) that both the residual voltage in the conducting position and the leakage current in the off-position will be α'/α'_i times smaller, if the base and the collector are connected in the switching circuit instead of the base and the emitter. α' is here the "normal" current amplification factor and α'_i that in the "inverse" direction. The latter is usually of the order of magnitude of 1.

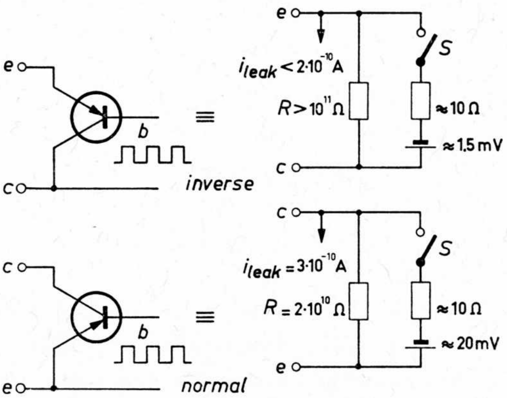

Fig. 35-12

Fig. 35-12 shows results of measurements with a silicon transistor BCZ10 used as the switch in the normal and in the inverse circuit, which illustrates what we have stated above. In this case the resistance in the inversed direction is more than 10^{11} Ω, whilst the forward resistance is approximately 10 Ω. The small leakage current and the part-compensation of the already small residual voltages, make it possible to switch voltages of a few tens of microvolts with sufficient accuracy according to the principle of Fig. 35-11. This may be done without taking many precautions and even when these voltages are supplied by sources with an internal resistance of a few kilo-ohms. The literature of recent years shows that a rapid technological development has taken place in this field. It is possible to obtain residual voltage levels of less than 0.1 μV, provided extensive precautions are taken. We should also note that the field-effect transistor will probably possess still better properties for this purpose.

Another analogue of the mechanical chopper involves photo-resistors. These exhibit a strong increase in conductivity of the material they are made of, when exposed to light. The ratio of the resistance in darkness to that on exposure to a neon tube can be of the order of 10^5. By alternating exposure of the photo-resistors in the circuit of Fig. 35-13, we obtain a square-wave voltage at the output, which is proportional to the input voltage. The response of photo-resistors is rather slow, especially in the direction of darkness, which restricts the switching frequency to approximately 100 c/s. The capacitive load of such a circuit can easily be made equivalent to a resistance of several megohms, whilst residual voltages of less than 1 μV are possible. The latter are mainly caused by impurities in the material and inaccuracies in the exposure.

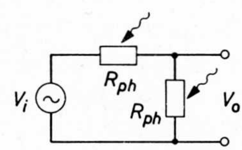

Fig. 35-13

Apart from chopping, there are a number of other methods for the measurement or amplification of d.c. voltages, which make use of a conversion from d.c. to a.c. voltage. Some of these utilize effects caused by the d.c. voltage which can be observed and used as a measure for that voltage. The "varactor" is an example. A semiconductor diode in the cut-off condition has a capacitance in parallel with the reverse resistance as a result of charge displacements at the barrier layer. Now the thickness of this layer, and hence the capacitance, is dependent on the value of the reverse voltage, and appreciable changes in capacity can be produced by changes of the order of millivolts in this voltage. Thus we can measure the small d.c. voltages by studying the changes in capacitance, and this can be done by inserting the element in an a.c. voltage bridge, or by combining it with inductance into a resonant circuit.

Another example is the "magnetic amplifier", where the signal voltage causes a displacement in the *B-H* curve of a magnetic material, which can be measured as a change either in the inductance of a coil or in the transfer of a transformer (Fig. 35-14). The input resistance of magnetic amplifiers is low, usually less than 1 kΩ, but the out-of-balance potential may be less than 1 μV.

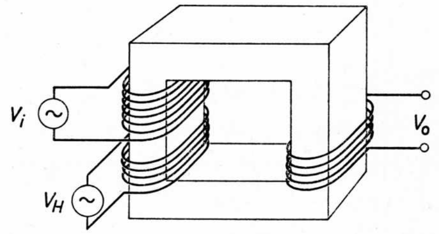

Fig. 35-14

The vibrating membrane electrometer (Fig. 35-15) incorporates an old principle for the measurement of d.c. voltages, where the load on the signal source must be particularly low. The d.c. voltage to be measured is applied to a capacitor C_{var} through a large resistor R. The capacitance of C_{var} is varied periodically by moving one of the plates to and fro. If resistor R is so large that the charge on the capacitor can hardly change during a cycle of the plate movement, the voltage across the capacitor will be inversely proportional to the capacitance ($V = Q/C$) and therefore proportional to the plate distance. The a.c. voltage resulting in this manner is proportional to the applied d.c. voltage, and can be amplified with a normal a.c. amplifier. This circuit behaves, theoretically, after the initial charging of the capacitor as an infinitely small load on the signal source; the mean voltage across the

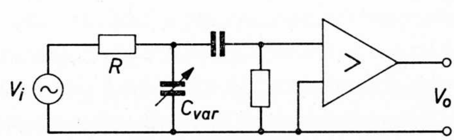

Fig. 35-15

capacitor equals the signal voltage. The input impedance is therefore mainly defined by the insulation resistance of the capacitors and can be very high with suitable design (10^{12}–10^{14} Ω).

As with the mechanical chopper but to a much higher degree, because of the higher impedance levels, we have to be aware of the danger that the drive may develop interference voltages in the measurement circuit.

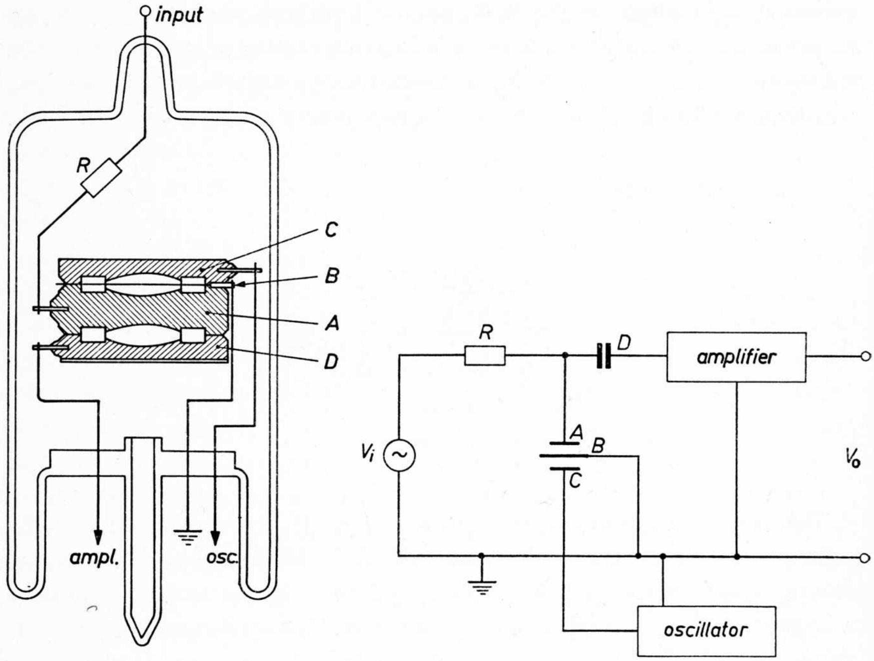

Fig. 35-16

However, the circuit of Fig. 35-16 allows a difference of 100 times or more between measurement and drive frequency, so that a filter can easily separate the measurement signal and the interference. Because the attractive force between the plates of a capacitor is proportional to the square of the voltage between the plates, plate *B*, which has the form of a membrane, may be driven by an h.f. voltage of, say, 1 Mc/s applied between plate *B* and the fixed plate *C*. The amplitude of this voltage is made to vary with the resonant frequency of the membrane. The latter amounts to a few kc/s, and the drive voltage is then a voltage modulated in amplitude (see Sections 36 and 41), which only contains frequencies in the immediate neighbourhood of 1 Mc/s, whilst the measurement frequency equals that of the membrane. By inserting capacitance C_{BC} in the feedback path of an h.f. oscillator (see Section 37), the variation of the oscillator voltage and hence the drive of the membrane can be made equal to the resonant frequency of the lathers. We shall not go into details of this mechanism. The variable measurement capacitance is C_{AB}, i.e. the capacitance on the other side of the membrane. By mounting the whole unit, including the coupling capacitance C_{AD} and possibly resistor R in an evacuated envelope, many advantages are apparent; high insulation resistance, low damping of the membrane and therefore small drive energy, constant contact potentials, and no ionization caused by local radioactivity or cosmic rays. The relatively high measurement frequency has also its advantages: greater bandwidth, low input impedance, and no interference from flicker noise of the a.c. amplifier used. The change in potential is mainly caused under these conditions by the effect of the temperature on the contact potentials of the capacitor electrodes. This effect is of the order of a few microvolts per °C.

When classifying the methods discussed according to input impedance, drift and bandwidth, we obtain the following survey:

	input impedance	residual potential/hour	bandwidth
Valve amplifier	high	200 μV	large
Transistorized amplifier	rather low	10 μV	large
Mechanical chopper	high	0.1 μV	very small
Electronic chopper	high	10 μV	small
Photo-resistor chopper	high	1 μV	very small
Varactor	high	10 μV	rather small
Magnetic amplifier	low	1 μV	small
Vibrating membrane electrometer	very high	1 μV	small

Sometimes for measurement purposes one needs a d.c. amplifier with a flat amplitude characteristic up to high frequencies (e.g. 0.1 Mc/s or more), with a high input impedance ($\gg 1\ M\Omega$) and also a very small voltage drift (constant within a few microvolts). Such an amplifier is, of course, also attractive for general use, but none of the amplifiers listed above combines all these useful properties. However, by combining different types it is possible to build up an amplifier which satisfies all the requirements mentioned. A good combination in this respect is that of a valve amplifier with one of the methods providing high input impedance and low drift, particularly the one with the mechanical chopper. Several ways of combining these two amplifiers are possible.

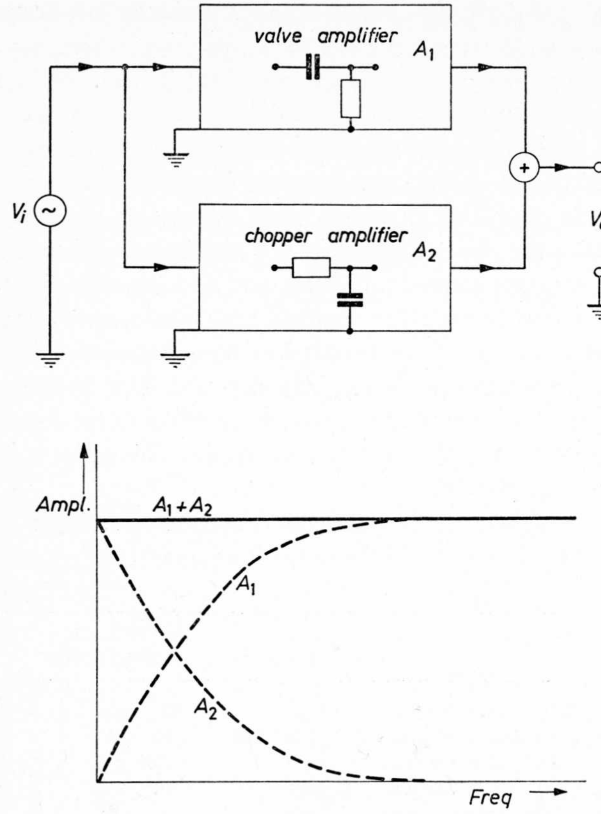

Fig. 35-17

In the first place, the amplifiers can be connected in parallel (Fig. 35-17). Here we use the valve amplifier as an a.c. amplifier, whilst the chopper amplifier only amplifies the low frequencies. The amplitude characteristic should be such that the drop in frequency characteristic of the a.c. amplifier is just compensated by the low frequency amplifier. The input impedance of such an amplifier is high, namely the parallel combination of the high input impedances of the individual amplifiers. The drift equals that of the chopper amplifier, whilst the limitation of the bandwidth is solely determined by the cut-off frequency of the valve amplifier. Heavy demands must be made on the amplification stability and the phase characteristics of the two amplifiers to ensure a flat amplitude characteristic of the combination, even against ageing effects. This is the principal reason why only recently a number of practical designs of this "hybrid amplifier" have been published.

The oldest and still most popular method is shown in Fig. 35-18 and is called the Goldberg-amplifier, after its inventor. Amplification A_2 of the chopper amplifier is here much larger than amplification A_1 of the valve amplifier; in many designs $A_1 = 1$ and $A_2 > 10^4$. The post-amplifier A_3 is also a d.c. amplifier, so that we can speak of pre-amplification of the low frequencies by the chopper amplifier, and the effect of the zero-point drift of A_1 and A_3 is reduced correspondingly. The uncompensated amplitude characteristic rises strongly at low frequencies but is flattened by applying strong feedback. This method has the disadvantage that the amplification also becomes

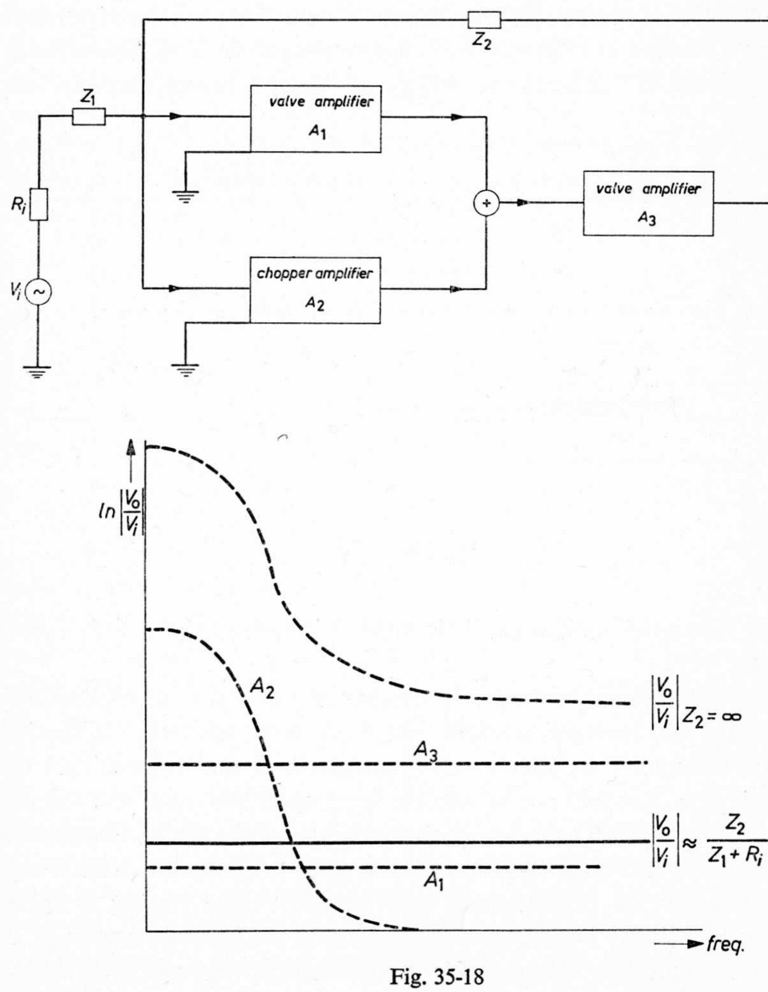

Fig. 35-18

dependent on the source resistance R_i, if this is not small enough compared to Z_1 to be negligible. This method can therefore only be used when R_i is negligible, or when it is constant. The input impedance of the amplifier approximately equals Z_1.

Landsberg's method is shown in Fig. 35-19. The amplification of the valve amplifier is restricted to value A_1 by means of feedback and is very constant at all frequencies. The chopper amplifier now compares the fraction $1/A_1$ of the output signal with the input signal. Any difference can only have been caused by the zero drift of the valve amplifier and is added in amplified form at a suitable point to this amplifier, so that the residual error will be largely compensated. The chopper amplifier here only comes into operation when a deviation is present, and has therefore exclusively a corrective function. Such a method of compensation may be compared to "adding what is lacking" (Section 27) and has comparable advantages, namely, there is less

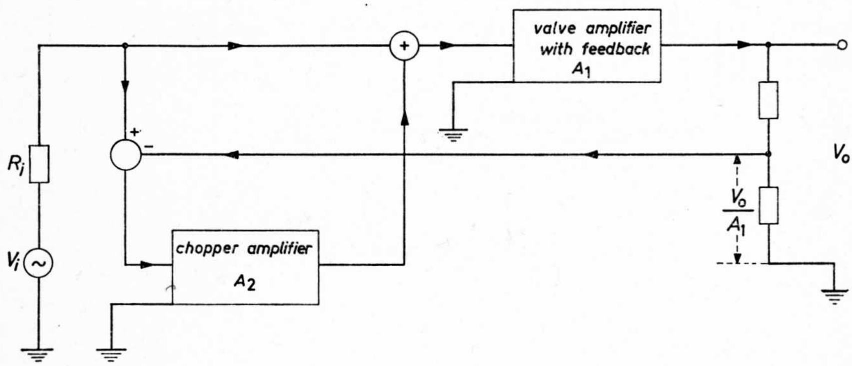

Fig. 35-19

danger of unwanted oscillation. This type of amplifier has a high input impedance.

The above-mentioned methods are suitable for fairly general application. Modifications are, however, possible, which can be of advantage in special cases. For example, if we have a valve amplifier with a zero-point drift of less than 1 μV/s, it would not be sensible to compare the input and output 50–100 times per second, as is the case for the circuit of Fig. 35-19. Dependent on requirements, it is often sufficient in such cases to make one check every 1–10 seconds. If the measurement may be interrupted during a short period, the check may be carried out by shorting the input by means of a relay and automatically counterbalancing the residual error at the output.

The wear on the switching elements is then obviously much smaller than with most choppers. If no measurement time can be lost, we can connect two of these amplifiers in parallel, so that one of the two can carry out the measurement, whilst the other is checked.

As appears from the above, all solutions for obtaining wide-band d.c. amplifiers are rather complicated. The situation would be considerably eased if it were possible to improve one of the earlier mentioned methods for d.c. amplification so that we no longer need a combination of amplifiers. As it is not likely that it will ever be possible to increase the bandwidth of one of the indirect methods to any great extent, we must concentrate on an improvement regarding zero-point drift of valve amplifiers, or input impedance of transistorized amplifiers. The situation with valves is not very hopeful, especially when we remember that a shift of 200 μV corresponds to a change of not more than 0.1 °C in the difference of the cathode temperatures in valves, otherwise assumed identical. However, there are indications that the further development of both junction transistors for very small currents and field-effect transistors will considerably assist in the solution of this problem.

36. Bandwidth and modulation

We have seen in the previous sections that we can design amplifiers with specific frequency characteristics. In practice it is also important to know what demands must be made on an amplifier or, in general, on a linear transfer system to obtain at the output a sufficiently true reproduction of the input signal. Further we must also know what distortion occurs as a result of the limitations imposed by realizable frequency characteristics.

As mentioned in Section 1, the signal in a measurement system is submitted, after initial linear amplification, to one or more processes before being recorded. It is obviously nonsensical to require that an amplifier transmits frequencies that will be eliminated at a later stage, i.e. during a further process or recording, and will not be reproduced in the final output. In the following discussion we shall start from the concept that it is sensible to attempt to make the output of an amplifier reproduce a faithful image of the input conditions. We shall therefore not discuss the frequency-dependence that is required in most practical circuits, either to compensate for certain shortcomings in the rest of the apparatus or to improve the ratio of the signal to noise or interference. Regarding the latter we should note that, whilst the determination of the optimum transmission characteristic for a given signal form and noise spectrum, is generally extremely complicated (and its discussion falls outside the framework of this book), it is not difficult to select the most suitable transmission characteristic for the majority of practical measurement circuits.

The demands made on the frequency characteristic of a linear transfer system are first of all determined by the frequency composition of the expected input signal. Sometimes this composition follows directly from the nature of the signal, but it often happens that our only available data concern the nature of the signal against time. The frequency composition can be derived in the latter case by means of Fourier analysis, which is based on the fact that any signal which is realized by physical means, may be considered as a, possibly infinite, number of sine and cosine functions, where the argument is proportional to the time. These periodic functions start at the instant in time $t = -\infty$ and continue until $t = +\infty$.

Because

$$\cos \omega t = \frac{e^{j\omega t} + e^{-j\omega t}}{2} \quad \text{and} \quad \sin \omega t = \frac{e^{j\omega t} - e^{-j\omega t}}{2j}$$

we can write the realizable time function $g(t)$ as a sum or integral of functions $e^{j\omega t}$ with positive and negative ω:

$$g(t) = \int_{-\infty}^{+\infty} h(\omega)e^{j\omega t}d\omega \tag{36.1}$$

where $h(\omega)$ is given by:

$$h(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} g(t)e^{-j\omega t}dt \tag{36.2}$$

Since $g(t)$ is a real function, we can write for $h(\omega)$:

$$h(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} g(t)\cos\omega t\,dt - \frac{j}{2\pi}\int_{-\infty}^{+\infty} g(t)\sin\omega t\,dt$$

so that

$$h(-\omega) = h^*(\omega) \quad \text{and} \quad |h(-\omega)| = |h(\omega)|$$

where $h^*(\omega)$ is the complex conjugated function of $h(\omega)$.

When discussing bandwidths, we should remember that $-\omega$ and $+\omega$ represent physically the same frequency:

$$2\cos\omega_0 t = e^{j\omega_0 t} + e^{-j\omega_0 t}$$

does not contain the two frequencies $-\omega_0$ and $+\omega_0$, but only the positive, physical, frequency ω_0.

In the special case that $g(t)$ is a periodic function with period T (this case often occurs in practice but then only over a finite period of time) the spectrum will only contain frequencies which are a multiple of $\omega_0 = 2\pi/T$. Equation (36.1) then reverts to:

$$g(t) = \sum_{-\infty}^{+\infty}{}_n\, c_n\, e^{jn\omega_0 t} \tag{36.3}$$

where n only assumes integers. The coefficient c_n is given by:

$$c_n = \frac{1}{T}\int_{t_0}^{t_0+T} g(t)e^{-jn\omega_0 t}dt \tag{36.4}$$

Written in sine and cosine functions, this gives the well-known Fourier series:

$$g(t) = b_0 + \sum_{n=1}^{\infty} b_n \cos n\omega_0 t \; + \; \sum_{n=1}^{\infty} a_n \sin n\omega_0 t \qquad (36.5)$$

where
$$b_0 = \frac{1}{T} \int_{t_0}^{t_0+T} g(t)\,\mathrm{d}t$$

$$b_n = \frac{2}{T} \int_{t_0}^{t_0+T} g(t) \cos n\omega_0 t\,\mathrm{d}t$$

$$a_n = \frac{2}{T} \int_{t_0}^{t_0+T} g(t) \sin n\omega_0 t\,\mathrm{d}t$$

and t_0 is any chosen reference instant in time.

For example, if we have the square wave function from $t = -\infty$ to $t = +\infty$ of Fig. 36-1, all coefficients b will equal zero, because $g(t)$ is here an odd function of t, $g(-t) = -g(t)$. By integration from $-\frac{1}{2}T$ to $+\frac{1}{2}T$ we find for the coefficients a_n:

$$a_n = \frac{4}{T} \int_0^{\frac{1}{2}T} \sin n\omega t\,\mathrm{d}t = -\frac{2}{n\pi} \cos n\,x \Bigg|_{x=0}^{x=\pi}$$

so that $a_n = 0$ when n is even, and $a_n = 4/\pi n$ when n is odd. Therefore

$$g(t) = \frac{4}{\pi}\left(\sin \omega_0 t + \frac{1}{3} \sin 3\omega_0 t + \frac{1}{5} \sin 5\omega_0 t + \dots\right)$$

The triangular function from Fig. 36-2 is obtained by integration from the square wave function and will therefore also exclusively contain odd harmonics. At the indicated selected zeropoint, the following equation is valid for this function:

$$g(t) = \frac{8}{\pi^2}\left(\sin \omega_0 t + \frac{1}{9} \sin 3\omega_0 t + \frac{1}{25} \sin 5\omega_0 t + \dots\right)$$

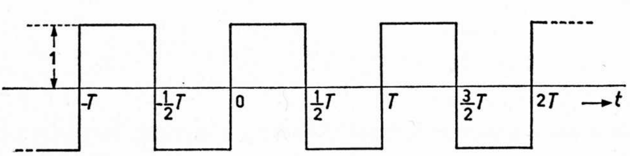

Fig. 36-1

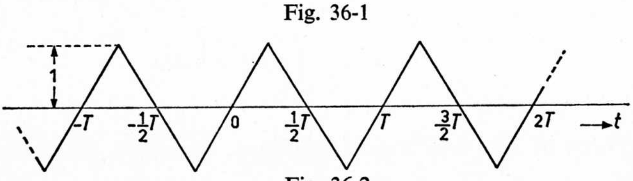

Fig. 36-2

A careless interpretation of the results of Fourier analysis easily leads to paradoxical conclusions. In practice, we have always to deal with signals that are either only present during a finite period of time $t_1 - t_2$, i.e. are then different from zero, or are only observed during that period of time. According to the above, these signals can be considered as a combination of sinusoidal signals that have been present from $t = -\infty$ and continue until infinity. In order to amplify each of these signal components, the amplifier should be in operation from $t = -\infty$ and continue to work until $t = +\infty$, i.e. even after the signal has been passed. It would be wrong to conclude from this that it is impossible to amplify a signal of finite duration faithfully: the amplifier does not have to respond to each signal component separately, but to the sum of all these as an entity, and this will obviously be zero outside the finite duration.

Another wrong conclusion which could be drawn from Fourier analysis would be that it is possible to determine the value of the components before the beginning of the signal by means of suitable filtering, and calculate $g(t)$ from this; this is tantamount to foretelling the future. The exact reverse is true; it follows from the fact that nothing can be predicted, that with physically realizable filters there must obviously be a connection between the amplitude and the phase characteristics.

An correct conclusion that can be drawn from Fourier analysis gives a better understanding regarding the requirements which a linear transfer system must satisfy: for each signal of finite duration and different from zero, the frequencies of the components continue to infinity but that at the same time the amplitudes are approaching zero at the higher frequencies. From this it follows that, although it is correct to state that the perfect rendering of signals of finite duration requires an amplitude characteristic which is flat up to infinitely high frequencies, the error introduced by transmitting only a finite frequency band need not be necessarily large. We shall now consider the order of magnitude of this error.

An infinitely long continuous signal can have a frequency spectrum with finite bandwidth: signal $g(t) = 1$ only contains frequency zero; signal $g(t) = \cos \omega_0 t$ only frequency ω_0 and the bandwidth is zero in both cases. For a signal $g_1(t)$ which only equals 1 during a period of time $2T$, and zero for all other times, we can apply equation (36.2) to obtain the Fourier spectrum and thus find

$$h_1(\omega) = \frac{1}{\pi\omega} \sin \omega T$$

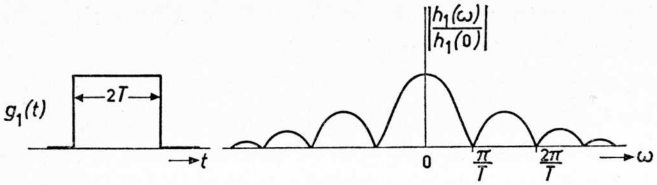

Fig. 36-3

The signal and corresponding spectrum are shown in Fig. 36-3. It will be seen that because of the finite duration of the signal, the frequency spectrum actually continues to infinity, whilst the amplitudes decrease by ω^{-1}.

Equation (36.2) can also be used to calculate the spectrum of signal $g_2(t)$, which equals $\cos \omega_0 t$ for a period of $2T$, and is zero for all other times. Better insight will be acquired when writing $g_2(t)$ as $g_1(t) \cos \omega_0 t$, so that a component $e^{j\omega t}$ of g_1 gives rise to $e^{j\omega t} \cos \omega_0 t = \frac{1}{2}e^{j(\omega+\omega_0)t} + \frac{1}{2}e^{j(\omega-\omega_0)t}$. The spectrum of g_2 will consist of two parts which follow from the spectrum shown in Fig. 36-3 by displacement over frequencies $+\omega_0$ and $-\omega_0$ (Fig. 36-4). The finite duration of the signal is here also taken into account in the expansion of the spectrum, compared to that of the corresponding signal of infinite duration. These examples also show that the frequency spectrum becomes wider if the signal is of shorter duration. In other words, the inaccuracy in

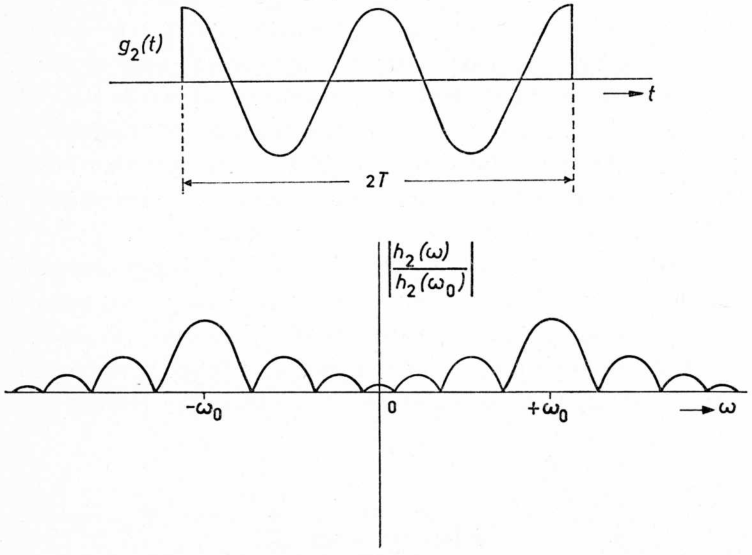

Fig. 36-4

determining the frequency is greater when the inaccuracy in the time determination is smaller.

This is analogous to "Heisenberg's Uncertainty Relation" in wave mechanics, which states that the accuracy in determining the energy of a particle becomes smaller as the accuracy in the determination of time or position becomes greater. Heisenberg gives the following definitions of the indeterminate values, which are logical and easily manipulated.

The "time gravity center" t_z is determined, using the notation of (36.1) and (36.2), by:

$$t_z = \frac{\int_{-\infty}^{+\infty} |(g(t)|^2 \, t \, \mathrm{d}t}{\int_{-\infty}^{+\infty} |g(t)|^2 \, \mathrm{d}t}$$

The "time inertia moment" I with respect to this point of gravity is given by:

$$I = \frac{\int_{-\infty}^{+\infty} |g(t)|^2 \, (t - t_z)^2 \, \mathrm{d}t}{\int_{-\infty}^{+\infty} |g(t)|^2 \, \mathrm{d}t}$$

The indeterminate value Δt in the time, i.e. the duration, is then defined by $\Delta t = \sqrt{I}$. The indeterminate value $\Delta\omega$ in the frequency is similarly derived from $h(\omega)$. We find then for the thus defined values: $\Delta t \cdot \Delta\omega \geqslant \frac{1}{2}$, where the minimum value $\frac{1}{2}$ is reached when the time function follows the Gaussian curve:

$$g(t) = g(t_0) \, \mathrm{e}^{-a(t - t_0)^2}$$

The frequency spectrum has then the same form (Fig. 36-5).

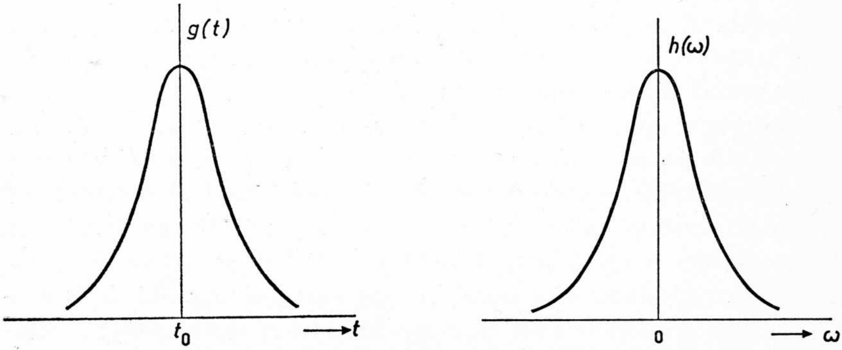

Fig. 36-5

Consider the case when a signal of very long duration $g(t)=1$ is measured with an amplifier over a short period from t_1 to t_2 (Fig. 36-6). If the amplification is equal for all frequencies and amounts to A_0, and the amplifier therefore has an infinitely large bandwidth, the output signal will be equal to A_0 during the period t_1-t_2, and the reproduction will be exact. However, if the amplification falls according to $A_0/(1+j\omega\tau)$, the output signal during this period will be $A_0(1-e^{-(t-t_1)/\tau})$. Whether one considers the mean value or the final value of this signal as the measured one, the error will be smaller in both cases when t_2-t_1 increases with respect to τ. Although the error will only become zero after an infinitely long observation time, it can nevertheless be made small enough by selecting a sufficiently long observation time. That frequencies must be passed for which the amplification decreases considerably, is in this case entirely the consequence of the finite observation time.

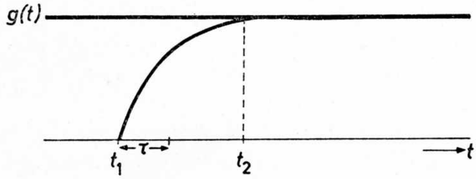

Fig. 36-6

The situation is different if there are frequencies in the range where the amplification decreases, which are not a consequence of the finite time of observation but which are an inherent part of the signal, and which will therefore also be present during long observation times. In this case the loss of a certain amount of essential information will be unavoidable, however long the observation. If the frequency content of the corresponding long-period signal is known, it is not difficult to determine the required bandwidth for a given period of observation and the admissable error.

In the case of signals of long duration we are able to make a distinction in bandwidth requirement for short or long observation times. This of course is no longer possible if the signal has a duration time which is short compared with the observation time. Pointing out the error due to the finite bandwidth of the amplifier is much more difficult here. If little is known about the frequency composition of a signal of long duration, one can draw conclusions with greater certainty from the response of the amplifier to step-function signals, i.e. by considering the step-function response. This was in the above case $1-e^{-t/\tau}$. Detail in the signal with a duration of approximately τ will be distorted because of this, but coarser details will be distorted to a smaller extent. If little is known of the nature of the observed signal, it is advisable to design the system so that the response time of the system (the time the system allows to elapse when processing a step-function from close to the initial value to close to the final value) is small compared to the duration of the details which one requires to observe or to process. It should be realized then that one can expect that details, occuring during shorter times, will be lost to a great extent.

Much use is made of the fact that when a signal is multiplied by sin $\omega_c t$, displacements by a shift ω_c in the frequency spectrum occur. This is called "modulation". For example, in radio technique it is in this way possible to transmit a low-frequency acoustical signal at high frequencies and ensure selective reception. In this case, the frequency of the "carrier wave" is usually much higher than that of the modulating signal components. If $v_s = V_s \cos \omega_s t$ represents one of the components of the l.f. signal and $v_c = V_c \cos \omega_c t$ the carrier wave, modulation will occur as indicated in Fig. 36-7, so that the

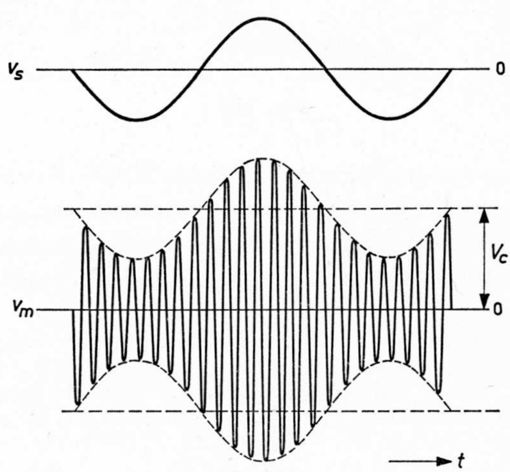

Fig. 36-7

carrier wave will be present also in the absence of a modulating signal. We then find for the instantaneous value of the modulated signal:

$$v_m = (V_c + kV_s \cos \omega_s t)\cos \omega_c t =$$

$$V_c\left(1 + k\frac{V_s}{V_c}\cos \omega_s t\right)\cos \omega_c t = V_c\,(1 + m \cos \omega_s t)\cos \omega_c t$$

where m = "modulation depth". To avoid difficulties in reception the modulation depth is kept smaller than unity for the largest modulating amplitude. As we shall see when discussing the application of modulation for measurement purposes, modulation is there often stronger and the carrier wave may even be completely suppressed.

A signal component with frequency ω_s causes two components in the modulated signal with frequencies $\omega_c - \omega_s$ and $\omega_c + \omega_s$, since the term $\cos\omega_s t \cos\omega_c t$ can be written as $\frac{1}{2}\cos(\omega_c - \omega_s)t + \frac{1}{2}\cos(\omega_c + \omega_s)t$. If the l.f. signal is composed of components with frequencies between $\omega_{\min}$ and $\omega_{\max}$ the modulated signal will contain the carrier wave as well as components with frequencies in the bands between $\omega_c - \omega_{\max}$ and $\omega_c - \omega_{\min}$, and between $\omega_c + \omega_{\max}$ and $\omega_c + \omega_{\min}$. These are called "side bands" (Fig. 36-8).

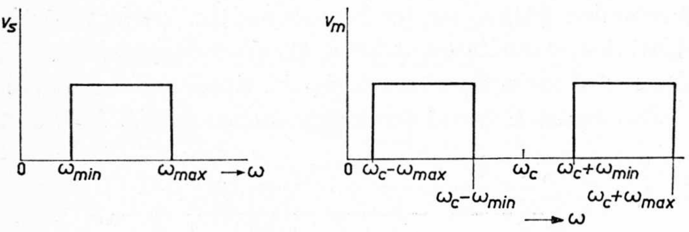

Fig. 36-8

There are many reasons for using amplitude modulation in measurement electronics. Perhaps the most frequent application is the avoidance of the necessity of amplifying very low frequency voltages and currents and hence the elimination of the effect of drift and flicker noise in amplifiers.

Fig. 36-9 shows a measurement arrangement where use is made of amplitude modulation.

We can measure the spectral distribution of a light beam by means of a thermo-couple. A rotating perforated disc periodically interrupts

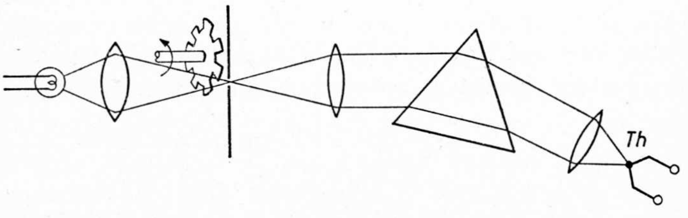

Fig. 36-9

the beam, and the thermo-couple will produce an alternating voltage, the amplitude of which is a measure of the light intensity. The frequency of the "carrier wave" must of course be chosen so that the thermo-couple cell can follow it and in practice it is restricted to a few tens of c/s. Considerably higher frequencies can be used for other optical transducers, e.g. photo-cells and photo-transistors. The variations to be measured are always relatively slow and usually correspond in the case of a thermo-couple to frequencies of less than 1 c/s. In order to obtain a maximum signal-to-noise ratio, only a very narrow band around the carrier-wave frequency must be passed. Because of the low frequency, this is not feasible by means of tuned circuits or filters, particularly when the frequency of the carrier wave may also change. In such a case we have recourse to "synchronous detection", as described in Section 40.

Measurements of this kind generally require a faithful transfer of the signal. Should signal v_s, which contains components with frequencies up to $\omega_{\max}$, be amplified without modulation, the amplifier must possess an amplitude characteristic which is flat up to $\omega_{\max}$ as well as a phase characteristic, where the phase is proportional to ω up to $\omega_{\max}$.

The complex amplification in the frequency range up to $\omega_{\max}$ must in this case satisfy $A(\omega) = A_0 e^{j\omega\tau}$, where A_0 and τ are constants.

With

$$v_s(t) = \int_{-\omega_{\max}}^{\omega_{\max}} h(\omega) e^{j\omega t} d\omega$$

the output signal will become:

$$v_o(t) = \int_{-\omega_{\max}}^{\omega_{\max}} A_0 e^{j\omega\tau} h(\omega) e^{j\omega t} d\omega = A_0 \int_{-\omega_{\max}}^{\omega_{\max}} h(\omega) e^{j\omega(t+\tau)} d\omega = A_0 v_s(t+\tau)$$

so that v_o will have the same form as v_s, but is displaced with respect to v_s by a time τ. Since the signal cannot appear at the output before it has entered the amplifier, τ must be negative or at the utmost zero.

What are the requirements if we wish to amplify a signal v_m, the amplitude of which is modulated by v_s? This question can be answered by considering a term of v_m

$$2 \cos \omega_s t \cos \omega_c t = \cos(\omega_c + \omega_s)t + \cos(\omega_c - \omega_s)t$$

If component $\cos(\omega_c + \omega_s)t$ is amplified A_+ times and undergoes a phase shift φ_+, and component $\cos(\omega_c - \omega_s)t$ is amplified A_- times and undergoes a phase shift φ_-, we obtain after amplification the term:

$$A_+ \cos[(\omega_c + \omega_s)t + \varphi_+] + A_- \cos[(\omega_c - \omega_s)t + \varphi_-]$$

For faithful transfer, this component must be changeable into a term of the form

$$2\,A \cos \omega_s(t + \tau) \cos(\omega_c t + \varphi)$$

where A, τ and φ are independent of ω_s; A because amplification must be independent of ω_s; τ because the same delay must occur for all values of ω_s; and φ because the carrier wave may only be submitted to a phase shift which is the same for all ω_s. When writing out both expressions in terms containing $\cos \omega_c t \cos \omega_s t$, $\cos \omega_c t \sin \omega_s t$, $\sin \omega_c t \cos \omega_s t$ and $\sin \omega_c t \sin \omega_s t$, and making the corresponding terms of this expansion equal to each other, we obtain:

$$\begin{aligned} A_+ \cos \varphi_+ + A_- \cos \varphi_- &= 2\,A \cos \varphi \cos \omega_s\tau \\ -A_+ \sin \varphi_+ + A_- \sin \varphi_- &= -2\,A \cos \varphi \sin \omega_s\tau \\ -A_+ \sin \varphi_+ - A_- \sin \varphi_- &= -2\,A \sin \varphi \cos \omega_s\tau \\ -A_+ \cos \varphi_+ + A_- \cos \varphi_- &= 2\,A \sin \varphi \sin \omega_s\tau \end{aligned}$$

Adding the first and last equations, we obtain:

$$A_- \cos \varphi_- = A \cos(\varphi - \omega_s\tau)$$

and after subtraction:

$$A_+ \cos \varphi_+ = A \cos(\varphi + \omega_s\tau)$$

Adding the second and third equations yields:

$$A_+ \sin \varphi_+ = A \sin(\varphi + \omega_s\tau)$$

and after subtraction:

$$A_- \sin \varphi_- = A \sin(\varphi - \omega_s\tau)$$

Apart from irrelevant variants it follows:

$$A_+ = A_- = A$$

and

$$\varphi_+ = \varphi + \omega_s\tau, \quad \varphi_- = \varphi - \omega_s\tau$$

so that in this case amplification must also be constant over the entire band, whilst the phase must again be linear with ω but is not required to be zero for the carrier-wave frequency ω_c. Compared to the non-modulated version, the frequency range over which these requirements for amplification must be satisfied, is now relatively much narrower, so that it is easier to meet the requirements.

Closely connected to the above and of practical importance is the fact that it is easy to derive a circuit with the same properties concerning the relative deviation β with respect to a central frequency ω_0 from another circuit which has the correct amplitude and phase characteristics against frequency ω in the range $0 \leqslant \omega \leqslant \omega_{max}$. If a capacitor with a value satisfying $C=1/\omega_0{}^2L$ is connected in series with a coil L, the impedance of the series resonant network will be $j\omega L+1/j\omega C=j\omega L+\omega_0{}^2L/j\omega=j\omega_0 L(\omega/\omega_0-\omega_0/\omega)=j\omega_0\beta L$, i.e. the behaviour of this impedance with respect to $\omega_0\beta$ is similar to that of the single coil with respect to ω. The admittance of a parallel circuit of a capacitor C and a coil with a value satisfying $L=1/\omega_0{}^2C$ will likewise have the same behaviour with respect to $\omega_0\beta$ as the admittance of C with respect to ω. Therefore, when we have a circuit with an impedance $Z(\omega)$, and when all coils and capacitors in this circuit are replaced in the manner described by series or parallel resonant circuits, the impedance of the new circuit will be $Z(\omega_0\beta)$. All properties of the original circuit in ω will transpose into the same properties for the new circuit in $\omega_0\beta$. The latter can be approximated for small values of β by $2(\omega-\omega_0)$. Fig. 36-10 gives an example of such a transformation.

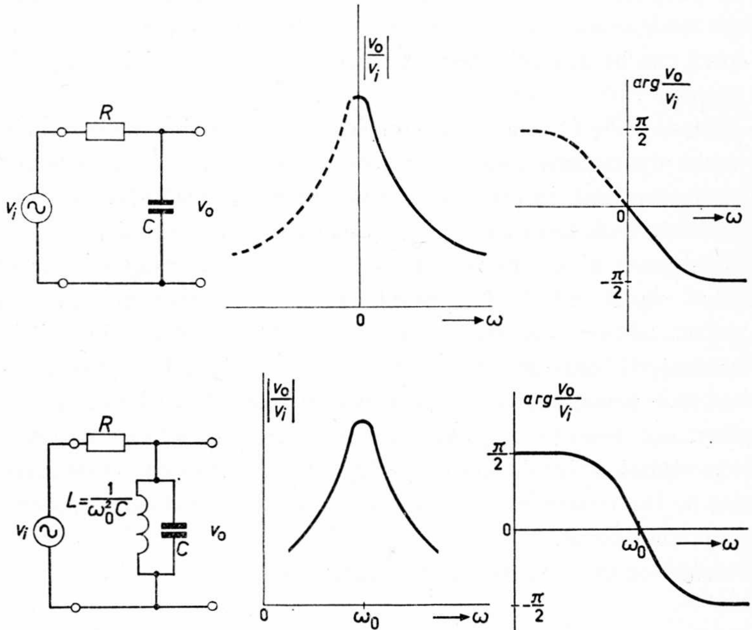

Fig. 36-10

Apart from amplitude modulation, where amplitude A of a carrier wave $A\cos(\omega t+\varphi)$ varies proportionally to the signal, there are other forms of modulation. The most obvious ones are those where frequency ω or phase φ

of the carrier wave are made functions of the signal. We speak then of frequency and phase modulation. Frequency modulation particularly, is often used, and we shall therefore restrict ourselves in the following to this type.

In telecommunications frequency modulation has the great advantage that it allows one to make the signal-to-noise ratio considerably better than with amplitude modulation. In measurement technique frequency modulation is, for example, used for recording a signal on magnetic tape. Frequencies below a certain frequency cannot be recorded directly on tape. The use of amplitude modulation would allow one to shift the entire frequency spectrum upward to a range which can be accepted, but it is then necessary to allow for amplitude variations of approximately 10 per cent caused by inhomogenities in the tape and variations in the distance between tape and recording head. In the case of frequency modulation, however, only fluctuations in the speed of the tape can produce distortion. These fluctuations can easily be made smaller than 1 per cent. If necessary, possible changes in the tape speed can be determined and corrected for by recording a signal with fixed frequency on a second track.

To determine the transfer of an f.m. signal when passing through a circuit, or to derive the requirements of a circuit to produce a transfer with specific distortion properties, in principle the same method could be used as with the amplitude-modulated signal, i.e. by using Fourier analysis.

We can speak of an instantaneous frequency in a frequency or phase modulated signal, which is defined as the derivative of the argument with respect to the time: $\omega_{\text{inst}} = \text{d(argument)}/\text{d}t$. To what calculations the Fourier analysis leads becomes, for example, obvious from the recurrent practical case where the following equation is valid for the instantaneous frequency: $\omega_{\text{inst}} = \omega_0 + \omega_d \cos \omega_s t$, with ω_0 = frequency of the carrier wave at zero modulation signal, ω_d = frequency sweep, i.e. the maximum deviation assumed by the instantaneous frequency against ω_0 and ω_s = frequency of the modulating signal.

The definition of ω_{inst} gives in this case:

$$\text{argument} = \int_0^t (\omega_0 + \omega_d \cos \omega_s t)\text{d}t = \omega_0 t + \frac{\omega_d}{\omega_s} \sin \omega_s t + \varphi_0$$

so that we find for the carrier wave:

$$f(t) = A \sin\left(\omega_0 t + \frac{\omega_d}{\omega_s} \sin \omega_s t + \varphi_0\right)$$

where A is the constant signal amplitude of the signal.

With $\varphi_0 = 0$ and $\omega_d/\omega_s = \delta$ this becomes:

$$f(t) = A \sin(\omega_0 t + \delta \sin \omega_s t)$$

for which can be written:

$$f(t) = A\,[\sin \omega_0 t \cos(\delta \sin \omega_s t) + \cos \omega_0 t \sin(\delta \sin \omega_s t)]$$

We have also:

$$\cos(\delta \sin \omega_s t) = J_0(\delta) + 2 \sum_{n=1}^{\infty} J_{2n}(\delta) \cos 2n\omega_s t$$

and

$$\sin(\delta \sin \omega_s t) = 2 \sum_{n=0}^{\infty} J_{2n+1}(\delta) \sin(2n+1)\omega_s t$$

where J_k is the Bessel function of the first kind and k-th order. This substituted in $f(t)$:

$$\begin{aligned} A^{-1}f(t) = {} & J_0(\delta)\sin \omega_0 t + J_1(\delta)[\sin(\omega_0 + \omega_s)t - \sin(\omega_0 - \omega_s)t] + \\ & J_2(\delta)[\sin(\omega_0 + 2\omega_s)t + \sin(\omega_0 - 2\omega_s)t] + \\ & J_3(\delta)[\sin(\omega_0 + 3\omega_s)t - \sin(\omega_0 - 3\omega_s)t] + \dots \end{aligned}$$

It follows that a modulation signal $V_s \sin \omega_s t$, which only gives two components with frequencies $\omega_0 + \omega_s$ and $\omega_0 - \omega_s$ in the Fourier spectrum of the carrier wave when using amplitude modulation, produces an infinite number of components with frequency modulation. The amplitudes of these components are Bessel functions with argument $\delta = \omega_d/\omega_s$ i.e. the ratio of the maximum frequency deviation in the carrier wave to the frequency of the modulation signal. Fig. 36-11 shows some of the frequency spectra for different values of the frequency width ω_d at a constant value of modulation frequency ω_s, whilst Fig. 36-12 shows spectra for different values of ω_s at constant ω_d. If the modulation signal consists of more components, the

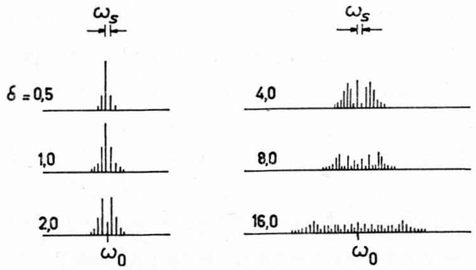

Fig. 36-11

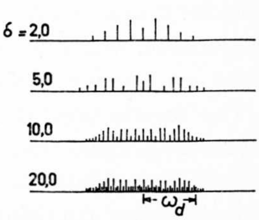

Fig. 36-12

frequency spectrum of the carrier wave will become particularly complex because of the occurrence of all sum and difference frequencies of these components and their harmonics. However, this analysis shows quite clearly that, with the same modulation signal, an f.m. signal usually occupies a much wider bandwidth than an a.m. signal. This is also the reason why carrier-wave frequencies for f.m. radio transmissions are rather high (approximately 100 Mc/s), to avoid overlapping of the frequency bands.

It is obvious that the above analysis, and in general the calculation of the transfer function of circuits, for f.m. signals is complicated by the fact that the instantaneous frequency is not a limitation but an extension of the concept of frequency used so far, and that relations and calculation rules which are valid for the latter can not be applied to instantaneous frequencies without correction. The concept of impedance, in particular, has no real sense here. An extensive, and often quite complicated literature exists for calculations involving f.m. signals. For measurement technique we are specially interested in the results derived in 1937 by Carson and Fry. They showed that for the calculation of the transfer function for an f.m. signal $f(t)$ by a linear circuit, for example an impedance $Z(\omega)$, it is permissible to consider the instantaneous frequency as a variable ordinary frequency, and to use the calculation rules valid for the latter, provided the derivatives of $Z(\omega)$ to frequency and of $f(t)$ to time are small compared to $Z(\omega)$ and $f(t)$ respectively. In other words the functions $Z(\omega)$ and $f(t)$ must not show abrupt changes. In the very simple case where a frequency-modulated current $i(t)$ passes through a coil L, no restrictions need to be applied to $i(t)$ to cal culate the voltage across the coil. Current $i(t) = i_0 \sin \int_0^t \omega_{\text{inst}}\, dt$ produces a voltage $v(t) = L\, di(t)/dt = Li_0\, \omega_{\text{inst}} \cos \int_0^t \omega_{\text{inst}}\, dt$ across coil L, which does not depend on the nature of $i(t)$ and is similar to the relation valid for a sinusoidal current.

Apart from amplitude, frequency and phase modulation, many other ways exist for transmission of information, such as pulse-width and pulse-height modulation, where the width or height respectively of a periodic rectangular pulse is a function of the modulation signal.

The reverse process of modulation, i.e. the extraction of the original modulating signal from a modulated signal, is known as demodulation or detection. We shall discuss circuits for the accurate modulation and detection of signals after considering some basic circuits used for these purposes.

37. Oscillation

When feedback systems were discussed, we noted that an output signal may occur without an input signal being present. We then assumed that the situation of "self-generation" or "oscillation" did not occur and that linear amplification was maintained. We shall now proceed to discuss circuits which do oscillate, not only to acquaint ourselves with the range of behaviour in this case, but also because we must know to what extent oscillation in non-oscillatory circuits can occur because of the ageing of components or instability of supply sources. The "stability criteria" which will be derived in the next section will supply information on whether a system will oscillate or not. In this section we shall discuss the mechanism of oscillation and the practical design of oscillators.

Oscillation is one of the most complex phenomena in electronics. It is therefore best to begin its study by starting from a relatively simple situation, such as a pentode with a parallel resonant circuit in its anode circuit. We assume, for greater ease, that the effect of the anode voltage on the anode current can be neglected. Apart from a possible input signal we also supply part of the anode signal to the grid and thus obtain the situation depicted in the circuit diagram of Fig. 37-1. Retaining the same symbols as used for feedback, we write for the fed-back part of the anode voltage $-kv_a$, where for the time being k is assumed to be constant and positive. Calculating the amplification from the input to the anode is simple. Without feedback we have:

$$A = \frac{v_a}{v_s} = \frac{-\mathrm{j}\omega SLR}{-\omega^2 LCR + \mathrm{j}\omega L + R}$$

where S is the transconductance of the pentode.
With feedback, we obtain:

$$A' = \frac{A}{1 + Ak} = \frac{-\mathrm{j}\omega SLR}{-\omega^2 LCR + \mathrm{j}\omega L(1 - kSR) + R} \tag{37.1}$$

and

$$|A'| = \frac{\omega SLR}{\{R^2(1-\omega^2 LC)^2 + \omega^2 L^2(1-kSR)^2\}^{\frac{1}{2}}}$$

$|A'|$ is given for a few values of kSR against frequency in the lower part of Fig. 37-1. The maximum is always attained for $\omega = \omega_0 = 1/\sqrt{LC}$ and has the value $|SR/(1-kSR)|$. For $kSR = 1$, the maximum will be infinite

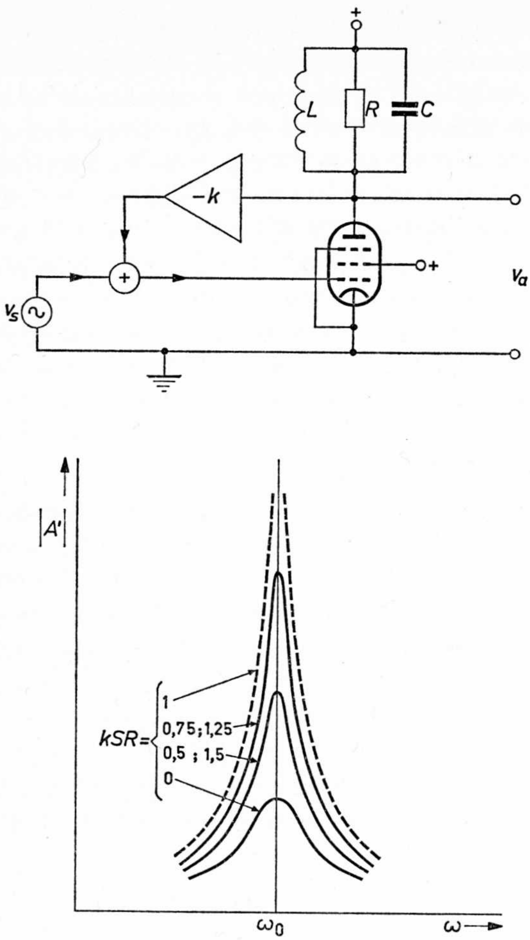

Fig. 37-1

(broken curve). As long as kSR remains smaller than 1, the absolute amplification will increase with k and the more so when the frequency is closer to ω_0. The relative width of the curve decreases at the same time i.e. the amplification becomes increasingly selective. The behaviour is just the reverse for values of k for which kSR is larger than 1: amplification and selectivity decrease with increasing k.

The infinitely large amplification for the frequency ω_0 when kSR equals 1, does not mean that a possibly present input signal v_s will produce an infinitely large signal at the output with that frequency (for this would require an infinitely large current through the valve), but that a finite output signal will be present without a signal at the input. In practice, there is always an input signal, although unintentional, because the circuit contains noise, for example from the valve, which may be imagined as represented by an equivalent signal at the input. This means that when signal v_s is absent, the amplification does not have to be infinitely great in order to supply a large output voltage. This voltage mainly contains components in a single narrow frequency band around ω_0 because of the very high selectivity, and will therefore be almost sinusoidal in shape as we have seen in Section 36, with slowly varying amplitude and frequency. This phenomenon, when an output signal is provided without the presence of an input signal, is called "oscillation".

Although it is not necessary for the occurrence of oscillation that the amplification is infinitely great and therefore kSR exactly equals 1, the required amplification is nevertheless so great that kSR must only differ from 1 to a very small extent. The order of magnitude of the permissible deviation can be estimated by considering that the amplification increases $1/(1-kSR)$ times because of feedback. This multiplication factor must therefore correspond to the ratio of the noise voltage on the anode with feedback, i.e. the wanted output voltage, to the noise voltage on the anode without feedback, in the same bandwidth. Normally we find a value for $1-kSR$ which is smaller than 10^{-8}.

Assuming (usually correctly) that the anode current noise contributes most to the total noise voltage, we find for the equivalent noise voltage v_i at the input $v_i{}^2{}_{,\mathrm{rms}} = 4\,k_BTR_{\mathrm{eq}}\mathrm{d}f = 2\pi^{-1}\,k_BTR_{\mathrm{eq}}\mathrm{d}\omega$ where k_B = Boltzmann's constant and T = absolute temperature. The absolute amplification is found from (37.1) so that the voltage across the anode is:

$$v_a{}^2{}_{,\mathrm{rms}} = \int_0^\infty \frac{2}{\pi}\,k_BT\,R_{eq}\,\frac{\omega^2\,S^2L^2R^2\,\mathrm{d}\omega}{R^2(1-\omega^2LC)^2+\omega^2L^2(1-kSR)^2} =$$

$$\frac{k_B\,T\,R_{eq}\,S^2R}{C(1-kSR)} = \frac{\omega_0\,k_B\,T(kSR)^2\,R_{eq}}{\omega_0\,RCk^2\,(1-kSR)}$$

which can be written, with $kSR \approx 1$ and $\omega_0RC = R/\omega_0L = Q_0$, the quality factor of the resonant circuit, as

$$v_a{}^2{}_{,\mathrm{rms}} = \frac{\omega_0\,k_B\,T\,R_{eq}}{Q_0\,k^2\,(1-kSR)}$$

We thus find the required value of $1 - kSR$:

$$1 - kSR = \frac{\omega_0\,k_B\,T\,R_{eq}}{Q_0\,k^2\,v_a{}^2{}_{,rms}} = \frac{\omega_0\,k_B\,T\,R_{eq}}{Q_0\,v_g{}^2{}_{,rms}}$$

where v_g is the oscillatory voltage on the control grid.

Another, rougher, method of calculation is to assume that the amplification $SR/(1 - kSR)$ is not only valid for the resonant frequency, but also for a small width Δf around it. Δf represents the half width of the fed-back circuit, so that $\Delta f = f_0/Q$, where Q = the apparent quality factor of this circuit. We therefore find: $Q = Q_0/(1 - kSR)$. The contribution outside this region is neglected. We then find:

$$v_a{}^2{}_{,rms} = \left(\frac{SR}{1 - kSR}\right)^2 4\, k_B\, T\, R_{eq}\, \frac{f_0}{Q_0}(1 - k\, SR) =$$

$$\frac{4\, f_0\, k_B\, T\, R_{eq}\, S^2R^2}{Q_0(1 - kSR)} = \frac{4\, f_0\, k_B\, T\, R_{eq}(kSR)^2}{Q_0 k^2(1 - kSR)} \approx \frac{4\, f_0\, k_B\, T R_{eq}}{Q_0 k^2(1 - kSR)}$$

The error caused by the simplification is thus a factor $\pi/2$, which is not very important because we are only interested in the order of magnitude.

With practical values $Q_0 = 100$, $v_{g,rms} = 0.1$ volt, $f_0 = 10^6$ c/s and $R_{eq} = 1$ kΩ, we find, with $k_BT = 4 \cdot 10^{-21}$ W sec:

$$1 - kSR \approx 2 \cdot 10^{-11}$$

and for the half-width of the transmitted frequency band:

$$\Delta f = \frac{10^6 \cdot 2 \cdot 10^{-11}}{10^2} = 2 \cdot 10^{-7}.$$

There is no need to emphasize that such a very small value of $1 - kSR$ cannot be obtained by adjustment, and certainly cannot be maintained. That oscillation can nevertheless occur as a permanent phenomenon is due to two circumstances. Firstly, kSR is then not constant but a function of the signal amplitude so that kSR is greater than 1 for small signal amplitudes and gradually decreases to less than 1 for greater amplitudes. Furthermore, oscillations can be started because for $kSR > 1$ the natural modes of the system are not damped but increase. kSR will decrease because of this increase in amplitude and approach unity, with the result that the noise will now also contribute. This causes a further decrease in kSR and, as will be explained later, a stable equilibrium will indeed only occur at a value of kSR which is slightly smaller than 1. We shall illustrate this mechanism by means of a few calculations.

The natural modes of the system follow from the differential equation:

$$LRC\frac{d^2v_a}{dt^2} + L(1 - kSR)\frac{dv_a}{dt} + Rv_a = 0 \qquad (37.2)$$

which is obtained from (37.1) by replacing $j\omega$ by the operator d/dt and putting $v_s = 0$. When kSR is not constant but dependent on the value of v_a in some way or the other, (37.2) is a non-linear differential equation which can only be solved explicitly for some special forms of the relation between

kSR and v_a. It is nevertheless possible to learn something about the solution in the case of interest to us, which is that kSR is larger than 1 for small amplitudes of v_a and gradually decreases to below 1 with increasing amplitude of v_a. We can then visualize the solution as a series of consecutive solutions of differential equations with different constant values of kSR. For constant values of kSR, the solution of (37.2) is:

$$v_a = c_1 e^{\alpha_1 t} + c_2 e^{\alpha_2 t}$$

where c_1 and c_2 are integration constants and α_1 and α_2 are given by:

$$\alpha_{1,2} = \frac{kSR - 1}{2\,RC} \pm \sqrt{\left(\frac{kSR - 1}{2\,RC}\right)^2 - \frac{1}{LC}}$$

kSR will be greater than 1 at a small signal amplitude, with the result that α_1 and α_2 have a positive real part, so that the natural modes will increase. This causes a reduction of kSR, and in the absence of noise the amplitude of the natural modes would become constant for $kSR=1$, whilst the frequency would be exactly $\sqrt{1/LC}=\omega_0$. However, in the presence of noise the mechanism becomes different; when kSR has almost reached unity, the noise in the narrow transmitted band will also be amplified and contributes to the output signal, thus causing a further decrease in kSR. Only when $k=1-\varepsilon$, where ε is a small positive quantity, can an equilibrium state be reached; the noise makes a large contribution to the output signal, whilst the natural modes are damped and will decrease. This will slightly increase kSR, i.e. it tends to become closer to unity; the noise contribution will therefore increase so that the output signal is stable in amplitude. The contribution of the resonant frequencies becomes increasingly smaller and that of the noise correspondingly greater, because kSR more closely approaches 1. It should be mentioned that the damping of the natural modes is an extremely slow process: with $1-kSR\approx 10^{-11}$ the damping term for the natural modes becomes of the order $1/10^{11}RC$, so that a time of $10^{11}RC$ is required for a reduction of e times. This is approximately 10^{12} periods with $RC=Q_0/\omega_0$! It is nevertheless possible to state that the oscillator can be considered to be more a selective noise amplifier than a generator of resonant frequencies. This clearly shows that even with an oscillator having ideal components, the amplitude and the frequency will show small changes. The latter can be demonstrated even with simple measurement methods, if we take a design that has a resonant circuit with a low quality factor, a small amplitude of the oscillation, and possibly very powerful noise sources. However, when the oscillator is designed normally, it will nearly always be possible to neglect the changes discussed above with respect to those caused by variations of the components.

It would also be possible to amplify the noise to a large output signal without feedback, provided we make the amplification A very large. The additional use of a filter element would then make the output signal sinusoidal. However, it is not possible to achieve the selectivity we have calculated here without using feedback, even when incorporating quartz crystals which are the best practical filter elements.

A particulary large open-circuit amplification is of no use when feedback is used, as far as oscillation is concerned. Because of other considerations which will be discussed later, it is rarely necessary to make A larger than a few thousands. This means that the difference between zero and the term $1+Ak$ from (37.1) will always be less than 10^{-8}, so that Ak will be exactly -1, or the loop gain $(-Ak)$ equal to $+1$. The requirements following from these considerations for the amplitudes and arguments of these values are known as "oscillation criteria".

Although the above purports only to be a simple description of the mechanism, it nevertheless explains a fundamental property of oscillators, that of synchronization. When an external signal is applied to an oscillator, it is possible that the frequency of the output signal will equal that of the external signal. This may be explained by imagining that to the input of the above described circuit a sinusoidal signal is applied with a constant, small amplitude, the frequency of which slowly approaches that of the oscillator. In this case we shall have the normal oscillation signal plus the amplified external signal, the latter with a slowly increasing amplitude. This additional signal will eventually affect the value of kSR. A decrease in kSR has little effect on the amplification of the external signal, but a great effect on that of the oscillator signal lying at the central frequency. Both signals will be present at approximately the same amplitude for a short time. After this, the amplified external signal will rapidly increase in value and the oscillator signal will rapidly decrease, without playing an important part any longer. The result is that the oscillator is synchronized. Synchronization occurs more rapidly when the external signal is larger and the resonant circuit has a lower Q. With normal oscillators and good-quality resonant circuits, the synchronization span will be relatively narrow for small external signals.

It is obvious from the above that non-linearity in the system plays an essential part. Because of the fact that it is not possible for signals to become infinitely great, each system possesses an inherently limiting mechanism which ensures the required drop in gain beyond a specific signal amplitude. However, when seeking amplitude stability, better results are obtained with limiting circuits or components inserted for this purpose. A wide range is possible even with the simplest design. We shall illustrate this by giving two examples of limiting, both more or less extremes which are nevertheless capable of giving oscillator voltages of high purity.

The first example encompasses the long-tailed pair circuit of Fig. 37-2. As mentioned in Section 19, the ratio of the direct current I_2 through the

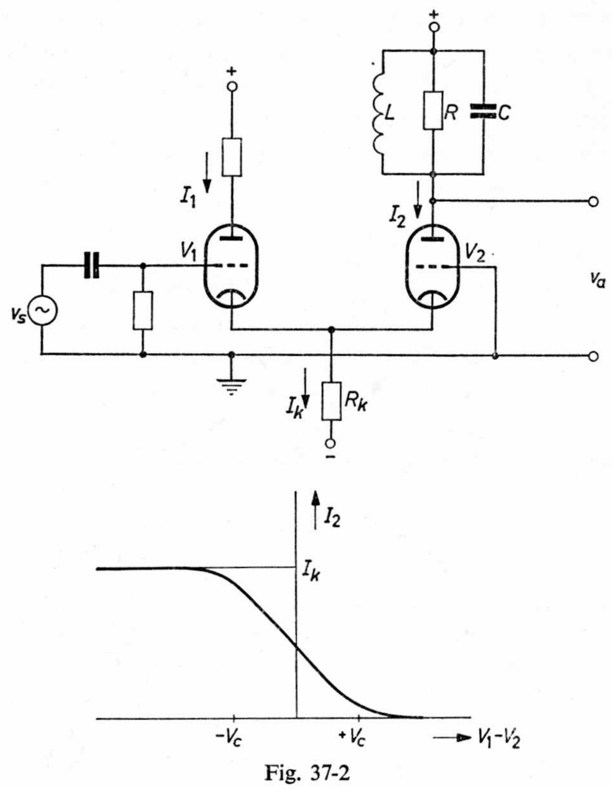

Fig. 37-2

second valve to the d.c. voltage difference ($V_1 - V_2$) between the grids is as shown in the lower part of this figure. The stage's behaviour is linear for small voltage differences, but the current is limited for out-of-balance voltages exceeding a few volts, or of a few tenths of a volt for transistors. If the left-hand grid is fed with a sinusoidal voltage v_s a constant distortion-free amplification will occur as long as v_s is sufficiently small. However, the more v_s approaches the knee voltage V_c, the more distortion will occur in the current waveform and the amplification of the fundamental component will decrease. The upper part of Fig. 37-3 shows the behaviour of the current through the second valve for various values of the input voltage. The rectangular form is the limiting case to which these waveforms approach at increasing amplitude of v_s.

When the frequency of the input signal equals the resonant frequency of the network in the anode circuit, and the latter has a good quality factor, the anode voltage will be mainly determined by the fundamental component of the current. At small input levels, this current will increase in proportion to the input signal, but for amplitudes of v_s larger than V_c, this increase will become very small, and the limiting amplitude for very large values of v_s becomes $4/\pi \approx 1.27$ times V_c. The result is a relation between v_a and v_s as shown in Fig. 37-3 (left-hand lower curve).

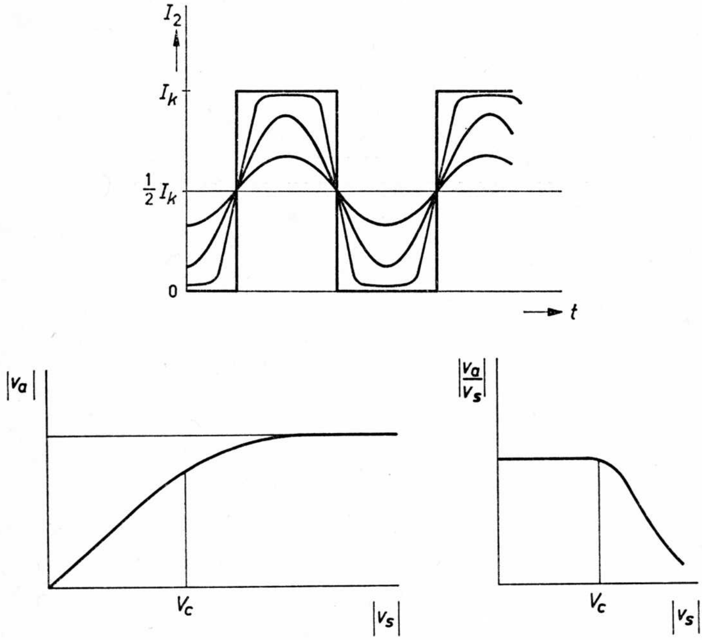

Fig. 37-3

Fig. 37-4 shows an example of the second method. Here we make use of a so-called NTC-resistor, i.e. a resistor with quite a large negative temperature coefficient, i.e. a few per cent per °C. By applying the signal voltage v_a across such a resistor and by ensuring that amplification depends sufficiently on the value of this resistor (i.e. $R_a \gg R_{NTC}$ and $1/\omega C \ll R_{NTC}$) the power which will be dissipated in the NTC-resistor will, when correctly designed, result in a smaller amplification through the reduction of the resistor's value. The degree of this effect depends, amongst other factors, on the dissipation constant of the resistor. The graph of Fig. 37-4 shows the relation between the resistor's value and the current through the resistor for a miniature NTC-resistor having a value of 100 kΩ at room temperature and no dissipation. The dissipation constant of such a resistor is approx. 0.25 mW/°C at a thermal conductance of about 1 mWs/°C. An example of the values to be expected for larger NTC-resistors is: dissipation constant 10 mW/°C and thermal conductance 1 Ws/°C. NTC-resistors are also known as thermistors.

With a long-tailed pair circuit as well as with the NTC-circuit, it is possible to reduce the amplification for large signals in order to obtain a stable oscillator circuit. There are nevertheless important differences between these methods. Whilst the circuit with the balanced stage operates in

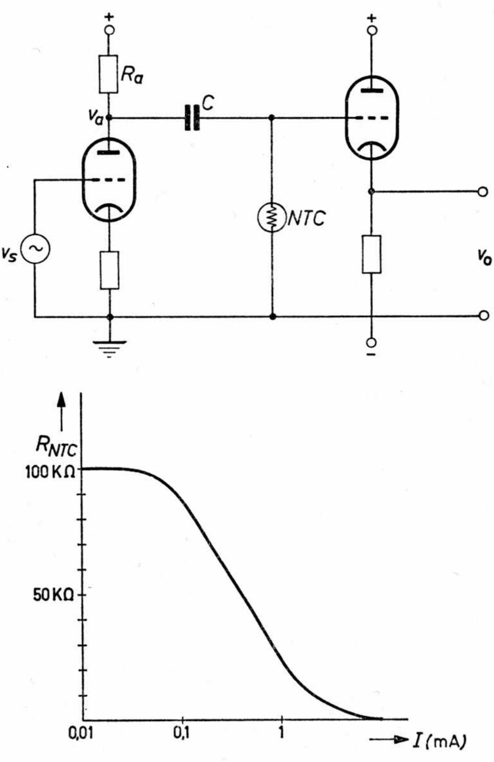

Fig. 37-4

an instantaneous fashion and therefore supplies a distorted signal which can be rendered sinusoidal in the following stages of the oscillator by filtering, the operation of the NTC-resistor is of an integrated form, the thermal time constant (a few seconds with the miniature resistor) is chosen very much larger than the oscillation period, so that the resistor value is not determined by the instantaneous signal amplitude, but by the mean signal value over a longer period. The resistor value over short durations can thus be considered as a constant and the circuit operates entirely linearly, so that only sinusoidal signals will occur. The difference stage does not possess a "memory", in contrast to the NTC-resistor, and this has a great effect on the oscillator's behaviour. With the instantaneous system, the signal rises to the value where the oscillation criterion is satisfied and remains at this value after that. With the integrating system, however, delays will occur; when the signal increases or decreases in amplitude, the amplification will only be adjusted after some time. This behaviour of thermal inertia shows a great likeness to that of an *RC*-time constant in an amplifier. For example, it would be possible by inserting three NTC-resistors in a circuit (analogous to what could occur in an amplifier with three *RC*-time constants) to obtain oscillation of the amplitude of the actual oscillator voltage. However, this phenomenon is very rare in practice, if only because a single NTC-resistor is used, and the remaining delay would have to be supplied by the typically smaller electrical inertias. On the other hand, the following situation is often met: because the loop gain is great for small signals, the signal will rapidly increase in amplitude until the maximum value has been reached. The regulation of the amplification is delayed so much that, when the amplitude of the oscillation begins to decrease, the regulation will continue until the oscillation ceases, after which the gain will once more slowly increase, and oscillation will re-occur as soon as the criterion is reached (left-hand side of Fig. 37-5). We call this blocking. When the delay caused by the integrating element is not as large, we obtain a signal with a periodically varying amplitude (right-hand side of Fig. 37-5).

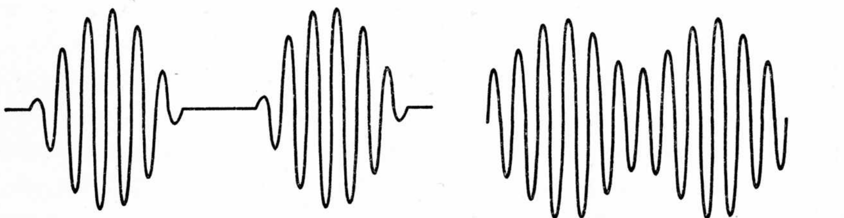

Fig. 37-5

That this form of oscillation, which shows great similarity to multivibrator action (Section 35), can occur in simple systems is shown by the following considerations. If in a feedback system the loop gain $-Ak$ varies in a sinusoidal fashion around the value 1, the amplitude of the oscillation will then vary as indicated in the lower part of fig. 37-6. As long as Ak is larger than 1, the amplitude will increase and the more rapidly, the greater the difference between $-Ak$ and 1. This is similarly valid for a decrease of the amplitude when $-Ak < 1$. Small changes in amplitude once again will be almost sinusoidal, so that a phase shift of 90° will occur between Ak and the amplitude adjustment. The remainder of the system then only needs to provide a further 90° phase shift for oscillation to occur. The amplitude changes will be smaller when the changes in the value of $-Ak$ occur more rapidly, so that the behaviour of the "transfer" function between $-Ak$ and the amplitude corresponds to the impedance of a capacitor. Normally, this effect is only troublesome, but use of it has been made for driving the vibration plate electrometer of page 296 at high frequencies.

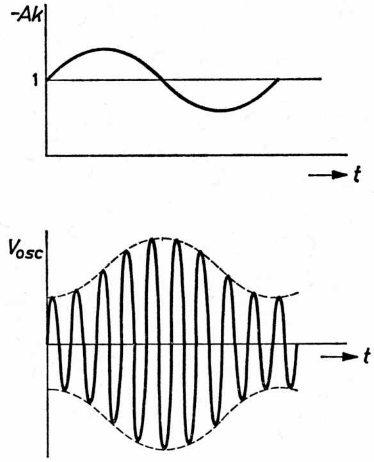

Fig. 37-6

Amplitude stabilization by means of a balanced stage (Fig. 37-2) has the great advantage that for sufficiently large input signals the magnitude of the output signal is very well defined. This is due to the fact that the latter is determined by the anode voltage and the current passing through the conducting valve, which in turn is mainly determined by the cathode resistance R_k or a current source replacing the latter. Since the action of filtering with passive elements or with active filters can allow a constant relation to be maintained between the amplitude of the fundamental component and the magnitude of the square wave, it is possible to design with this circuit oscillators having a very high voltage stability.

Amplitude stabilization by means of an NTC-resistor or a tungsten filament lamp possessing a rather large positive temperature coefficient, is applied in very many cases, particularly because the circuitry required can be kept very simple.

Apart from these two more or less extreme stabilization methods, there are several others, which usually have more in common with the integrating method than with the instantaneous one. The method of grid detection is frequently applied in radio technique. If a triode without grid-cathode bias is used (Fig. 37-7), grid current will pass during the positive half of an a.c. voltage applied to the grid. If the product of capacitance C and grid leak R_g is made large with respect to the period of the signal, the same rectifying effect will be obtained as with a diode: the grid will acquire such a negative bias that the charge passing for a small portion of cycle will compensate the discharge during the remainder of the cycle (right-hand side of Fig. 37-7). In practice this means that the grid bias is almost equal to signal amplitude v_s, and also that at increasing signal amplitudes, the time during

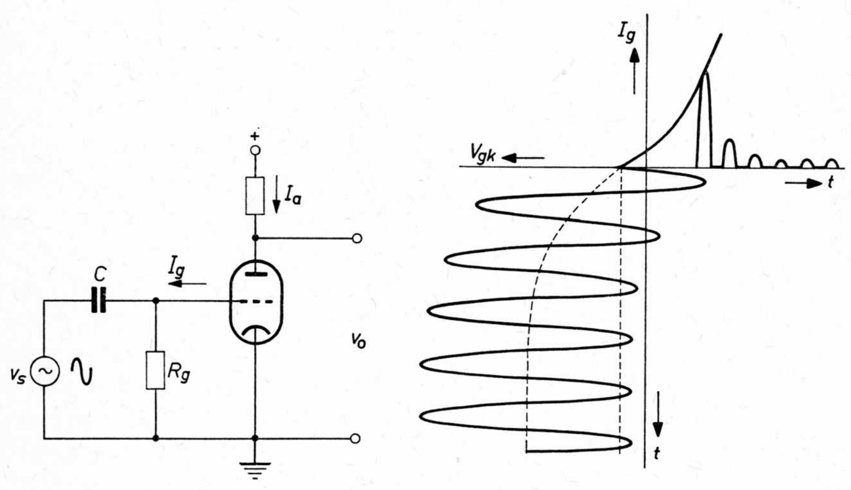

Fig. 37-7

which the valve is conductive is necessarily shorter because the grid voltages for which the value conducts (V_r in Fig. 37-8) are then experienced in a smaller portion of the period. Since the height of the anode current pulses always have approximately the same value I_{a0}, and since with a periodic signal consisting of small pulses, the amplitude of the fundamental is proportional to its area, the fundamental component $(i_a)_\mathrm{I}$ in the anode current will decrease beyond a specific input signal value (right-hand side of Fig. 37-8). This explains the limiting action of this circuit.

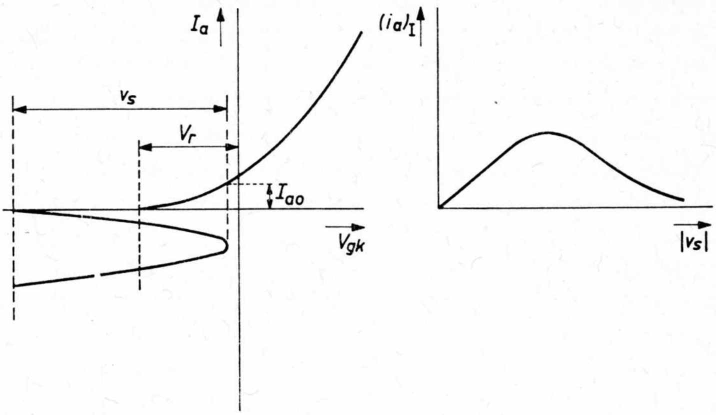

Fig. 37-8

Fig. 37-9 shows a basic limiting circuit, of much greater sensitivity than the one with grid detection. Whilst with the latter the control voltage was supplied by the a.c. signal on the grid, in the circuit of Fig. 37-9 we operate on the changes in the larger anode signal. As long as the amplitude of this signal remains smaller than the reverse bias on the diode (in this example

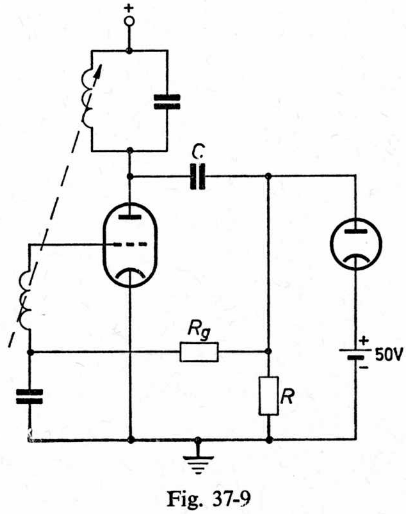

Fig. 37-9

50 volts) there will be no change in the working point of the valve and hence in amplification. However, as soon as the amplitude becomes larger than the reverse bias, the diode conducts during the positive peaks and will charge the capacitor to such a negative voltage that the charge supplied will be the same as that which is removed. At a sufficiently large value of the time constant RC, the negative grid bias will increase by the difference between the amplitude of the anode signal and the diode's bias. Such a clipping diode allows a particularly sharp drop in amplification above a certain signal value, which results in a good amplitude stability.

Another simple amplitude limiting circuit can be obtained by making the anode load of a valve, triode or pentode, large (Fig. 37-10). Beyond a specific amplitude of the control grid signal, the anode voltage will reach its minimum value during the grid's positive peaks, so that limiting occurs, resulting in

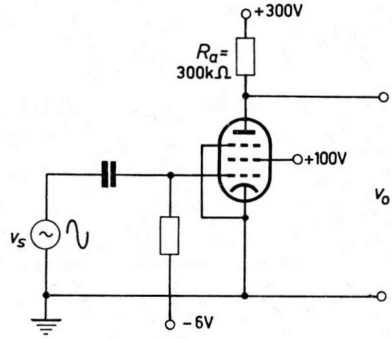

Fig. 37-10

a decreasing amplification of the valve. With a suitable choice of the working point the negative peaks of the grid signal will cause the valve to be cut off and in this manner a relation between the anode and grid voltage is obtained, as indicated on the right-hand side of Fig. 37-11 which is similar to the relation found with the long-tailed pair circuit. Using a pentode instead of a triode has the advantage of not taking a grid current despite the very low anode voltage, and thus not imposing an additional load on the signal source.

Another limiting circuit is shown in Fig. 37-12. As long as current I_a through the valve is less than 1 mA(=100 volts across 100 kΩ), diode D_1 will conduct, which makes the anode voltage V_A approximately 150 volts and V_P about 160 volts. As soon as I_a becomes greater than 1 mA, both

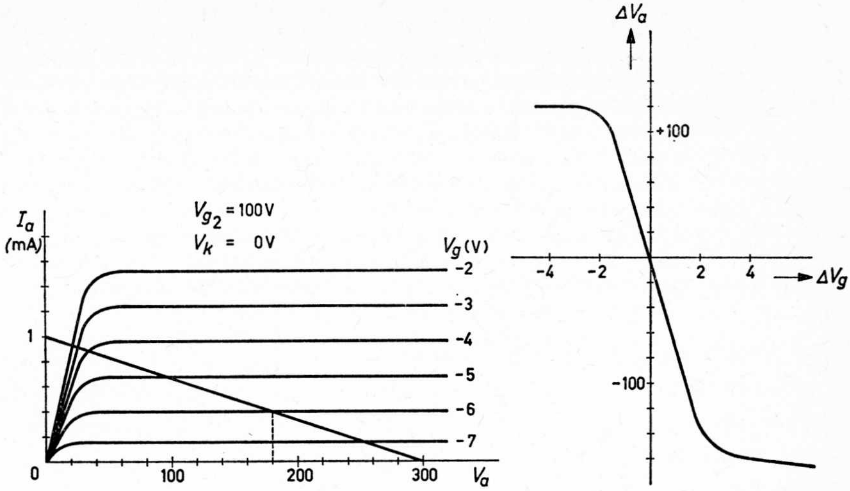

Fig. 37.11

points will drop in voltage until V_P has reached 150 volts, when D_2 starts to conduct. This occurs at $I_a \approx 1.1$ mA, and the relation between I_a and V_P is then as shown on the right-hand side of Fig. 37.12. When connecting a sinusoidal voltage with a sufficiently large amplitude to the grid, a squared voltage with an amplitude of approx. 11 volts will be found on the anode. In this case the amplification of the fundamental component will be inversely proportional to the signal value.

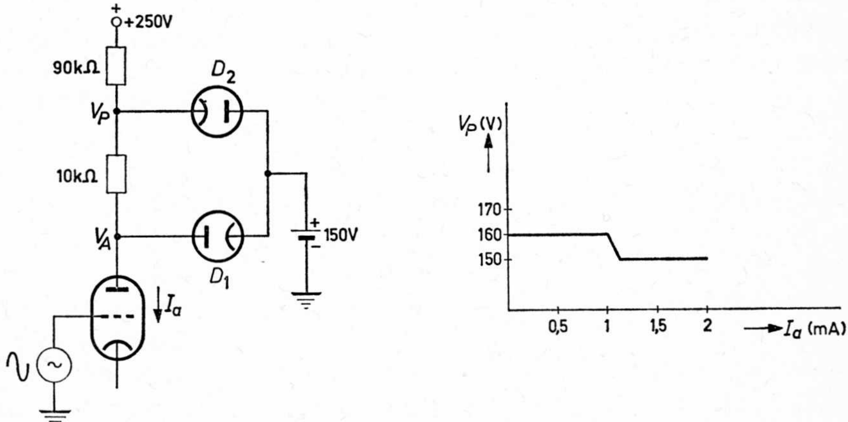

Fig. 37-12

For the limitation of periodic signals, the battery in Fig. 37-12 can be replaced by a sufficiently large capacitor (Fig. 37-13). The d.c. voltage across the capacitor will then acquire such a value that the quantities of supplied and removed charge will be equal, whilst the distance between the limiting levels is determined by the d.c. voltage difference between points a and P. As long as the valve amplification remains linear, the mean valve current will not change by external grid drive. Since all of this current passes through resistor R_1, the same applies to the mean voltage level of P. The result is that, at sufficiently large overdrive, the limiting action of the signal voltage at this point will be such that the areas of the cut-off parts above and below will be equal. Because the anode voltage still changes slightly when diode D_2 conducts, the limitation of V_a will be less symmetrical than that of V_P.

It is obvious that this same circuit can also be used for asymmetrical waveforms.

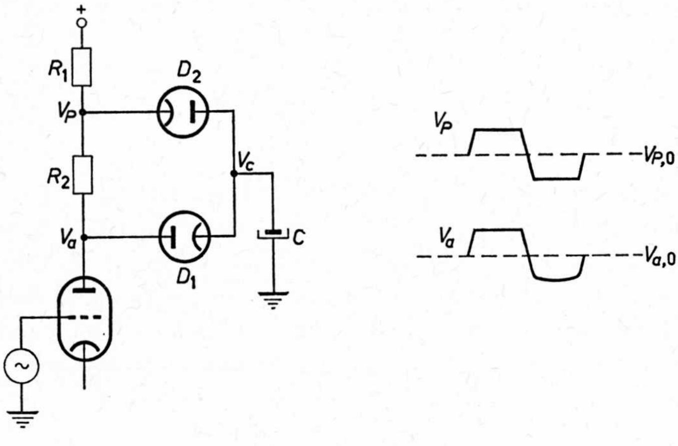

Fig. 37-13

Fig. 37-14 shows an interesting variation of this circuit. By making the capacitors sufficiently large the voltages at the points P_1 and P_2 are sufficiently smoothed. In the absence of a signal, and assuming

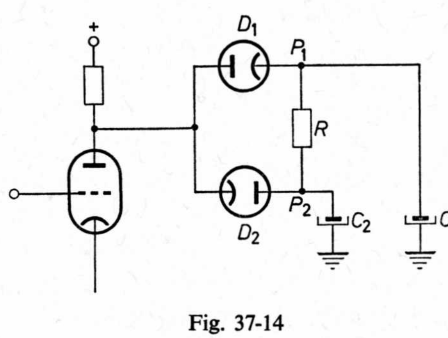

Fig. 37-14

ideal diodes, the voltages at the anode and these points would be the same. However, if the valve current increases the anode voltage will drop. This is counteracted by diode D_2, which supplies the additional current from C_2 and thus the voltage of point P_2 must drop.

The voltage of P_1 will increase in a similar fashion during the other half cycle. In the equilibrium state, the voltages at points P_1 and P_2 will be so far apart that the d.c. current through R will be the same as the mean current through diodes D_1 and D_2, which is proportional to the input signal. The difference in voltage between P_1 and P_2 and the amplitude of the anode signal (which will be almost square with correct design), will thus be proportional to the input signal. This circuit therefore also cuts off, but is not suitable for stabilizing the amplitude of an oscillator.

The operation of the balanced stage as a limiter can be improved by using two diodes as indicated in Fig. 37-15: the diodes will not conduct as long as the potential difference between the anodes is smaller than the drop in voltage across resistor R and the stage will then amplify normally. But if, for example, the left-hand valve carries so much more current than the right-hand one that V_{P_1} has become equal to V_{a_2}, diode D_1 will conduct. This situation is shown above right in Fig. 37-15. Under further overdrive, voltages V_{P_1}, V_{a_2} and V_{P_2} will no longer change, and V_{a_1} will slowly drop to value $V_{a_2}-I_kR$. When using the very steep potential difference $V_{P_1}-V_{P_2}$ as output (lower part of Fig. 37-15), this signal will always be a good square wave for a sufficiently large sinusoidal input overdrive; its amplitude will be approximately $\frac{1}{2}I_kR$, so that the amplification of the fundamental component will then be inversely proportional to the input.

In almost all practical oscillatory circuits, the loop gain can be imagined to be split into a linear frequency-selective amplification and a signal limiting action which, as shown by the examples given, may be assumed to be frequency-independent, certainly in the immediate region of the oscillation frequency. Ak can then be written as the product of an amplitude-dependent and a frequency-dependent part:

$$Ak = F(v) \cdot G(\mathrm{j}\omega)$$

where v = amplitude of the oscillatory signal. The oscillation criteria are then:

$$F(v) \cdot \mathrm{Re}\left\{G(\mathrm{j}\omega)\right\} = -1 \quad \text{and} \quad \mathrm{Im}\left\{G(\mathrm{j}\omega)\right\} = 0$$

The oscillation frequency follows from the latter equation and, when substituted in the former, gives the signal amplitude.

An oscillator's quality is primarily determined by the amplitude and frequency stability. The limiting circuits using the balanced stage and NTC-re sistor are also illustrative for the amplitude stabilization. Figs 37-2 and -3 show

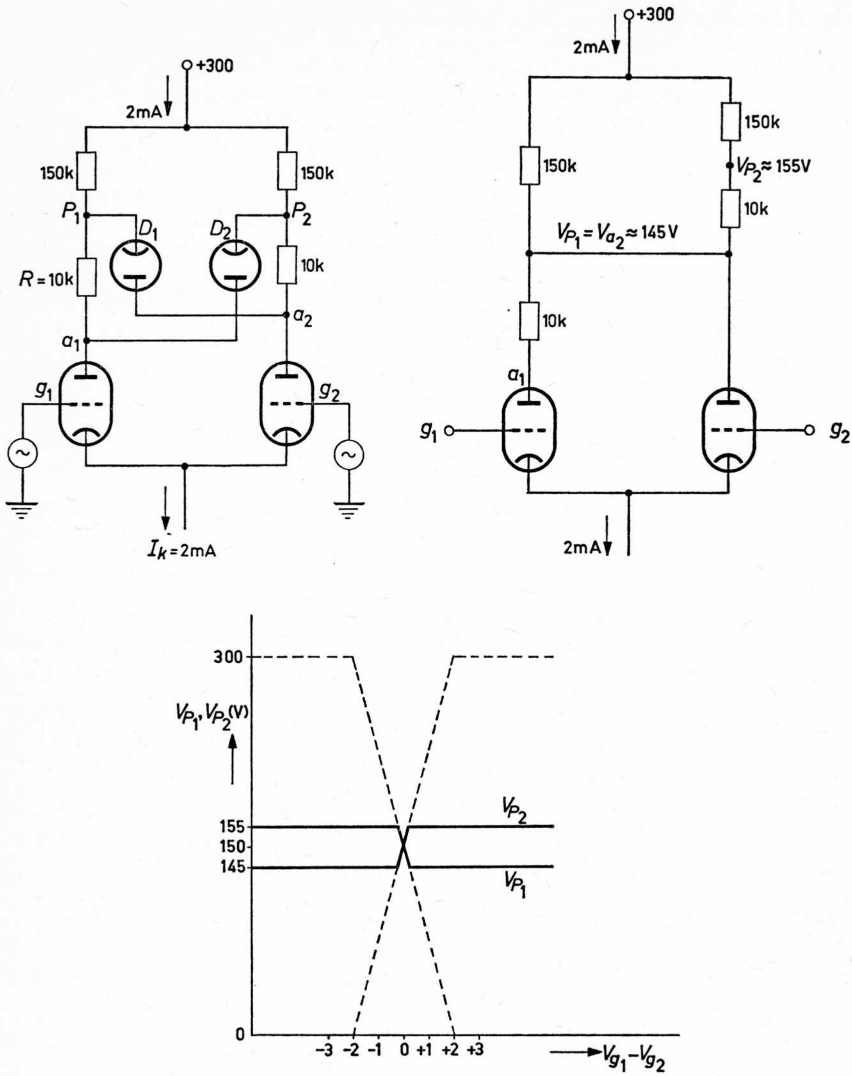

Fig. 37-15

how the first method, for a sufficiently large input signal, gives a squared output signal, the height of which is solely determined by the supply voltages used and the ratio of anode to cathode resistance. On obtaining the fundamental sinusoidal waveform from this square-wave voltage by filtering, its amplitude stability will be determined by possible changes in these filters, supply voltages and resistor values of the limiting circuit.

Fig. 37-16 shows a practical design of an oscillator using a difference stage as limiter. The two active filters (B_2–B_3 and B_4–B_5–B_6) belong to the type discussed in Section 33. A twin-T filter (F_1 and F_2) is used in the feedback loop, so that a signal of a frequency equal to the central frequency of these filters will be passed, theoretically not attenuated. The filtered signal is then fed back to the input of limiter B_1, completing the oscillator's circuit loop. The distortion present in the signal after the second filter is so small that it is no longer allowed to use a simple cathode follower, but the more extensive circuit B_5–B_6 is required, which introduces very little distortion. The amplitude stability of the output voltage v_u is mainly determined (assuming well-stabilized supply voltages) by a possible difference in tuning of the filters, which causes a slight attenuation with possible fluctuations in the oscillation signal.

The output voltage for an 80 c/s design was approx. 10 V_{rms}, and the changes in voltage were less than 0.01 per cent over a period of hours. The distortion was found to be: $d_2 = 2.5 \cdot 10^{-5}$, $d_3 = 1.5 \cdot 10^{-5}$ and for the total of the remaining harmonics less than $0.4 \cdot 10^{-5}$. Frequency fluctuations were of the order of 10^{-4}.

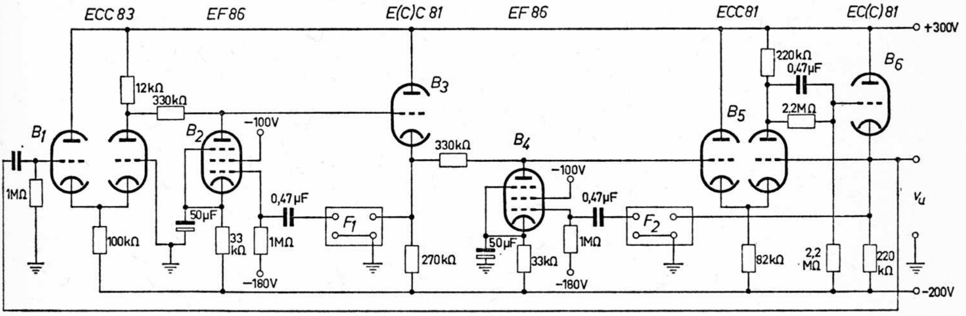

Fig. 37-16

When using an NTC-resistor, as shown in Fig. 37-4, the voltage across it is sinusoidal but not independent of the resistance value. As the purpose of adjusting this value is to maintain the loop gain at 1, possible changes in the open-circuit amplification will also be expressed in the voltage. By inserting a normal resistor in series with the NTC-resistor, one can obtain the effect that the voltage across the combination in a specific region is almost independent of the value of the NTC-resistor. The Wien bridge oscillator is an example which will be discussed later. With complete compensation, the amplitude is still determined by the thermal equilibrium of the NTC-resistor, i.e. by the ambient temperature. Depending on the precautions taken, an amplitude stability between a few per cent and approx. 0.1 per cent can be obtained with this method.

In contrast to amplitude stability, it is possible to discuss frequency stability in a more general way. It follows from the condition $\text{Im}\{G(j\omega)\}=0$ that, apart from the usually negligible fluctuations due to noise, the frequency would be perfectly constant if the components in the frequency dependent network would not show any changes. However, these do occur, due to such causes as temperature, shock, ageing or replacement of valves or transistors. The oscillation frequency will therefore also show changes. If a good frequency stability is required, we must therefore attempt to ensure that the frequency is mainly determined by the most stable components in this respect, and not appreciably affected by components which are subjected to larger changes.

The circuit of Fig. 37-17 shows how this can be achieved. This is the basic circuit of the phase-shift oscillator we shall discuss later. The *RC*-circuit, without a coil in the anode circuit, forms a capacitive load on the valve, so

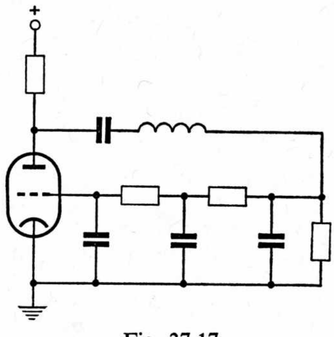

Fig. 37-17

that when the internal resistance of the valve changes, the phase of the loop gain will also change and hence the frequency. This dependence can be eliminated by inserting a coil of such a value that the anode impedance at the oscillation frequency becomes real.

We must also ensure that any change in phase only results in a slight change in frequency. This is the case when the components which mainly determine the frequency, only need a small frequency change to correct this phase change. Most oscillators therefore incorporate a frequency-selective network consisting of very high stability components, where the phase is very highly dependent on the frequency of oscillation.

Expressed as equations, we can say that for a frequency change $\Delta\omega$, the change in the imaginary part of $1+Ak=1+F(v)G(\mathrm{j}\omega)$ equals:

$$F(v)\left[\frac{\mathrm{d}\,\mathrm{Im}\{G(\mathrm{j}\omega)\}}{\mathrm{d}\omega}\right]_{\omega_0}\Delta\omega$$

With $F(v)\cdot[\mathrm{Re}\{G(\mathrm{j}\omega)\}]_{\omega_0}=1$, the coefficient of $\Delta\omega$ will be:

$$\frac{\left[\dfrac{\mathrm{d}\,\mathrm{Im}\{G(\mathrm{j}\omega)\}}{\mathrm{d}\omega}\right]_{\omega_0}}{\left[\mathrm{Re}\{G(\mathrm{j}\omega)\}\right]_{\omega_0}}$$

and since $[\mathrm{Im}\{G(\mathrm{j}\omega)\}]_{\omega_0}=0$, this quotient equals

$$\left[\frac{\mathrm{d}\arg\{G(\mathrm{j}\omega)\}}{\mathrm{d}\omega}\right]_{\omega_0}$$

Variations in frequency, caused by some disturbance, will therefore be smaller when the argument of $G(\mathrm{j}\omega)$ changes more rapidly with ω.

If the frequency-selective network consists of a parallel resonant circuit, we can write:

$$G(\mathrm{j}\omega)=\frac{1}{1+\mathrm{j}Q\beta}$$

The argument equals $-Q\beta=-2Q\Delta\omega/\omega_0$, so that the change in ω has the value $2Q/\omega_0$, which is proportional to the quality factor of the resonant circuit.

Some components combine an inherent high stability with a very high value of d(arg)/dω. Suitably cut quartz crystals with electrodes mounted on either side are an example. By inserting such a crystal instead of a conventional *LC*-resonant circuit in an oscillator, particularly constant frequencies will result.

The fundamental resonant frequency of these crystals can be from a few kc/s to 100 Mc/s. Quality factors of $3\cdot10^4$ and more are quite usual. Mounting the crystal in a vacuum enclosure offers values of around 10^6. The temperature coefficient is dependent upon the axis across which the mechanical resonance is developed and depends in practice on the faces of the crystal which are used as electrodes. Depending on this axis the temperature coefficient will vary from zero to approx. $5\cdot10^{-5}$ per °C over a rather wide temperature range. By placing a quartz crystal with a small temperature coefficient in a thermostatically controlled oven, it is possible to achieve a frequency stability of the order of magnitude 10^{-9}–10^{-8} over periods of hours, and 10^{-8}–10^{-7} over still longer times.

It is sometimes a requirement that the frequency of the oscillator be freely variable over a wide range, e.g. in an *LC*-circuit by adjusting a variable capacitor. Quartz crystals obviously cannot be used for this purpose. We then have to return to variable *LC*-or *RC*-circuits which give a much smaller change of phase with frequency, especially the latter. However, it follows from the above that frequency stability can be improved by reducing the real part of $G(j\omega)$. This can be achieved by using bridge circuits. Oscillators using this principle are usually called Meacham oscillators, after the prototype by Meacham in 1938. Fig. 37-18 shows its frequency-determining network. Transformers can be avoided by using a difference amplifier, as shown in Fig. 37-19.

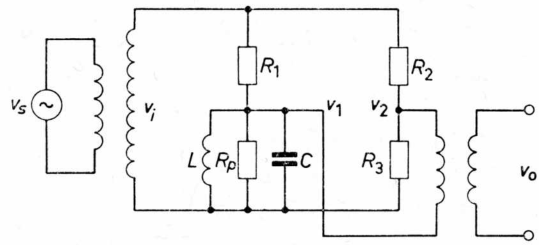

Fig. 37-18

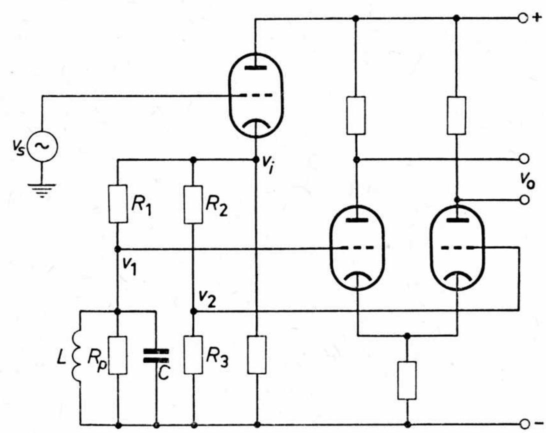

Fig. 37-19

We can write for the impedance of the circuit:

$$Z = \frac{R_p}{1 + jQ_0\beta}$$

where $$\beta = \frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}, \quad \omega_0 = \frac{1}{\sqrt{LC}} \text{ and } Q_0 = \frac{R_p}{\omega_0 L}$$

Therefore:

$$v_1 = \frac{R_p}{R_1 + R_p + jR_1Q_0\beta} v_i \quad \text{and} \quad v_2 = \frac{R_3}{R_2 + R_3} v_i$$

In Fig. 37-20, v_1 and v_2 are shown in the complex plane. This shows that the phase of $v_1 - v_2$ changes much more strongly than that of v_1, the more as v_2 is closer to $v_1(\beta = 0)$. By making R_2 an NTC-resistor and making the loop gain large, we can force v_2 very close to $v_1(\beta = 0)$. We thus achieve a very good frequency stability, even when the selectivity of $v_1(\beta)$, and hence of the resonant network, is small.

An example of an *RC*-circuit which can replace the bridge circuit of Fig. 37-18 is the Wien bridge dating from about 1890 (Fig. 37-21). With equal resistors and capacitors in the frequency-dependent left-hand branch, we find:

$$v_1 = \frac{v_0}{3 + j\beta} \quad \text{and} \quad \omega_0 = \frac{1}{RC}$$

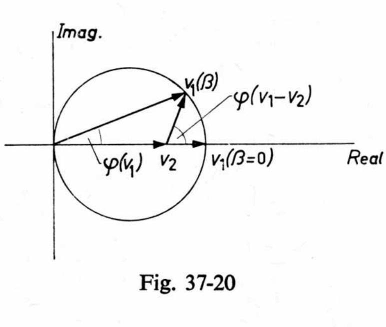

Fig. 37-20

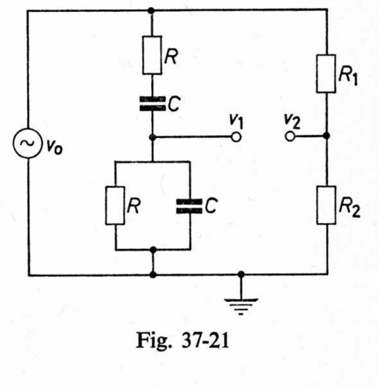

Fig. 37-21

This makes the effective quality factor $\frac{1}{3}$. If we now make R_1 an NTC-resistor, its value will adjust itself to approx. $2R_2$ when the loop gain is sufficiently large. This not only gives a high effective quality factor of the bridge voltage $v_1 - v_2$, but the amplitude of the signal will also be constant.

Fig. 37-22 shows the circuit of a Wien bridge variable oscillator for very low frequencies. The lowest frequency reached is about 0.1 c/s for the capacitor values indicated.

In order to avoid less constant phase-shift in the amplifier, the latter

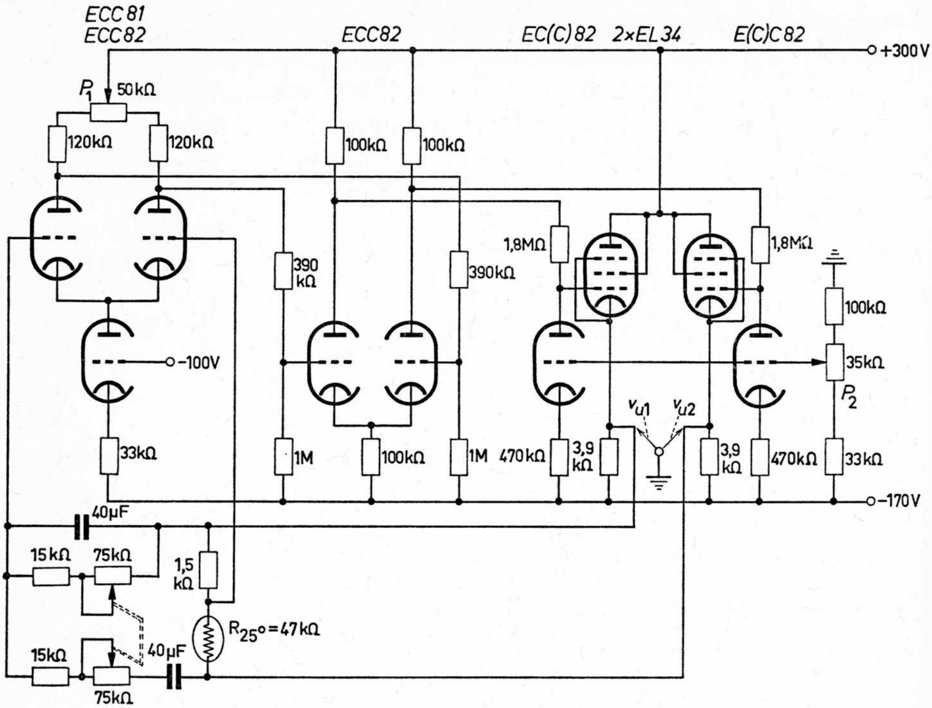

Fig. 37-22

is direct coupled. A simple calculation shows that a phase change of $\Delta\varphi$ in the amplifier produces a relative frequency change of about $\Delta\varphi/5A$, where A is the amplification of the amplifier without feedback. The zero levels of both input and output terminals are adjusted by means of potentiometers P_1 and P_2. The NTC-resistor must have a large thermal time constant so that its resistance value does not follow the instantaneous signal amplitude, even at the lowest frequency. This value is about 3 kΩ in the equilibrium state. The differential resistance is then negative and differs little from 1.5 kΩ so that the voltage across the NTC-resistor together with the fixed 1.5 kΩ resistor is less sensitive to changes.

When using moderately stabilized supply voltages, a frequency and amplitude stability of approx. 0.1 per cent can be obtained with this type of oscillator.

Apart from amplitude and frequency stability, the distortion of the output signal is also an important measure of an oscillator's quality. With "linear" oscillators having an integrating limiter we obtain exclusively, apart from a possible amplitude modulation caused by the inertia of this adjustment, distortion due to the non-linear characteristics of the active elements. For example assuming that the loop gain in the circuit of Fig. 37-1 has been adjusted in this manner, the distortion of the anode current will be determined by the value of the grid signal, and under conditions of normal drive this amounts to no more than a few per cent. However, harmonics are strongly reduced in the anode and grid voltages: $Q_0\beta$ times which approximately equals nQ_0 for the n-th harmonic. When good resonant circuits are used, this distortion thus need not exceed 0.01 per cent.

If we now satisfy the oscillation condition in this circuit by applying grid current detection (Fig. 37-7), the anode current will consist of small peaks, in which the harmonics are as strongly represented as the fundamental. In this case, distortion is solely determined by the quality of the resonant circuit. The total distortion in the output signal is then approximately a fraction $1/Q_0$ of the oscillation signal. In the circuit of fig. 37-16 instantaneous limiting with a balanced stage has been applied; the distortion of the output signal will be determined by the extent to which the two filters reduce the distortion in the squared waveform. As mentioned, a total distortion of approx. 10^{-5} can be achieved in this way.

We should finally note that most oscillator circuits only allow relatively small loads, a separate output amplifier is often needed if more power is required. We can reduce the effect of the distortion which could occur in this case by means of feedback. We thus actually obtain a stabilized power supply, with an a.c. instead of a d.c. voltage as reference. In the case of non-linear loads, we should always use such an isolation stage or circuit.

Fig. 37-23 shows the circuit of this type of output amplifier, capable of supplying an output voltage of about 50 V_{rms}. Part of the output voltage v_o is compared to the reference voltage v_{ref} of approx. 10 V_{rms}. Calculations have shown that the difference amplifier which must be used for this purpose, ought to have a rejection factor of at least 10^4 in order to limit its contribution to distortion to 0.01 per cent. The final stage consists of a single-ended push-pull circuit with an output impedance of about 1000 Ω. This is reduced to about 0.5 Ω by feedback. The adjustment of this stage is automatically maintained by means of d.c. voltage feedback. When using the output signal of the circuit of Fig. 37-16 as reference voltage, we can thus obtain an l.f. oscillator of the following specification:
output voltage 40–60 V;
maximum power: 2 watts;
distortion in the output voltage by a non-linear load, which takes a distortion current of 8 mA, i.e. 20 per cent of the maximum continuous current, is smaller than 0.01 per cent;
maximum permissible current 120 mA;
variation of the output voltage over a period of hours < 0.03 per cent;
internal resistance $< 0.7\ \Omega$.

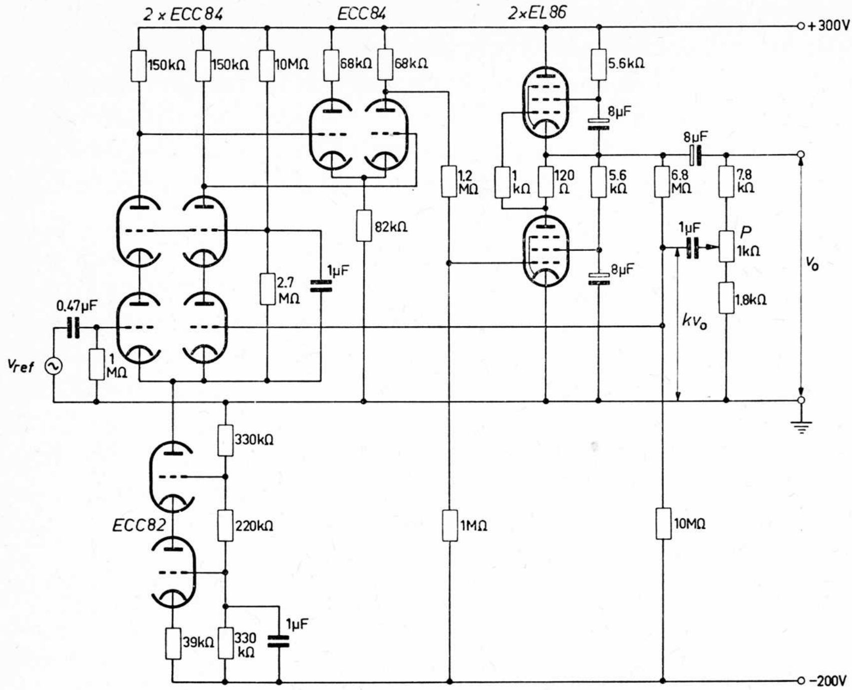

Fig. 37-23

The simplest h.f. oscillator circuits such as used in radio technique, consist of one valve or transistor and usually one and sometimes two resonant circuits. The anode and grid circuits are coupled together, while the cathode is earthed so that we have the general group of basic circuits shown in Fig. 37-24. A simple calculation shows the requirements which Z_1, Z_2 and Z_3 must satisfy to ensure oscillation in this type of circuit. Neglecting the effect of the anode voltage on the valve current, we can write, with $i = Sv_g$:

$$v_g = \frac{Z_3}{Z_2 + Z_3} v_a = - i \frac{Z_3}{Z_2 + Z_3} \cdot \frac{Z_1(Z_2 + Z_3)}{Z_1 + Z_2 + Z_3} = \frac{-Z_1 Z_3 S}{Z_1 + Z_2 + Z_3} v_g$$

so that:

$$\frac{Z_1 Z_3 S}{Z_1 + Z_2 + Z_3} = -1$$

This implies that the argument of $Z_1 Z_3/(Z_1 + Z_2 + Z_3)$ must be 180°. Restricting ourselves to impedances consisting of passive elements which,

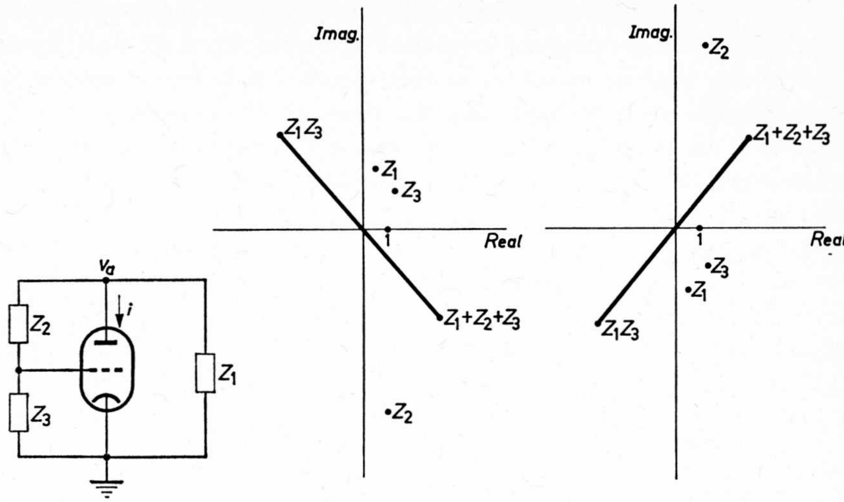

Fig. 37-24

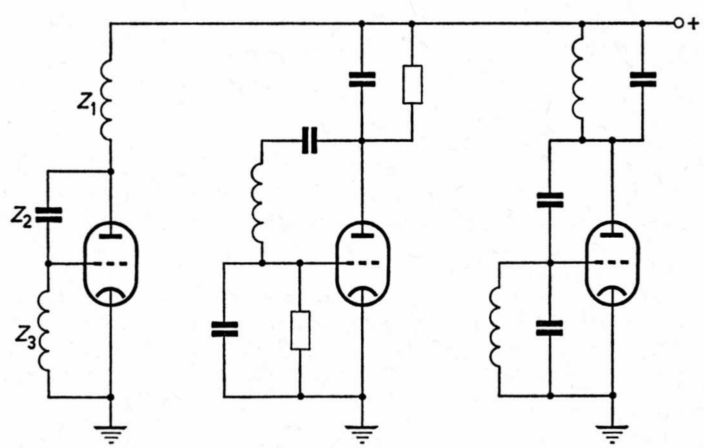

Fig. 37-25

due to the positive energy consumption, have a positive real part, only two combinations are possible: Z_1 and Z_3 in the first quadrant of the complex plane (inductive) and Z_2 in the fourth quadrant (capacitive) (centre of Fig. 37-24), or, the other way round, Z_1 and Z_3 capacitive and Z_2 inductive (right-hand side of Fig. 37-24).

Fig. 37-25 shows three possible basic circuits, which are known in the literature as Hartley, Colpitts and tuned anode – tuned grid oscillators respectively. The feedback capacitor of the Colpitts oscillator must be made large enough for Z_2 to remain inductive. The resonant circuits in the right-hand circuit must be inductive, so that the oscillation frequency will be lower than the resonant frequencies of the circuits. The term tuned anode – tuned grid oscillator is misleading in so far one might conclude that both resonant circuits should be tuned to the oscillation frequency.

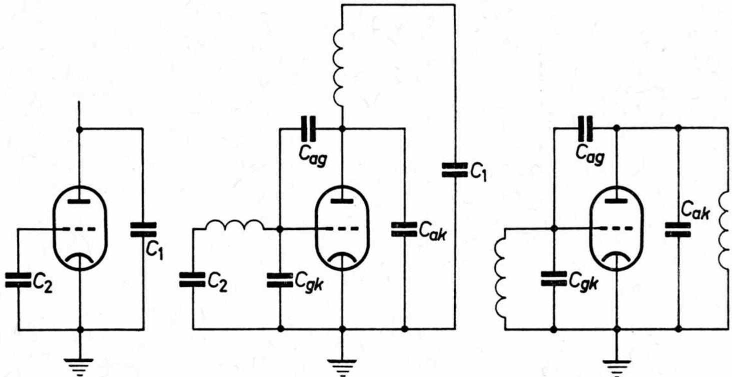

Fig. 37-26

From the above follows how it is possible for an amplifier stage where anode and grid are decoupled by large capacitances (Fig. 37-26) to oscillate at high frequencies. The inductances of cables and even those of internal connections cannot be neglected at very high frequencies. This results in the situation shown in the centre of Fig. 37-26. C_1 and C_2 are effectively short-circuits at these high frequencies, so that the circuit is now reduced to that on the right-hand side of Fig. 37-26, i.e. the tuned anode – tuned grid circuit. Oscillation can be prevented by making the grid circuit have a poor quality factor, and this is accomplished by inserting a resistor of 100–1000 Ω ("grid-stopper") in the grid lead.

"Parasitic" oscillation can quite easily escape our notice, particularly in low frequency circuits where the measuring equipment used is not intended to show h.f. phenomena. However, the d.c. levels are usually seriously disturbed by h.f. oscillation. Thus if the d.c. potentials in a circuit deviate considerably from the theoretical values, and this cannot be explained by accidental short- or open-circuit or incorrect component values, it is a good rule to check the possibility of h.f. oscillation.

LC-oscillators are difficult to implement at low frequencies because of the high values of inductance; in this case the use of selective *RC*-circuits is to be preferred.

We have seen that their inherent low Q can be increased by using a bridge circuit. We can then obtain similar circuits to those of the Wien bridge oscillator of Fig. 37-22. The use of an integrating limiter in these oscillators, combined with a relatively large time constant, allows one to keep the distortion extremely small. This "linear" operation becomes essential when a bridge circuit is not used and a reasonably sinusoidal waveform is required, because distortion is hardly reduced by the selective network. Phase shift oscillators (Fig. 37-27) belong to this class of simple RC-oscillators. If the resistors of the RC-circuit are very much higher than the anode resistor R_a, the grid-anode amplification will be real and negative, so that the loop gain will be real and positive for the frequency at which the RC-circuit gives a phase shift of 180° in the transfer from anode back to grid. If the amplification of the valve more than compensates for the attenuation of the filter, the circuit will oscillate at this frequency. Limiting can be effected by using an NTC-resistor for all or part of the anode resistance.

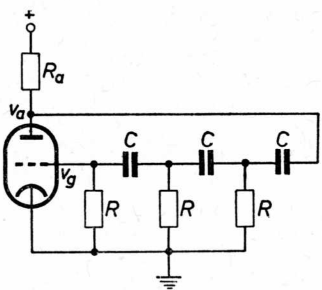

Fig. 37-27

Because of the loading by the RC-circuit, the amplification from grid to anode will not be real, and the phase shift in the circuit will be correspondingly changed. Both this load and that of the RC-filter can be minimized by isolation with cathode followers or amplification stages.

Apart from the chosen example, we can also use other RC-circuits consisting of three or more sections. The required amplification as well as the attenuation of the harmonic components may differ quite considerably in these cases.

We can illustrate this by calculating the transfer of the circuit used in Fig. 37-27. Putting $X = 1/\mathrm{j}\omega C$ we can write:

$$v_g = \frac{R^3}{R^3 + 6R^2X + 5RX^2 + X^3}\, v_a = \frac{v_a}{\left(1 - \dfrac{5}{\omega^2\tau^2}\right) - \mathrm{j}\left(\dfrac{6}{\omega\tau} - \dfrac{1}{\omega^3\tau^3}\right)}$$

where $\tau = RC$.

Therefore ω_0 follows from $$6 - \frac{1}{\omega_0{}^2\tau^2} = 0 \text{ or } \omega_0 = \frac{1}{\tau\sqrt{6}}.$$

The attenuation at this frequency has the value $1-5/\omega_0{}^2\tau^2 = -29$. When $R_a \ll R$, the valve must amplify at least 29 times.

If the sections did not load each other, each would have to give a phase-shift of 60°, which will be the case at a frequency $\omega_0 = 1/\tau\sqrt{3}$. The reduction per section is then 2, and in total 8 times.

With a practical design (Fig. 37-28), the sections load each other only very little, and an amplification of approx. 12 is necessary. Using a miniature NTC-resistor of 100 kΩ at room temperature, the output voltage was about 8 volts, and the second and third harmonic contents were both approx. 0.03 per cent. The total content of the remaining harmonics was less than $5 \cdot 10^{-5}$.

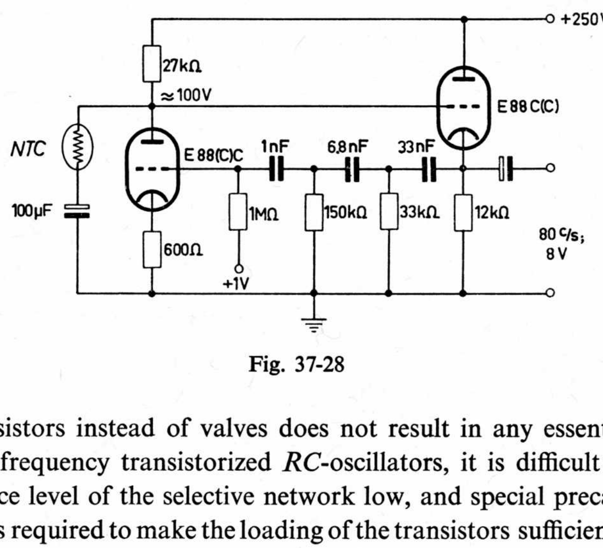

Fig. 37-28

The use of transistors instead of valves does not result in any essential change. With low-frequency transistorized *RC*-oscillators, it is difficult to make the impedance level of the selective network low, and special precautions are sometimes required to make the loading of the transistors sufficiently small.

With the oscillators discussed in this section, the sinusoidal voltages were produced as the "natural" solutions of systems, which satisfy to a very good approximation the second-order differential equation:

$$x + \frac{\mathrm{d}^2x}{\mathrm{d}t^2} = 0$$

It is, however, also possible to derive sinusoidal signals more or less accurately from other signal forms. For example, we shall discuss in Section 42 a circuit with which a sinusoidal voltage can be derived from a symmetrical triangular waveform. With such an artificial method, we shall be able to avoid the transient phenomena. The latter generally occur with ordinary oscillators: an impulse disturbance gives rise to natural vibrations which need many cycles to damp out because of the good quality of the frequency-determining network. This can be a troublesome property, particularly at low frequencies. The beat-frequency oscillator is an in-between in this respect. Use is made here of the fact that the product of two sinusoidal voltages with frequencies f_1 and f_2 is the sum of two sinusoidal voltages with frequencies $f_1 - f_2$ and $f_1 + f_2$ respectively. By using a low-pass filter we can ensure that only the signal with the difference frequency remains. The usual practice is to keep one of the frequencies constant and to vary the other one. It is thus theoretically possible to obtain signals of very low frequencies. A disadvantage is that a relatively small change in either f_1 or f_2 produces a large relative change in the difference frequency. We are therefore forced to resort to one of the more artificial solutions, as will be discussed in Section 42.

38. Stability criteria

When discussing oscillation we assumed that an oscillator would "build-up" if for small signals the modulus of the loop gain Ak is larger than 1 and the phase has the correct sign, and then, due to some kind of limiting action, that it would finally oscillate with a constant amplitude at frequency ω_0. For ω_0 is valid: $1+A(j\omega_0)k(j\omega_0)\approx 0$. We shall discuss in this section the behaviour of feedback systems. We shall consider the phenomena which occur when oscillators "build-up" and also how the stability of an amplifier can be judged.

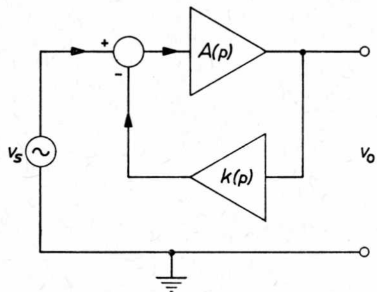

Fig. 38-1

Let us once again consider the feedback system of Fig. 38-1, where A and k are, in general, quotients of polynomials in the operator $p := \mathrm{d}/\mathrm{d}t$

$$A(p) = \frac{P(p)}{Q(p)} \quad \text{and} \quad k(p) = \frac{R(p)}{S(p)}$$

so that we can write for the amplification:

$$\frac{v_o}{v_s} = \frac{A(p)}{1 + A(p)k(p)} = \frac{P(p)S(p)}{Q(p)S(p) + P(p)R(p)} = \frac{T(p)}{N(p)} \tag{38.1}$$

The build-up of an oscillator has its origin in the fact that in the linear range (i.e. with small signals) modes with increasing amplitude can be produced. This means that such a solution exists for v_0 in the differential equation which corresponds to (38.1):

$$N\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) v_o = T\left(\frac{\mathrm{d}}{\mathrm{d}t}\right) v_s$$

for $v_s = 0$, and therefore with the right-hand term equal to zero.

The solution of the linear differential equation $N\left(\frac{\mathrm{d}}{\mathrm{d}t}\right)v_0 = 0$

is:
$$v_o = \sum_{k=1}^{n} c_k e^{p_k t}$$

where $c_1 \dots c_n$ are constants and $p_1 \dots p_n$ the roots of the n-th degree equation $N(p) = 0$.

For a mode of increasing amplitude to occur, at least one of these roots must have a positive real part. Solving an n-th order equation with complex roots is usually extremely complicated for $n \geqslant 4$. Hurwitz (1895) derived the conditions which the coefficients of a higher-power equation must satisfy so that none of the roots possesses a positive real part.

For $N(p) = a_0 + a_1 p + a_2 p^2 + \dots + a_n p^n = 0$, where $a_0 > 0$, must then apply:

$$a_1 \geqslant 0, \quad \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix} \geqslant 0, \quad \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix} \geqslant 0, \quad \begin{vmatrix} a_1 & a_0 & 0 & 0 \\ a_3 & a_2 & a_1 & a_0 \\ a_5 & a_4 & a_3 & a_2 \\ a_7 & a_6 & a_5 & a_4 \end{vmatrix} \geqslant 0$$

etc, where $a_k = 0$ for $k > n$.

Example: The equation $p^4 + 7p^3 + 19p^2 + 23p + 10 = 0$ has no roots with a positive real part:

$$a_1 = 23; \quad \begin{vmatrix} a_1 & a_0 \\ a_3 & a_2 \end{vmatrix} = \begin{vmatrix} 23 & 10 \\ 7 & 19 \end{vmatrix} = 367; \quad \begin{vmatrix} a_1 & a_0 & 0 \\ a_3 & a_2 & a_1 \\ a_5 & a_4 & a_3 \end{vmatrix} = \begin{vmatrix} 23 & 10 & 0 \\ 7 & 19 & 23 \\ 0 & 1 & 7 \end{vmatrix} = 2040.$$

Polynomials satisfying these conditions are called Hurwitz polynomials, and play an important part in circuit theory, because the numerator and denominator of an impedance always consist of such polynomials. In practice, the usefulness of applying Hurwitz criteria is rather restricted, particularly because it is impossible to indicate in the case of a stable system how far removed it is from instability.

In 1932 Nyquist derived a criterion for feedback systems which is extremely suitable in practical application, because it establishes a relation between the behaviour of the system and the easily measured loop gain Ak. This criterion has been derived by using some properties of functions of a complex variable, which we shall therefore consider first.

If $F(z)$ is a polynomial in the complex variable z, we can write:

$$F(z) = \lambda(z - z_1) \dots (z - z_n)$$

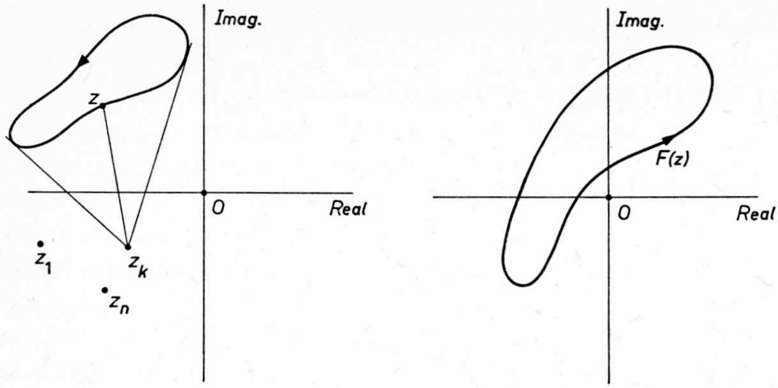

Fig. 38-2

where λ is a constant and $z_1 \ldots z_n$ the roots of $F(z)=0$. We shall restrict ourselves to the case that $F(z)$ only possesses single roots, because the presence of multiple roots does not make any difference to the conclusions.

When the locus of z is a closed curve in the complex plane which does not enclose any of the zeros $z_1 \ldots z_n$ of $F(z)$, left-hand side of Fig. 38-2, the arguments of the terms $(z-z_1) \ldots (z-z_k) \ldots (z-z_n)$ will vary, but resume their original values on the return of z to the starting point. In this case the argument of $F(z)$ will be unchanged after a closed excursion of z, which means that the origin 0 must lie outside the area enclosed by the corresponding curve $F(z)$, right-hand side of Fig. 38-2. The situation is entirely different when the z-curve encloses one or more zero-points of $F(z)$. For example, if this is the case for z_k, the argument of $(z-z_k)$ after a closed excursion of z will have increased or decreased by 2π, depending on the direction of travel (Fig. 38-3).

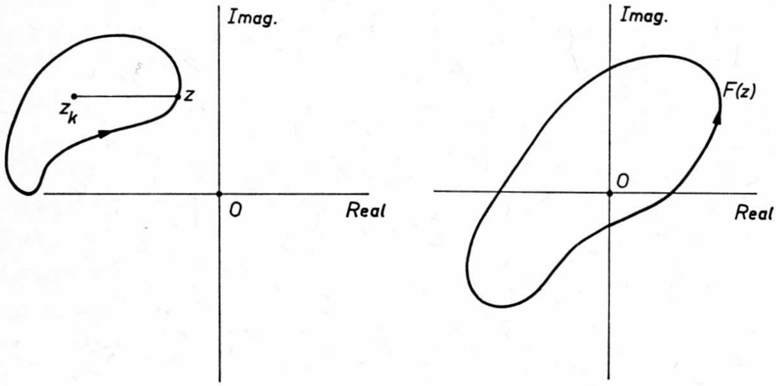

Fig. 38-3

If there are m zero-points of $F(z)$ in the area enclosed by the z-curve, and this area is a single enclosed non-intersecting self-area, the argument of $F(z)$ will be changed by $2\pi m$ after a closed excursion of $F(z)$. This implies that the $F(z)$-curve will enclose the origin m times.

We should note here that the enclosure of the origin by the $F(z)$-curve occurs in the same sense as the enclosure of the zero-points by the z-curve. This follows from the fact that the argument of $F(z)$ equals the sum of the arguments of $(z-z_1)\ldots(z-z_n)$ and that the sign of a possible change 2π in the arguments of these terms is exclusively determined by the direction of travel of z.

If we now take the right-hand half-plane for the inner area of the z-curve, the number of times that $F(z)$ encloses the origin corresponds to the number of zero-points of $F(z)$ which lie in the right-hand half-plane. In this case, we take the z-curve which is indicated in Fig. 38-4: from the origin along the imaginary axis to $+j\infty$, from $+j\infty$ along the semi-circle $|z|=\infty$ to $-j\infty$, and from there along the imaginary axis back to the origin. The right-hand half-plane lies therefore on the right of this loop, so that the inner area of the $F(z)$-curve also lies on the right of this curve.

Let us apply the above to the function $N(p)$ from (38.1) by letting $p=\alpha+j\omega$ enclose the right-hand half-plane in a clockwise sense. If the $N(p)$-curve now encloses the origin in a clockwise sense, $N(p)$ will possess roots with a positive real part, which would mean that the system will oscillate.
However, it follows from (38.1):

$$N(p) = Q(p)S(p)\left\{1 + A(p)k(p)\right\}$$

and since both A and k are each assumed to be stable circuits, the de-

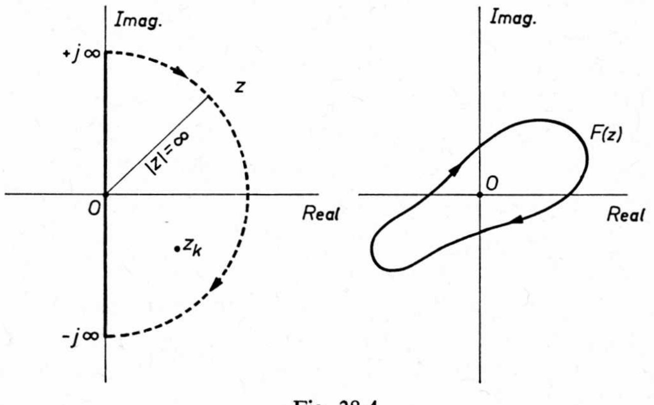

Fig. 38-4

nominators $Q(p)$ and $S(p)$ of $A(p)$ and $k(p)$ will only possess zero-points in the left-hand half-plane. The possible zero-points of $N(p)$ in the right-hand halfplane will therefore coincide with those of $1+A(p)k(p)$. We can therefore restrict ourselves to examining the behaviour of this expression. If p on the periphery of the right-hand half plane follows the part 0 to $+j\infty$ of the imaginary axis, $1+A(p)k(p)$ will be identical with $1+A(j\omega)k(j\omega)$, where $0 \leqslant \omega < \infty$.

It follows from the fact that Ak is stable, that $A(p)k(p)$ is finite for all values of p, therefore also for $|p|\rightarrow\infty$. Thus the highest power of p in the numerator of $A(p)k(p)$ cannot be higher than that in the denominator, so that $A(p)k(p)$ will in any case be real for $|p|\rightarrow\infty$, namely the ratio of the coefficients of the highest powers in numerator and denominator, if these are equal, or otherwise zero. The semi-circle $|p|=\infty$ therefore changes to a single point on the real axis.

It follows from the fact that the real part $R(\omega)$ of $A(j\omega)k(j\omega)$ only contains even powers of ω and the imaginary part $I(\omega)$ exclusively odd powers, that $A(-j\omega)k(-j\omega)=R(\omega)-jI(\omega)$, so that the function $A(-j\omega)k(-j\omega)$ for $-\infty<\omega\leqslant 0$ is the mirror-image of the function for $0\leqslant\omega<\infty$ with respect to the real axis.

The Nyquist diagram gives the path of $A(p)k(p)$ instead of $1+A(p)k(p)$, which means that the central point is not the origin but point $(-1,0)$. We can therefore state:

If in a feedback system the open loop gain Ak is stable, the closed system will oscillate if the Nyquist diagram, i.e. $A(j\omega)k(j\omega)$ for $-\infty<\omega<+\infty$, encloses the point $(-1,0)$ on its right. Figs 38-5, -6 and -7 give a few examples of Nyquist diagrams. In future we shall only indicate part $0<\omega<+\infty$.

The Nyquist diagram is easy to determine in practice, because it is simply a polar representation of the frequency characteristic of loop gain Ak. A great advantage of the Nyquist diagram is that it supplies clear indication concerning the nature of the stability of a non-oscillating system. For example, the Nyquist diagram of Fig. 38-5 shows that the system does not oscillate but that it will start oscillating when the total amplification increases slightly for one reason or another. This may happen when replacing a valve or transistor, but also if, due to the curvature of a valve characteristic, the mean slope increases when the valve is driven by relatively large signals. In the case of Fig. 38-7, we also have a stable situation, but here the system will begin to oscillate when the amplification decreases sufficiently to bring point $(-1,0)$ within the loop. This may occur when the valves have aged, reducing their slope. Furthermore, such a system has the danger that during the build-up period the amplitude of the natural modes is sufficiently

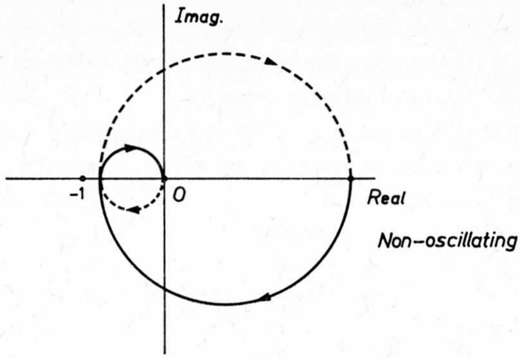

Fig. 38-5

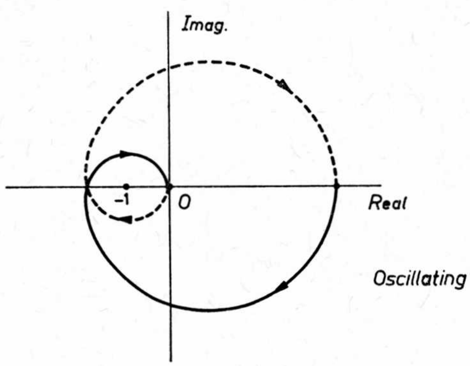

Fig. 38-6

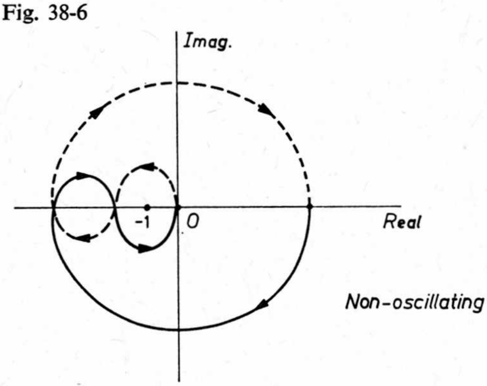

Fig. 38-7

large for the loop gain to be reduced because of limiting so that point $(-1,0)$ remains within the loop and oscillation continues.

We distinguish in this respect between the following definitions for feedback systems:

a. stable: this is the case when the system will not begin to oscillate under any circumstances which can occur under normal changes in conditions;
b. conditionally stable: when the system is normally stable, but can oscillate under specific operational circumstances;
c. unstable: any system which oscillates, intentionally or otherwise.

The following example shows that relatively small changes in a circuit can have considerable effect on its stability. The gain of the amplifier stage in Fig. 38-8 is:

$$\frac{v_o}{v_s} = \frac{A_0}{1 + j\omega RC}$$

where A_0 is the d.c. amplification and R the internal resistance of output when capacitor C is absent.

The polar diagram for this is shown on the left-hand side of Fig. 38-9.

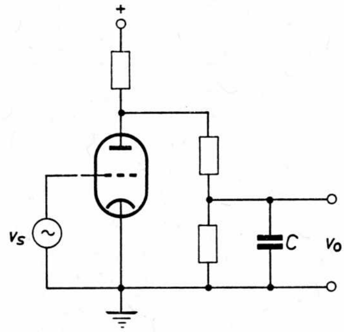

Fig. 38-8

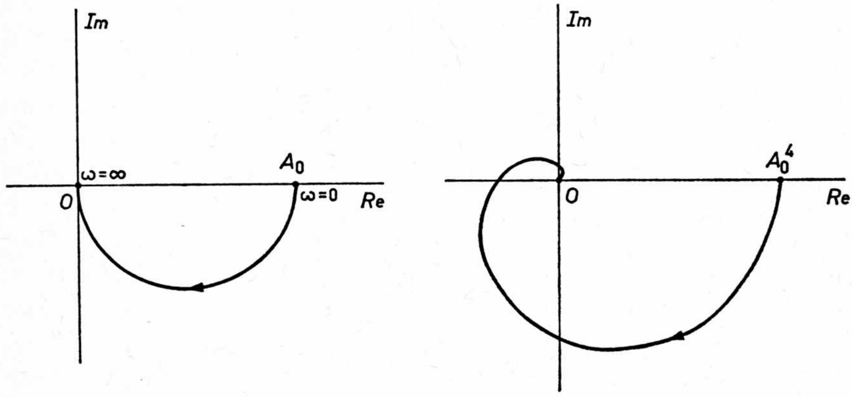

Fig. 38-9

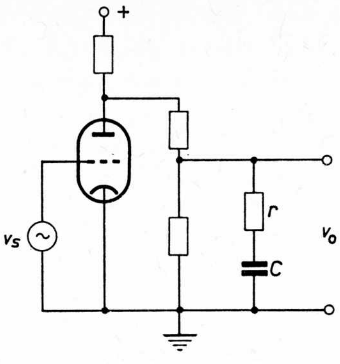

Fig. 38-10

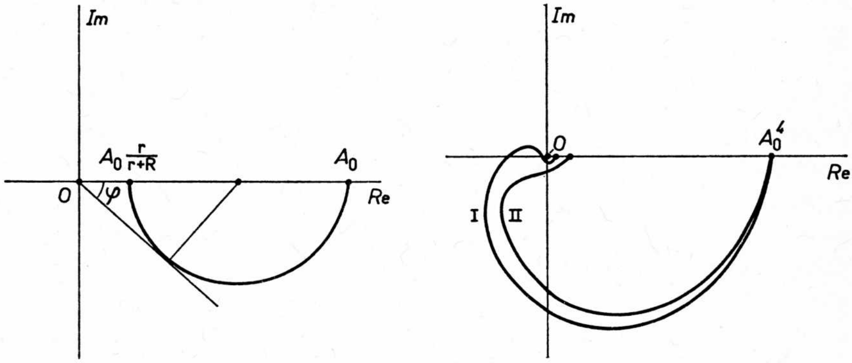

Fig. 38-11

In the case of four similar stages in cascade, the diagram of the total amplification will be as shown on the right-hand side. It follows that the system will begin to oscillate when the output is fed back to the input and A_0 is larger than a certain minimum value. The operation of the phase-shift oscillator is based on this principle.

Now taking the circuit of Fig. 38-10 instead of that of Fig. 38-8, where a resistor r, which is small compared to R, is connected in series to capacitor C, we find for the stages' amplification:

$$\frac{v_o}{v_s} = A_0 \frac{1 + j\omega\tau_1}{1 + j\omega\tau_2}$$

where $\tau_1 = rC$ and $\tau_2 = (r+R)C$.

The polar diagram of this circuit is shown in Fig. 38-11 (left-hand side). For four cascade-coupled stages a diagram which shows an entirely different form is obtained, right-hand side of Fig. 38-11. At a small value of r/R it depends on amplification A_0 whether point $(-1{,}0)$ will be enclosed (I). This is no longer possible at larger values of r/R, even with large values of A_0 (II), which shows that feedback may be applied to this type of amplifier without any risk of oscillation.

The maximum phase-shift given by a single stage follows from the diagram on the left-hand side of Fig. 38-11: $\sin \varphi = 1/(1 + 2a)$, where $a = r/R$. Oscillation can no longer occur with four stages if φ is smaller than 45°, or if a remains larger than 0.2. In the case of three sections, this becomes $\varphi < 60°$, therefore $a > 0.07$.

The open loop gain can be measured by interrupting the feedback loop (Fig. 38-12) and measuring the amplification between "input" AB and "output" $A'B'$. Care should be taken to terminate this output in the same way as when the loop is closed, i.e. impedance Z_{AB} between points A and B of the open system must be inserted between A' and B'. Dependent on the point where the loop is interrupted, Z_{AB} can be very large (input of a valve amplifier) or very small (e.g. the feedback resistor r when the interruption takes place at CD).

We must determine both the amplitude and the phase of the open loop gain. A detailed discussion of this will be found in Section 40. Another method consists of determining in the amplified signal both the value of the component in phase with the input and that of the component shifted by 90°. This, too, will be discussed in Section 40.

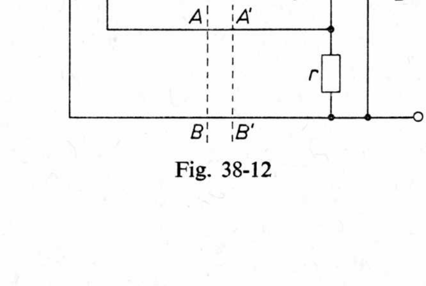

Fig. 38-12

39. Relaxation circuits

Having concluded the discussion on the various problems encountered in linear amplification, we shall now consider the various operational circuits. For these circuits as well as for signal sources, not only is there a need for generators of sinusoidal voltages but also for sources which can produce other signals which are or are not periodically repeated, such as step functions, square waves and sawtooth waveforms. Since many of these waveforms can be obtained by means of "relaxation circuits", we shall discuss these first. Because of their application in digital electronic computing techniques, an extensive literature exists on this subject. We shall therefore restrict ourselves to a discussion of the principles and of the most important aspects for use in analogue measurement circuits.

Let us first consider the difference stage of Fig. 39-1, where R_k is supposed to be very large, which contains two basically identical valves and in which the anode of the left-hand valve I is connected to the grid of the right-hand valve II by means of capacitor C. This connection provides very strong positive feedback to the circuit and there will be an unstable position when both valves pass the same current. When the current through valve I is decreased for one reason or another, e.g. noise perturbation, the anode voltage V_{a1} will rise and consequently also grid voltage V_{g2} via the capacitor. The current through valve II is therefore increased and because the total current passing through both valves remains constant, the current through valve I will further decrease. If the loop gain exceeds unity, this process will continue until all the current is passed by valve II and valve I is cut off. Grid voltage V_{g2} is then positive with respect to earth. However, this situation is not stable because voltage V_{g2} will drop exponentially to zero with a time constant $C(R_{a1}+R_g)$. As soon as this voltage has become small enough for valve I once again to pass current, the above process will be repeated, but this time in the opposite sense: V_{a1} decreases, as well as V_{g2}, and the current through valve I will increase by cathode coupling until valve II is cut off. V_{g2} is then negative with respect to earth. However, this again is not a stable situation because after V_{g2} has become sufficiently small, the right-hand valve will again conduct and the cycle will be repeated. If we neglect the effect of all stray capacitances, voltages V_{a1} and V_{g2} will therefore behave as indicated in the lower part of Fig. 39-1, where R_p is the parallel combination of R_{a1} and R_g, and $I_k \approx V_-/R_k$. This circuit is known as the cathode-coupled multivibrator and produces a periodic non-

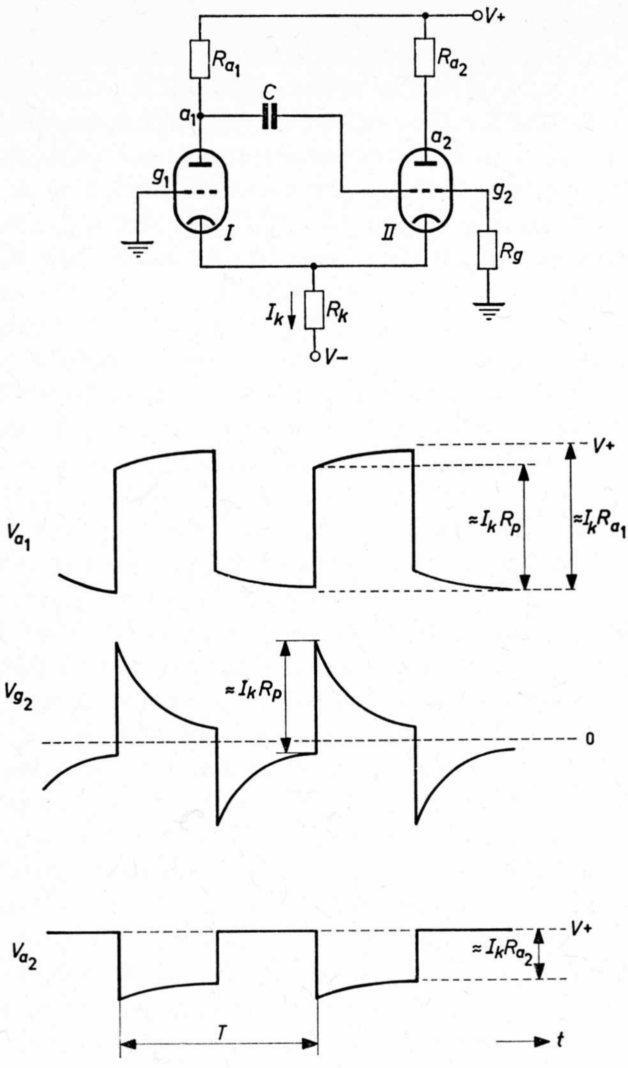

Fig. 39-1

sinusoidal output. The first half-cycle of the latter is determined by the time required for the decay of the charge of a capacitor. We should note that, theoretically, the time constants of the positive and negative portions of V_{g2} are not equal. $C(R_g + R_{a1})$ determines the time constant for the positive part and $C(R_g + R_{a1}//r_a)$ for the negative part, where r_a is the anode resistance of conducting valve I. However, in this case this resistance is very much larger than R_{a1}, because of the use of a large cathode resistance R_k. Its effect is therefore negligible.

With the circuits discussed in this section, the active components are used as switches and the capacitors are usually discharged through resistance circuits. It is relatively easy to calculate the waveforms at the various points in the circuit when only one capacitor is present, neglecting stray capacitances. We only have to remember that the ultimate potential of a circuit point will equal the potential it would acquire if there were no capacitors present, and that this situation is reached at an exponential rate with a time constant which equals the product of the relevant capacitance and the internal resistance of the circuit between the points across which the capacitor is connected. For example, let us consider the circuit of Fig. 39-2 and see what happens when current I is switched off. The voltage jumps at points a and b will be $I \cdot R_1R_2/(R_1 + R_2)$. Subsequently, the charging current of the capacitor produces a change in potential at points a and b at an exponential rate with a time constant $C(R_1 + R_2)$ and final voltages of V_+ and zero respectively. We shall later refer to the more complicated calculations which are necessary when stray capacitances are present and taken into account.

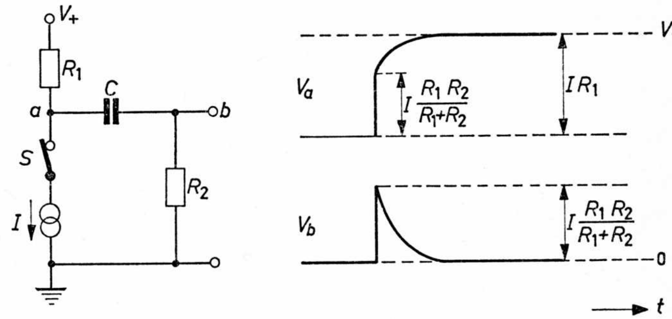

Fig. 39-2

We shall now first consider the similarities existing between such a multivibrator circuit and a conventional oscillator. We have seen that there is a difference in operation between certain types of oscillator circuits. One group had such a highly selective network that the remainder of the circuit could have a strong non-linear behaviour. The network of Fig. 37-1 is an example, where the anode current may consist of short current pulses in which the amplitude of the harmonics is approximately equal to that of the fundamental. The great selectivity of the resonant circuit removes these harmonics effectively, so that the voltage across the resonant circuit is almost purely sinusoidal. As the selectivity of the resonant circuit decreases, less distortion can be allowed in the remainder of the circuit in order to obtain a purely sinusoidal signal. In oscillators with *RC*-networks as frequency-determining elements we therefore preferably use a linear amplifier. Although the relation between amplification and amplitude is non-linear, as required for an oscillator, the amplification is controlled by a component which behaves – measured over a short time – as a linear element, namely the NTC-resistor or filament lamp.

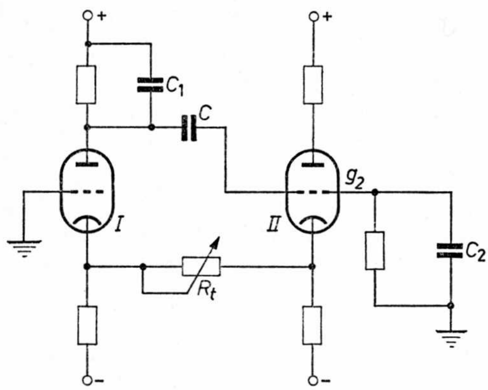

Fig. 39-3

If an active filter is used (Fig. 39-3) and we make the feedback so strong by means of R_t that oscillation occurs, an impure sinusoidal waveform results because of the relatively small selectivity of the circuit ($Q \leqslant 0.5$). Assuming we make R_t zero, the feedback is so strong that valve I opens and closes abruptly and the pulsed current will give rise to a waveform V_{g2} that has a considerable deviation from a sinusoidal form. In a square wave, the amplitude of the third harmonic is one third of the amplitude of the fundamental, whilst the selective attenuation of the filter ($\sqrt{1+Q^2\beta^2}$, where $Q \leqslant 0.5$ and $\beta = 8/3$) will not be more than 1.6. This attenuation can be made still smaller by degrading the quality of the resonant circuit. As we have seen in Section 33, we can achieve this by making coupling capacitor C much greater than C_1 and C_2. However, in this case we obtain the multivibrator circuit of Fig. 39-1.

Before considering circuit variations and improvements for the multivibrator, we shall first discuss the other two basic circuits. This will give us an idea of the possibilities offered by relaxation circuits.

If we disturb the symmetry of the circuit of Fig. 39-1, for example by making the grid voltage of valve I about 10 volts more positive than that of valve II (Fig. 39-4), the circuit will be stable, in contrast to the multivibrator. In this situation valve I carries all the current and valve II is cut off. If we now make grid g_1 sufficiently more negative for a short period, so that part of the current is transferred to valve II, the circuit will behave like a multivibrator: valve II will carry all the current and valve I will be cut off. However, this is not a stable situation, and after a certain time has passed, proportional to time constant $C(R_a + R_g)$, the circuit will return to the original stable situation. The latter will be maintained as long as no fresh negative initiating pulse is injected on the left-hand grid. This circuit with only one stable state is known as the monostable trigger, univibrator or one-shot multivibrator. It is chiefly used because of its feature of producing a pulse

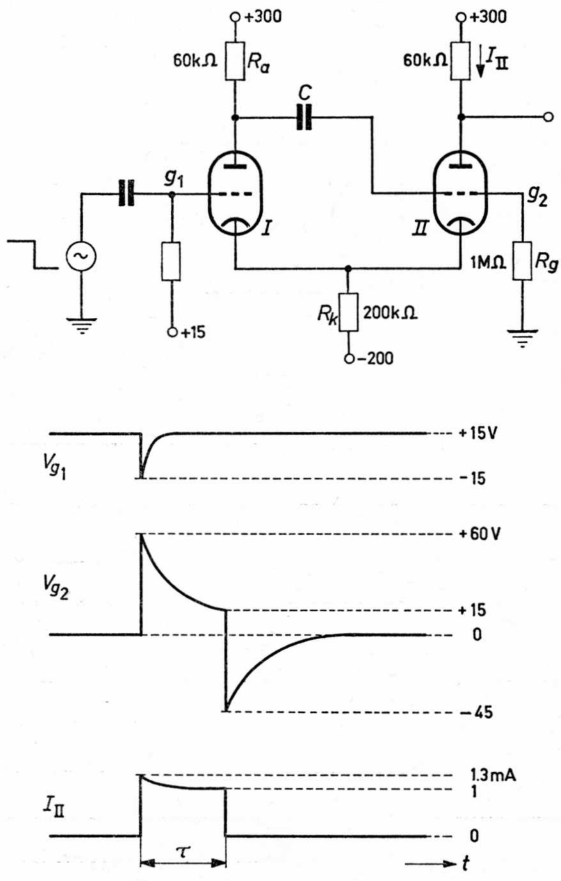

Fig. 39-4

of fixed width. By using the trailing edge of this pulse for triggering a following circuit, a delay of time constant τ is obtained. Other applications use the property that valve II passes a fixed charge per pulse. We can obtain a voltage which is proportional to the number of trigger pulses received, by feeding these constant charge pulses to a capacitor.

If we bridge capacitor C in the circuit of Fig. 39-4 with a resistor R, whose value has been selected so that the grid of valve II will just reach earth potential in the stable state when valve I is conductive (Fig. 39-5), this will not cause any change in the stable state. However, when forcing the circuit to the other state by means of a negative pulse, the voltage of the right-hand grid will no longer leak away to earth but to the voltage which is determined by the voltage division of R_a, R and R_g, and which is about 23 volts above earth. Valve I will therefore not take over the current, and the new state

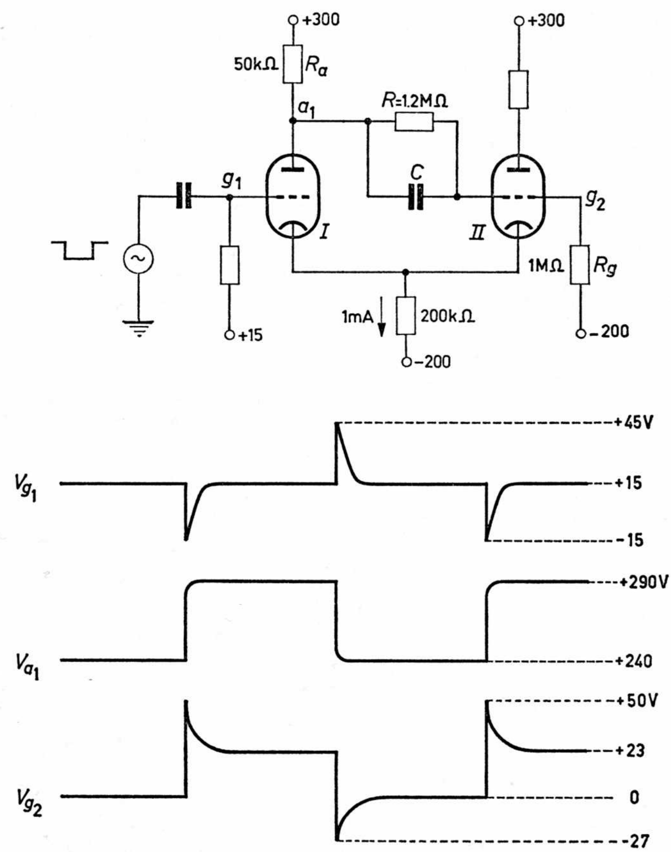

Fig. 39-5

will also be stable. If it is desirable to return to the original state, one way of achieving this is by means of a sufficiently large positive pulse on the left-hand grid. This circuit therefore features two stable states and is known as the "Schmitt trigger" or "bistable trigger".

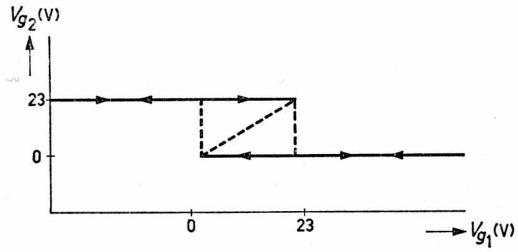

Fig. 39-6

The Z-shaped characteristic of Fig. 39-6 is typical for this type of circuit. It shows the relation between the d.c. voltage levels of two corresponding points, in Fig. 39-6 the grid voltage V_{g2} against V_{g1}. Two stable states are possible in a certain voltage range of V_{g1} (here between 0 and 23 volts). In which of the two states the circuit is at a given moment is determined by the direction from which this range is approached. True enough, there is also a state of equilibrium on the diagonal, but this is unstable. All applications of the Schmitt trigger, which we shall discuss later, are based on this hysteresis phenomenon.

Having described the operation of the three basic circuits, we shall discuss the circuit variations for each of them in greater detail, as well as their respective properties.

The original multivibrator circuit by Abraham and Bloch (1919) is symmetrical in design, and the cathodes are directly connected to earth (Fig. 39-7). The latter feature results in the first place in the fact that the voltages across both grids cannot exceed approximately earth potential. For as soon as one of the grids becomes positive with respect to earth, the current drawn by the grid will charge the capacitor. Also, since the grid-cathode diode of most valves has a resistance not exceeding a few kilo-ohms, resistors R_{g1} and R_{g2} are then effectively shorted out. The time constants for the positive parts of the grid signals will thus be considerably smaller than those of the negative parts. Moreover, because of this positive grid signal, an additional negative pulse will occur in the voltage of the corresponding anode as well as on the other grid (middle and lower part of Fig. 39-7, where the anode resistances are assumed to be small with respect to the grid resistances.) The cycle of the multivibrator is determined by the duration of the two negative parts, therefore by time constants $C_1\,(R_{g2}+R_{a1}//r_{a1})$ and $C_2(R_{g1}+R_{a2}//r_{a2})$, where r_{a1} and r_{a2} are the internalresistances of the conducting valves, as well as by the ratio of the voltages on the anodes and the grid swings of the valves. In contrast to the circuit with a large common cathode resistor, where the current through the conducting valve and hence the anode voltage waveforms is precisely defined, this is far from true here. In the conducting state, the current is determined by the valve characteristic at $V_{gk}\approx 0$, and large differences in behaviour can occur for different specimens of the same type. Since the time constants are also dependent on the internal resistance of the valves (which is not magnified

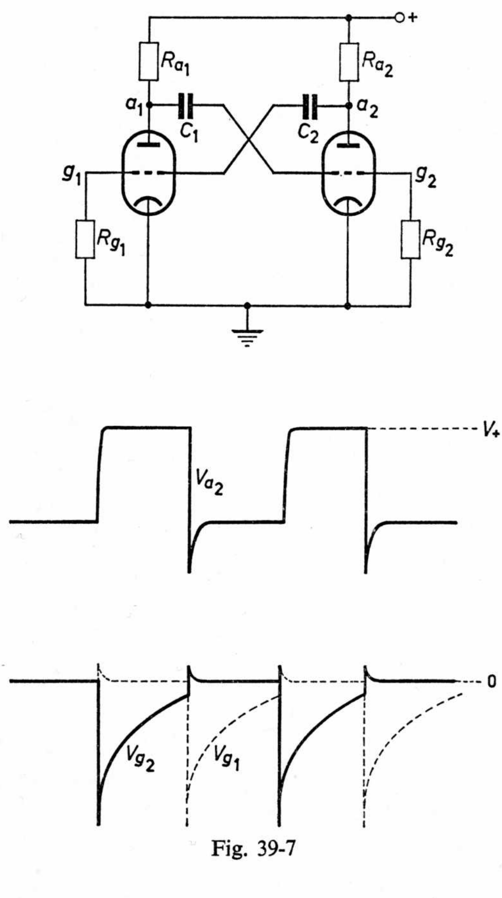

Fig. 39-7

in this case), the frequency stability of this type of multivibrator is rather poor.

Although the circuit of Fig. 39-1 has stable waveforms and stable time constants, the frequency can still show quite considerable fluctuations. This is due to the fact that the slope of grid voltage V_{g2} is not very steep at the change-over points, with the result that these points will not be firmly fixed in time. Changes in supply and heater voltages have thus a rather large effect on the frequency of the multivibrator. This can be improved by ensuring that the slope of the grid voltages at the change-over points is steeper. For example, we can achieve this with the symmetrical circuit of Fig. 39-7 by connecting

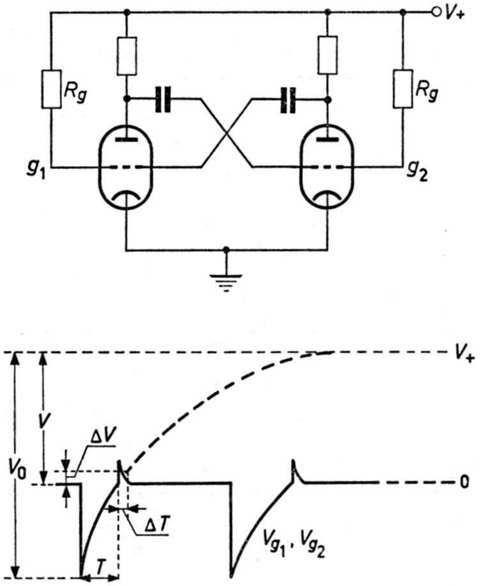

Fig. 39-8

the grid resistors to the positive supply rail, not to earth (Fig. 39-8). The resistors must then be so large that, in the conducting state of the valve, the grid current drawn is limited to a permitted value: $i_g \approx V_+/R_g$.

Since this makes the cycle shorter, the improvement is only effective for certain relations between the voltages. Using the symbols of Fig. 39-8 (lower part) the change ΔT in time T because of a change in level ΔV will be $\Delta T = \tau \Delta V/V$ where τ is the relevant RC time constant. This gives for $V_0 e^{-T/\tau} = V$: $\Delta T/T = \Delta V/V \cdot ln\,(V_0/V)$, so that for a certain change ΔV and a given V_0, the relative change in T will be inversely proportional to the product $-x \ln x$, if $x = V/V_0$. With $0 < x < 1$, this function has a maximum for $x = e^{-1} = 0.37$ with the value 0.37, but the function's value will only be reduced to half of this at $x \approx 0.07$ and $x \approx 0.8$ x should thus preferably have a value of between 0.3 and 0.5.

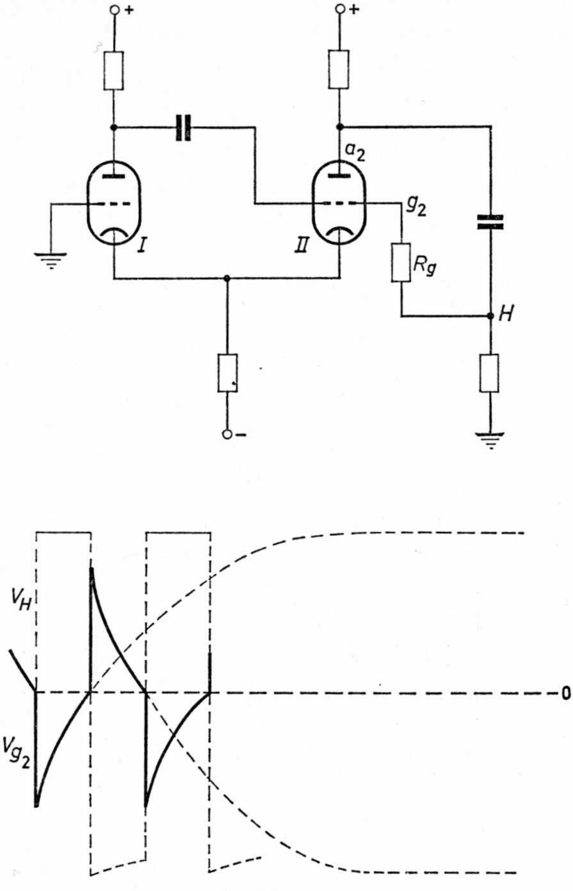

Fig. 39-9

This improvement is also possible with the asymmetrical circuit of Fig. 39-1 by making use of the voltage waveform on the right-hand anode a_2. This gives the circuit of Fig. 39-9, where grid resistor R_g is connected to a point H which is decoupled to a_2 by means of a large capacitor and thus has the same voltage changes.

A second improvement can be obtained by using the circuit of Fig. 39-10 instead of a single RC-circuit. We have shown this same circuit in a previous section as the simplest illustration of the principle of "adding what is lacking". In the case of a step function as input signal, the voltage at B will have an almost exponential waveform, but the voltage at A will remain approximately constant for quite a time, and will then drop rapidly. With a triple circuit, the ratio of the flat to the sloping portion will be still larger. The application of such networks in multivibrator circuits is therefore one way of considerably improving frequency stability.

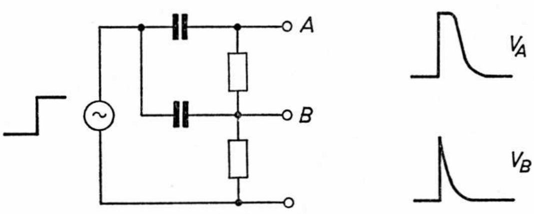

Fig. 39-10

We can also make use of the fact that the position of the change-over points in the circuits of Figs 39-1 and -7 can be easily influenced by "synchronizing" the frequency of a multivibrator with that of another periodic signal. By synchronizing is meant that the cycle of a multivibrator will be a whole number of times greater than that of the synchronization signal. This number should not usually exceed 20 in practice, with a view to stability of performance. The synchronizing signal can be applied to the left-hand grid of the circuit of Fig. 39-1, which gives us Fig. 39-11.

Assuming that the circuit of Fig. 39-11 changes over at $V_{g1} = V_{g2}$ (Fig. 39-12), this change will occur during the k-th period if V_{g2} is still greater than A at the $(k-1)$th peak and drops during the next cycle to below this value. The limiting cases then correspond to curves I and II. Although the starting points differ by a full cycle, this difference is much less for the points of intersection with V_{g1}.

For the time difference Δ between the points of intersection of I and II with the k-th period of V_{g1} applies: $e^{-T/\tau} = \cos\varphi$, where $\varphi = 2\pi\Delta/T$. With the usual small values of T/τ this gives $1 - T/\tau \approx 1 - \frac{1}{2}\varphi^2$ or $\Delta/T \approx \sqrt{2T/\tau}/2\pi$. With $T/\tau = 0.2$ this becomes $\Delta/T \approx 0.1$.

The value 1 has been given to T/τ in the figure for the sake of clarity.

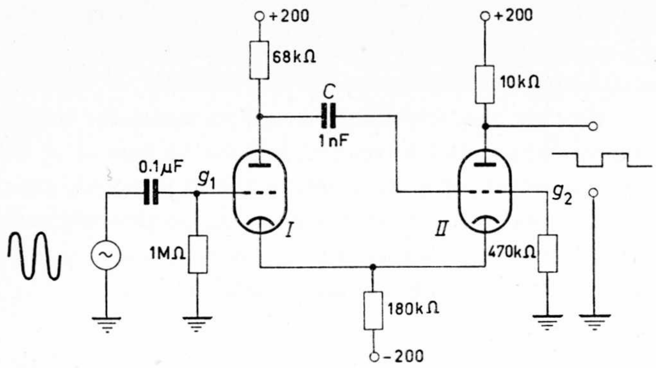

Fig. 39-11

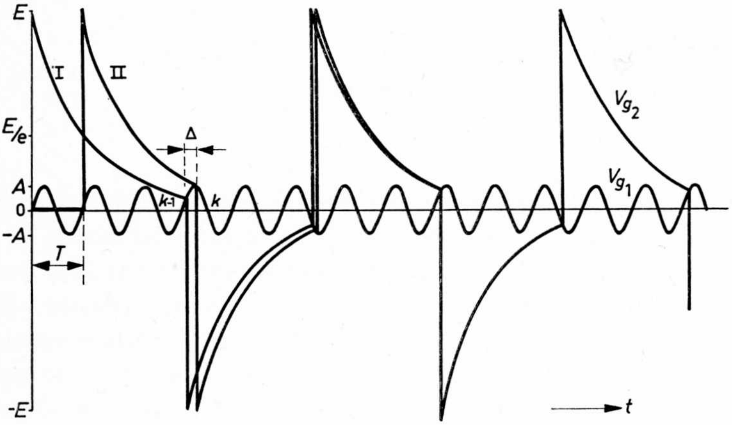

Fig. 39-12

As the same applies to negative signals, the starting points of the next cycles of I and II will be much closer together. This results, after a few cycles, in the switching-over point always occurring at the same point of the sinusoidal voltage V_{g1}, provided other conditions remain unchanged. Which point this will be is determined by E/A and τ/T. A division by the odd number $2n+1$ will occur in the idealized case of Fig. 39-12 when:

$$E\,\mathrm{e}^{-\frac{(2n-1)}{2\tau}T} > A > E\,\mathrm{e}^{-\frac{(2n+1)}{2\tau}T}$$

or:
$$2n - 1 < f_{\mathrm{sync}}/f_0 < 2n + 1$$

where f_{sync} is the frequency of the oscillator and f_0 the constant $(2\tau \ln E/A)^{-1}$.

Slightly different numerical relations are obtained in practical circuits because the assumption that there is an abrupt and complete take-over of the current for $V_{g1} = V_{g2}$ is strictly not true. There is, however, no qualitative difference, i.e.:

a. at a given value of E/A, the factor by which τ will be divided will shift in value at certain frequencies; division by even numbers will also occur over small frequency bands because of asymmetries;
b. the circuit will show its lowest stability in the immediate proximity of these limiting frequencies;
c. because of considerations of stability, the exponential fall should preferably be not less than 15–20 per cent per cycle, therefore $\tau/T < 5$. Since $\ln E/A$ can be made hardly larger than 2.5–3, this means a maximum value of the denominator of 20–30.

In the circuit of Fig. 39-11, f_0 is approx. 500 c/s, so that the frequencies at which the denominator alters will lie at a separation of approx. 1000 c/s. Because of asymmetry, it can be expected that division by even numbers will occur over frequency widths of approx. 100 c/s.

It is required in many applications that the waveform available very closely approximates to the ideal square wave. The waveforms of multivibrators show two shortcomings in this respect: 1) the pulses deviate from the rectangular form, and 2) there is a difference in width between the positive and the negative going pulses.

For point one, we can state that the rise and fall times of the waveform are determined by the speed at which the voltage at a given point can change, i.e. at given current and resistance values by the capacitance of that point to its surrounding. This implies that the parasitic capacitances will also determine the maximum speed. Comparing the circuits of Figs 39-1 and -7, we see in the former that for both anode signals these consist exclusively of stray wiring capacitances and the valve's inter-electrode capacitances, magnified by the Miller effect. One method to reduce the latter is to use pentodes. Because there is no need to connect the right-hand anode to other electrodes a very small capacitance giving short rise and fall times can be achieved in principle.

The circuit of Fig. 37-7 shows an additional effect, usually the most important: during the positive voltage pulses on the anodes, the grids are almost connected to earth by the grid-cathode diode. This increases the capacitance of the anode to earth by the cross-coupling capacitor, which will make the edges much less steep.

These capacitive effects are, of course, also encountered with the corresponding single-shot and bistable circuits and sometimes restrict their application. Apart from avoiding the grid current, as can be achieved in the circuit with a large common cathode resistance, the slope of the edges of the waveform can also be increased by restricting the use of the waveform to the initial part of the exponentially rising voltage. This is achieved by "clipping", i.e. limiting by means of a diode, as illustrated in Fig. 39-13. Another way would be to interrupt the connection between anode and grid during the positive rising pulse by means of a diode, Fig. 39-14, where R_b is usually larger than R_a.

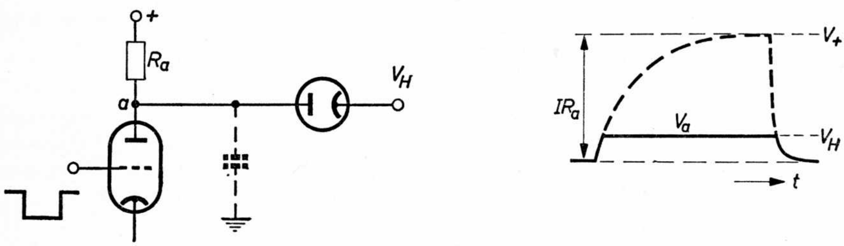

Fig. 39-13

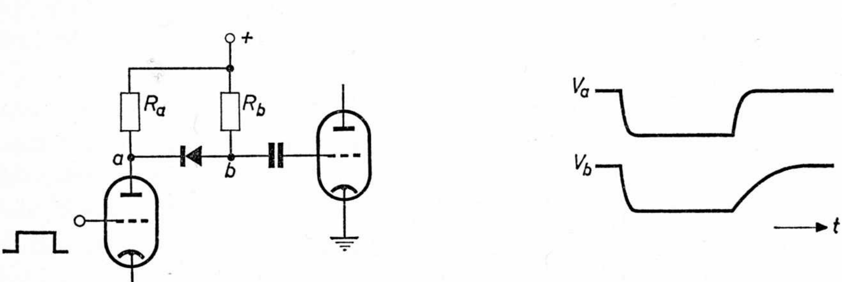

Fig. 39-14

In so far as there is any need for the exact calculation of the switching phenomena, it is seldom that more extended equivalent circuits than that of Fig. 39-15 will be encountered. Considerable simplification can usually be made because, for example, C_k may be much larger than

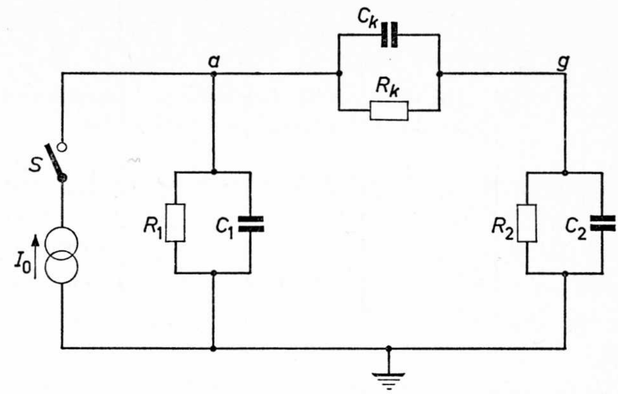

Fig. 39-15

C_1 and C_2. In the general case solutions are derived with Heaviside's formulae (Section 22).

As an example of how to make a simplification, we will calculate the voltage V_g in Fig. 39-15 when the switch S is closed ($t = 0$), first exactly and then using a sensible simplification. Here $C_1 = C_2 = 0.1C_k = C$ and $R_1 = R_2 = R_k = R$. With $Z_1 = R_1/(1 + R_1C_1p)$ etc., we have:

$$v_g = \frac{Z_1Z_2}{Z_1 + Z_2 + Z_k} \cdot I_0U(t)$$

where $U(t)$ is the unit step function at instant $t = 0$.
This yields, substituting the above equalities and $\tau = RC$:

$$v_g\{p(21\tau^2p^2 + 24\tau p + 3)\} = I_0R\{p(10\tau p + 1)\}U(t)$$

The roots of the coefficient of v_g are $p_1 = 0$, $p_2 = -1/\tau$ and $p_3 = -1/7\tau$, so that

$$V_g(t) = I_0R\left(\frac{1}{3} - \frac{1}{2}\mathrm{e}^{-t/\tau} + \frac{1}{6}\mathrm{e}^{-t/7\tau}\right)$$

With the approximated calculation method C_k is assumed to be a short-circuit for the rapidly changing waveform. v_a and v_g at the beginning then both equal $I_0U(t) \times (Z_1//Z_2) = \frac{1}{2}I_0R\ (1 + \tau p)^{-1} \cdot U(t) = \frac{1}{2}I_0R(1 - \mathrm{e}^{-t/\tau})$.
Capacitances C_1 and C_2 can be neglected for the slow changes, so that only the charging current of C_k remains.
Here the time constant is the product of C_k and the parallel combination of R_k and $R_1 + R_2$, in our case $6\frac{2}{3}\tau$. The final value of V_g becomes $\frac{1}{3}\ I_0R$, so that the contribution of the charging current will be

$$-\frac{1}{6}\,I_0R(1 - \mathrm{e}^{-t/6\frac{2}{3}\tau})$$

which gives:

$$V_g(t) = I_0R\left(\frac{1}{3} - \frac{1}{2}\,\mathrm{e}^{-t/\tau} + \frac{1}{6}\,\mathrm{e}^{-t/6\frac{2}{3}\tau}\right)$$

The difference between these two expressions is exceedingly small, especially for the leading and trailing edges.

The circuits of Figs 39-1 and -7 also give a good idea of the waveform in the plateau regions. In Fig. 39-1 the top of the V_{a1} waveform slopes upwards, due to the charging of capacitor C by R_a and R_g; the size of this effect is determined by the ratio R_a/R_g. Voltage V_{a2} on the other anode does not have this shortcoming, but instead voltage V_{g2} is determined by the ratio R_a/R_k.

This can be avoided by replacing R_k by a current source; we can thus obtain flatter plateaux. At very high repetition frequencies, the stray capacitance of the common cathode will counteract the good effect of the current source. If necessary this can be compensated by inserting a small capacitance in the anode. In the circuit of Fig. 39-7 both anode signals will show at their upper and lower plateaux a rounding-off as V_{a1} in Fig. 39-1. It is possible to improve these waveforms by the use of limiting diodes, as shown in Figs 39-13 and -14.

Since the height of the pulse in Fig. 39-1 is determined by the ratio R_a/R_k, or in the case of a current source by I_kR_a, the height is particularly well defined. On the other hand, we see in Fig. 39-7 that the valve characteristic for $V_{gk}\approx 0$ determines the height, and great fluctuations are consequently possible. Improvement can be achieved by clipping off the upper and lower plateaux.

The above applies both to multivibrator and mono- and bistable circuits. However, an additional requirement for multivibrators is that the positive and negative going pulses must be of equal width. Since the duration of both pulses is determined in the circuit of Fig. 39-7 by two different time constants, this can cause large differences if no adjustment is made. Even if this is not so, as in Fig. 39-1, a considerable difference is still possible due to possible differences in the valve characteristics. We can minimize this difference by applying the improvement of Figs 39-9 or -10, but if stringent requirements exist in this respect, it is better to use the combination of a multivibrator and a "flip-flop". This device gives a square-wave voltage whose positive and negative parts have exactly the same width.

It follows from the above that little improvement is needed for the monostable circuit of Fig. 39-1. If necessary, more extensive RC-circuits can be used in order to fix more accurately the waiting times in the unstable state, i.e. the pulsewidth. Due to the unequal grid voltages the switch-over point is already much better defined than in the case of the multivibrator of Fig.

39-1. It may of course be useful to replace R_k by a current source in this circuit as well, so that the signal on the right-hand anode will become a better approximation to the rectangular form.

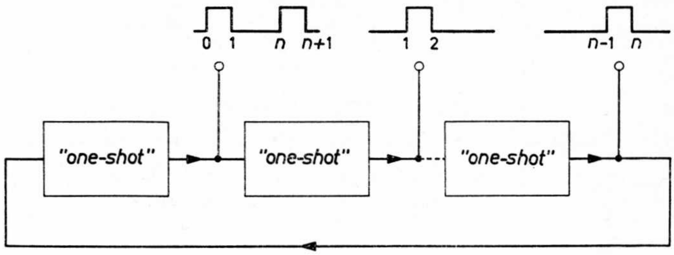

Fig. 39-16

Fig. 39-16 shows an application of monostable stages where a very constant pulsewidth may be of importance. By making a closed loop of these circuits, we see that each is triggered by the trailing edge of the output signal of the preceding stage. This results in pulses at the several points which are mutually shifted in time and are repeated at a certain frequency. The stability of this frequency will be determined by that of the pulses. This type of circuit is often used for measurements where a given sequence has to be followed and where the pulses are used to initiate the various actions of the sequence.

The symmetrical form of the Schmitt trigger is usually designated by the term "flip-flop". Fig. 39-17 shows the circuit with a resistor in the common cathode circuit, where control is effected symmetrically on the cathode by means of an additional valve (III). The d.c. grid bias of valve III is chosen so that this valve does not normally conduct. However, if this voltage is increased for a short time so that valve III takes over the current of the two other valves, the grid voltage of the non-conducting valve of the flip-flop system will make such a large positive step that this grid becomes positive with respect to the other one. By making the coupling capacitors so large that this situation still prevails when the flip-flop system again takes over the cathode current, the valve which was closed first will now conduct. In other words, the flip-flop has been switched. The capacitors between anodes and grids have therefore essentially a "memory" function. Since the control from valve III has no preference for letting conduct either valve I or valve II this memory function is essential.

The short positive pulse on the grid of the auxiliary valve can be obtained, for example, from a differentiated square wave. Valve III can be omitted if

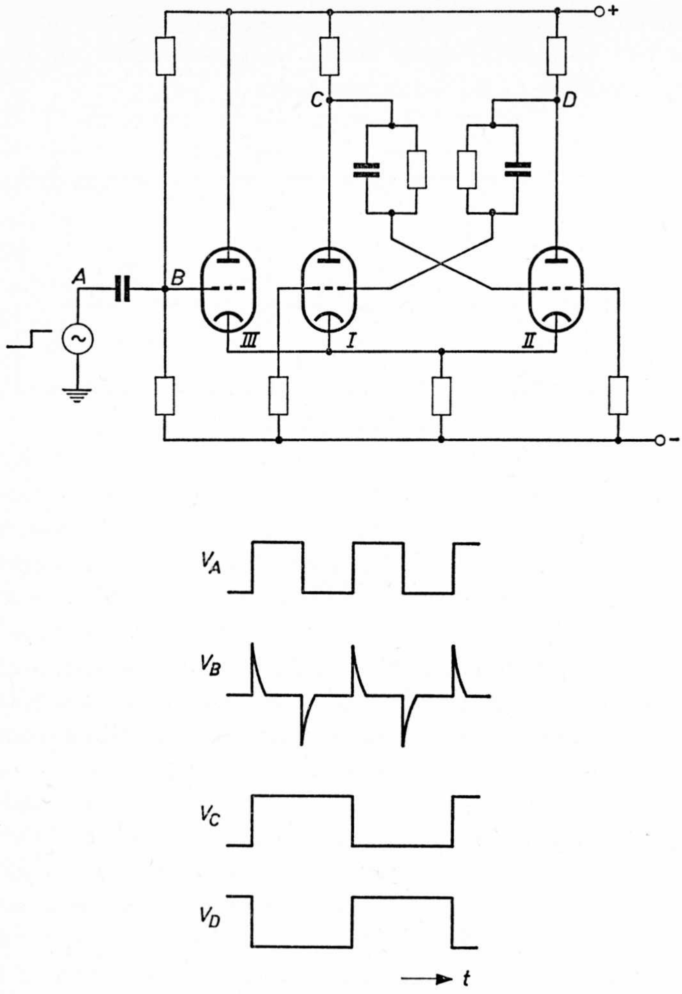

Fig. 39-17

the load on the signal source is acceptable. The source can then be directly connected to the common cathode through a capacitor. Another possibility is that the two grids are controlled simultaneously.

Square waves will occur on anodes *C* and *D* of the flip-flop, but with twice the period of the original square-wave voltage at *A*, so that this circuit can be used for frequency division. We should mention here that whilst the difference between the positive and negative halves of the square-wave voltage at *A*, if originating from a multivibrator circuit, can be quite considerable, this difference will be automatically reduced to zero in the square waves at *C* and *D*, apart from parasitic effects.

Apart from frequency division, the flip-flop circuit can also be used for counting pulses, as illustrated in Fig. 39-18, where three schematically indicated stages of Fig. 39-17 are connected in cascade. Let us indicate by the figure 1 for each stage that the right-hand valve conducts current, and by

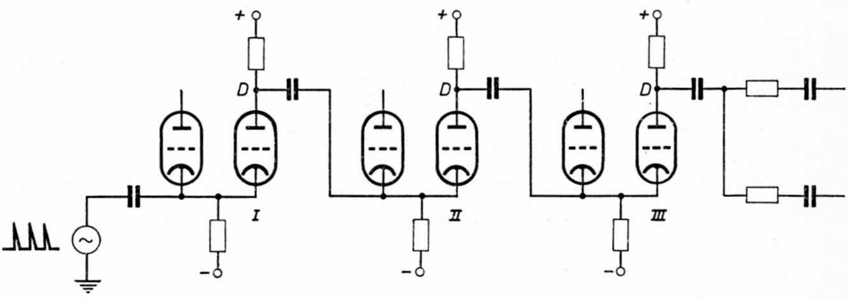

Fig. 39-18

the figure 0 the other situation. We then successively obtain, starting from 000, the following combinations for stages III, II and I, when a series of pulses is applied to the input of valve I: 000, 001, 010, 011, 100, 101, 110, 111, 000. However, these combinations just represent in binary form the number of pulses applied to the input. The flip-flop is for that reason the basic counting circuit used in digital electronic computers.

It is also possible to divide by numbers other than powers of 2. For example, decimal dividers are obtained as indicated in Fig. 39-19. The top circuit is so designed that when the first stage changes from situation 1 to 0, the fourth stage will also change over if it is in situation 1, whilst the feedback from IV to II ensures that the latter is blocked when IV is in situation 1. Starting from situation 0000, the counting of the first 7 pulses will take place in the same manner as before. At the 8th pulse, the fourth stage will change over by the effect of the third: 1000, whilst at the 9th pulse only the first stage will change over: 1001. At the 10th pulse, the first stage will again change over, but now also the fourth, whilst the feedback to the second stage will maintain the latter in situation 0: 0000.

The bottom circuit ensures that by changing over from situation 0 to 1 of the fourth stage, the second and third stages receive an additional pulse. The circuit works the same as the previous one up to the seventh pulse: 0111, but at the eighth, this is not changed into 1000 but into 1110. The ninth now gives 1111 and the tenth 0000. In this circuit, situation 1 of the fourth stage has therefore value 2.

There are also valves where a beam of electrons or a gaseous discharge has ten stable states, which can be controlled by means of pulses. These valves give a direct indication in the decimal system.

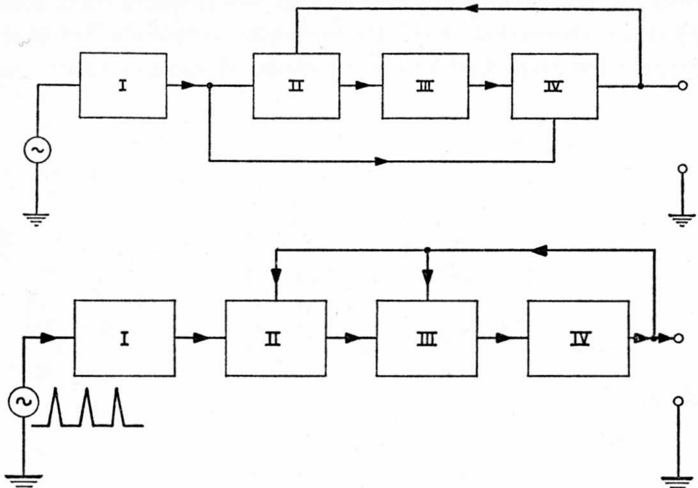

Fig. 39-19

It is obvious that the flip-flop system must meet high requirements of reliability. The circuit must not miss any input pulse, neither be switched by possible interference pulses. When triggered by rectangular pulse waveforms, we must also ensure that counting only takes place on one of the pulse's edges. All of these demands must still be met throughout the life of the circuit, even taking into account changes in components which often have rather wide tolerances. Fig. 39-20 shows the principle of a circuit where diode

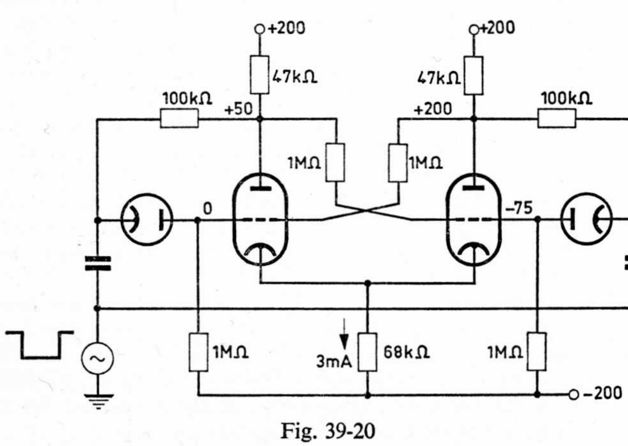

Fig. 39-20

control is used in the input circuit which allows a greater input tolerance. The numbers in the circuit diagram indicate the mean voltages at different points for the case that the left-hand valve is conducting. The diodes are then reverse-biased to 50 and 275 volts respectively and the valve grids are certainly unaffected by the positive going edges of the squared waveform. For a negative going edge with a value of approx. 100 volts, the left-hand valve must necessarily conduct, and the right-hand grid will be unaffected by this edge, so that the circuit is bound to change over and will always do so, even if one or more components are considerably aged.

The original flip-flop circuit of Eccles and Jordan (1919) which is still the basis of most current types, does not have a common cathode resistance (Fig. 39-21). One means of control as shown in this diagram is by narrow negative pulses or differentiated square waves applied to the grids. Highly reliable control can be achieved by the use of diodes. When the valve conducts there is hardly any potential difference between grid and earth, so that the corresponding diode is hardly reverse-biased. On the other hand, the reverse bias on the other diode amounts to many tens of volts. The circuit's operation has proved to be fully reliable with the component values indicated in Fig. 39-21, for heights ranging from 1 to over 40 volts.

In the non-conducting state, the valve is an almost ideal switch; the leakage current remains small even at high anode-cathode voltages. On the other hand, the valve has a serious shortcoming in the conducting state in so far as an anode voltage of many tens of volts is required to pass the usual currents, even when the grid-cathode bias is reduced to zero. This voltage

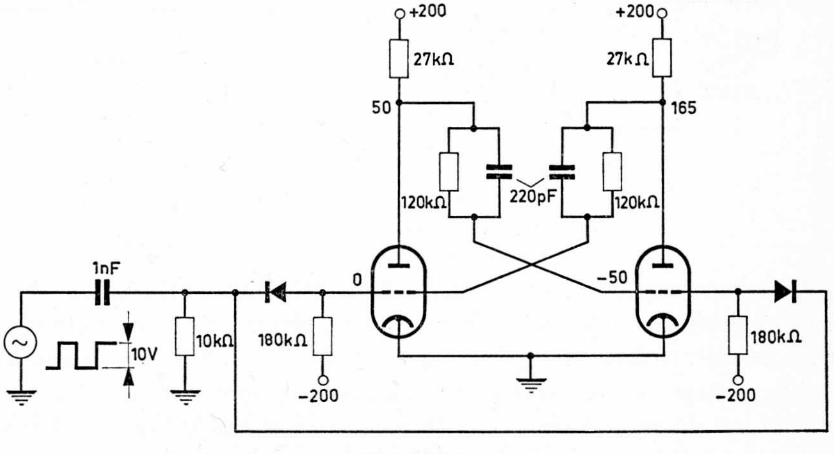

Fig. 39-21

is therefore 10–40 per cent of the normal supply voltage; it is not very constant and unfortunately dependent on the heater voltage and the age of the valve. Therefore circuits in which valves are used in this way produce a variable signal height, and, for example, flip-flop circuits with more than two stable states are difficult to design. Analogous to the ordinary flip-flop, each grid in such a circuit is connected to the anodes of all other valves through a resistance divider; switching of one of the valves must then produce a significant change in all the grid voltages. This, however, is a particularly difficult condition to realise, because of the loosely defined anode voltages when the valves conduct. Improvements are possible with the introduction of diodes, but it would be very difficult to design a flip-flop with ten stable states in this manner.

A property of the conducting valve which does correspond to an ideal switch is that for $V_{gk} < -1.5$ volt, i.e. in the absence of grid current, the anode current is exactly equal to the cathode current. This property is utilized in circuits with a large common cathode resistance and these therefore obtain a better signal form than can be obtained with other designs. However, the power dissipated by this type of circuit is considerable.

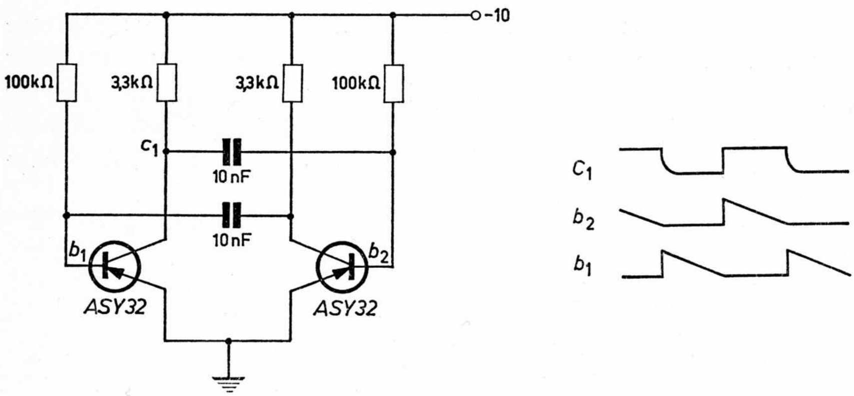

Fig. 39-22

As we have seen when discussing d.c. amplifiers (Fig. 35-12) the transistor is much closer to an ideal switch than the valve, particularly in the conducting state. In order to pass the usual currents, the transistor only needs a collector-emitter voltage of a few tenths of a volt, i.e. not more than approx. 1 per cent of the supply voltage. A first consequence of this is that simple and reliable transistorized multistable circuits are easier to design. Another

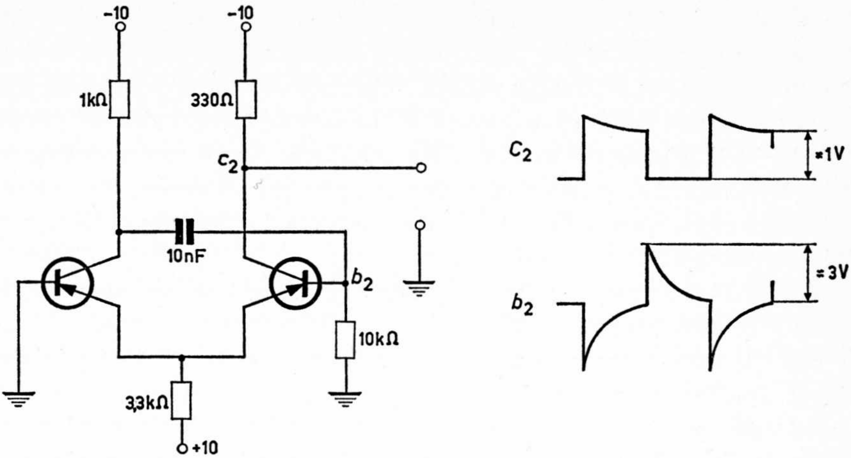

Fig. 39-23

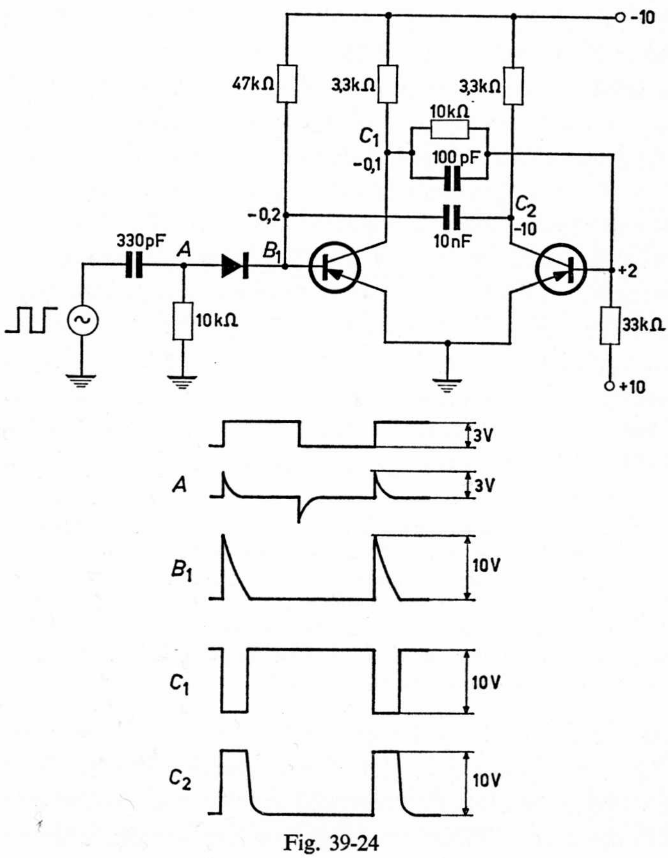

Fig. 39-24

consequence is that the collector signal height is very well defined in the multivibrator circuit of Fig. 39-22, which corresponds to Fig. 39-8, and is almost the same as the supply voltage. When a transistor is conducting, the collector current is not exactly equal to the emitter current and the circuit of Fig. 39-22 is preferred to that of Fig. 39-23 (the circuit corresponding to Fig. 39-1) regarding stability of output pulse height. However, the mutual loading is much lower in Fig. 39-23, so that very good waveforms are obtained without using additional diodes.

Fig. 39-24 is an example of a monostable circuit of two transistors which are not emitter-coupled.

For the same reasons as apply in Fig. 39-21, the use of diodes in the input circuit of a transistorized flip-flop is particularly efficient. The very small difference in potential between collector and base in the conducting transistor is such that the corresponding diode is hardly reverse-biased, while the other diode will have a reverse bias of more than 10 volts. The permissible margin in the amplitude of the control signals is then relatively large. This is illustrated in Fig. 39-25 where the voltages indicated are valid when the left-hand transistor is conducting. The collector potential of the conducting transistor is so low that the voltage across the corresponding diode (D_1) even reverses its polarity. The reverse voltage across the other diode D_2 is 10 volts. By means of diodes D_3 and D_4 we ensure that points a_1 and a_2 will rapidly reach the value of -0.1 volt after the corresponding transistor has just begun to conduct. This considerably shortens the "recovery time" during which there is no reaction to any following input signals. This time is a few microseconds in the circuit of Fig. 39-25.

The maximum repetition frequency of transistorized circuits is chiefly determined by the transistors themselves. In this respect, the use of transistors at very low collector voltages is disadvantageous because of the "hole storage" effect. This effect will therefore be avoided for very fast circuits.

The hole storage effect occurs because the base-collector diode first conducts at low collector voltages and must then be reversed; this reversal involves a great change in the concentration of the minority charge carriers at the junction and this takes a certain time. These new charge concentrations can be reached more quickly by incorporating specific impurities in the base and/or collector material. This is done in special switching transistors.

Because of the much lower impedance level, transistorized circuits are far less sensitive to capacitively induced interference voltages. On the other hand, low-frequency multivibrators and monostable circuits with a large pulsewidth are more difficult to design because, among other reasons, of

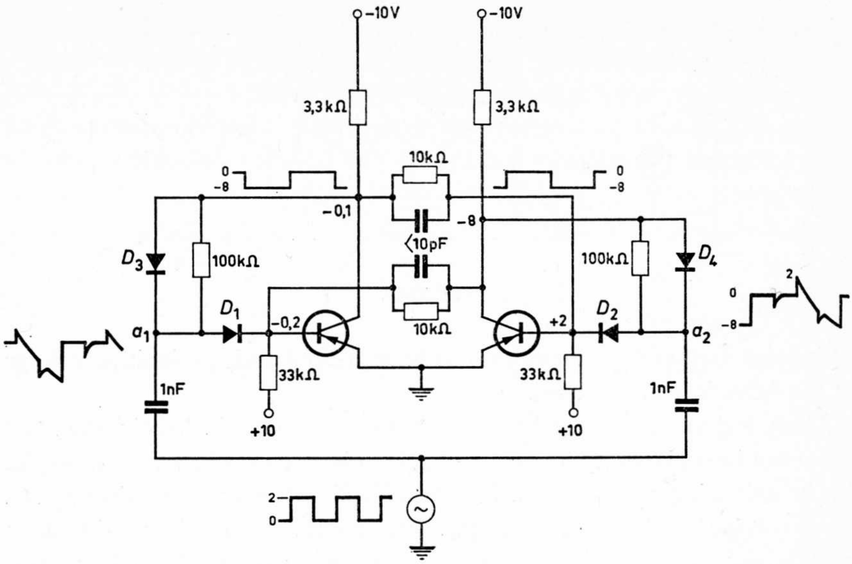

Fig. 39-25

the relatively large base current. The field-effect transistor gives a great improvement in this respect.

It used to be an inherent disadvantage of transistors that the maximum permissible reverse voltage across the base-emitter junction was very low, so that the height of voltage pulses in this direction nearly always had to be limited. Although progress has been made in recent years, this drawback still exists for very fast transistors.

The repetitive "sawtooth" waveform is often used. It consists of a ramp part ascending linearly with time, and followed by a rapid return drop to the reference level. This waveform is obtained by charging a capacitor linearly, then rapidly discharging it. The circuit of Fig. 39-26 gives a rough first approximation and utilizes the difference in potential between the burning and ignition voltages of a neon stabilizer valve. For example, for the 85A2 valve these voltages are 85 and approx. 115 volts respectively. If the valve is not burning, the voltage across the capacitor will increase more or less linearly with time. As soon as the valve strikes, a considerable discharge will occur as a result of the low internal resistance of the valve. The voltage across the capacitor will then drop rapidly to about the burning voltage. If the current passing through resistor R at this instant is smaller than is necessary to maintain the discharge in the neon valve, the latter will

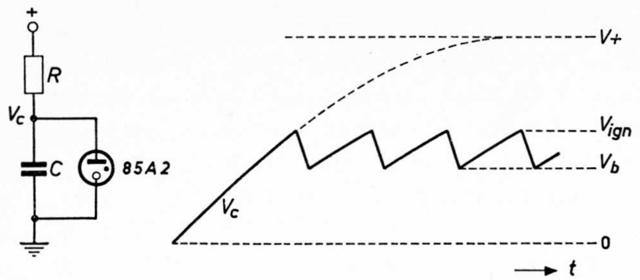

Fig. 39-26

extinguish itself and the capacitor will be recharged until the ignition voltage of the valve is again reached.

Many improvements can be added to this basic circuit. Firstly, the neon valve can be replaced by a directly coupled bistable circuit, as, for example, the Schmitt trigger (Fig. 39-27). This circuit will make valve B conducting at a given voltage V_H across the capacitor. This enables the capacitor to discharge rapidly, and the valve will be made non-conducting again below a lower voltage V_L. The reversal voltages are thus solely determined by the two stable states of the Schmitt trigger and can therefore be defined very accurately. We shall deal with these circuits in Section 41.

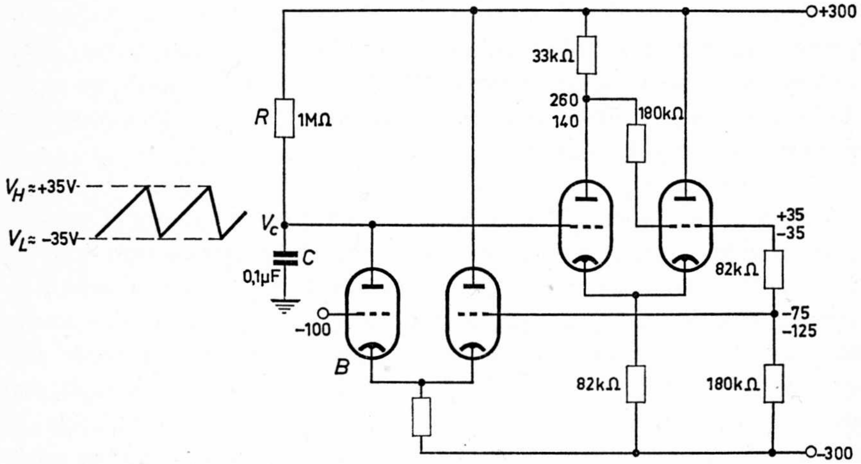

Fig. 39-27

Charging through a resistor only gives good linearity if only use is made of the beginning of the exponential curve. An improvement is possible by using a constant current source instead of the resistor. The simplest solution is obtained by using a transistor with a large emitter resistance (Fig. 39-28). The maximum allowed voltage is sometimes too small for application, and the valve circuit of Fig. 39-29 can be used when large amplitudes are required. The voltage waveform at point *A* is here almost entirely reproduced at point *B* by means of the cathode follower and the neon valve, so that resistor *R* is apparently increased by a factor which might easily be as much as 20 times. In most cases it is possible to replace the neon valve by a sufficiently large capacitor.

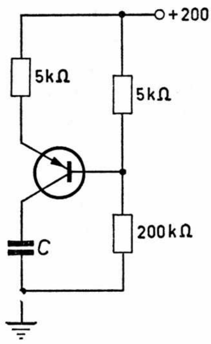

Fig. 39-28

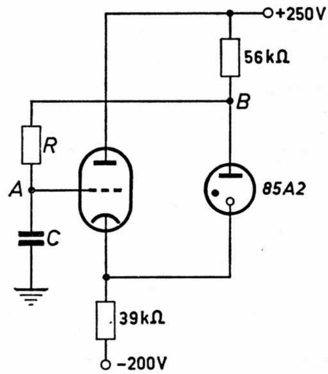

Fig. 39-29

One use of a sawtooth is as a time base X-deflection voltage in an oscilloscope. Synchronization is also possible here. As we found for the synchronization of a multivibrator, the exact instant of fly-back can be determined by an external signal, and a jitter-free image can thus be obtained with a periodic signal.

40. Amplitude and phase measurements

One of the commonest forms of "processing" a signal is the determination and possible recording of the amplitude of an a.c. voltage which varies relatively slowly. This is found for example in telecommunications when demodulation is carried out, i.e. the derivation of an information signal from an amplitude modulated waveform. The information is expressed by the amplitude variations in the modulated signal. In measurement electronics, the amplitude of an a.c. voltage is often proportional to the original effect to be measured, so that measurement of the amplitude leads directly to the required information. The required degree of accuracy, however, is usually much higher here than in telecommunications. In this section we shall now discuss the various factors which have a bearing on amplitude measurement.

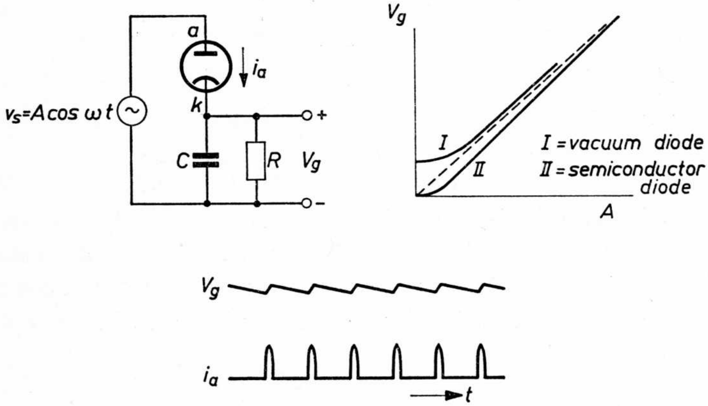

Fig. 40-1

Most methods of measuring amplitudes are based on the transformation of an a.c. voltage into a d.c. voltage, whose value is proportional to the amplitude of the a.c. voltage. One of the simplest circuits to achieve this reasonably is the rectifier circuit of Fig. 40-1. The relation between amplitude A of the a.c. voltage and d.c. voltage V_g across the capacitor has a reasonable linearity as will be seen from the following calculations. This applies as long as the signal value exceeds a certain minimum value, and the leakage resistance of the capacitor is sufficiently large. At small signal amplitudes, the difference between the characteristics of the valve and semiconductor diodes and the ideal diode becomes noticeable, and quite large deviations from linearity will occur (Fig. 40-1).

For fairly accurate results, the circuit with a valve diode can be used for signals of a few or more volts, and the circuit with a semiconductor diode for signals exceeding approx. 0.5 volt.

How much the a.c. voltage source will be loaded by this type of circuit is also important for many measurements. The diode conducts for short periods so that the diode current i_a consists of short narrow pulses (bottom of Fig. 40-1). If the signal source has a negligibly small internal resistance for the higher harmonics in this current, so that only the fundamental of i_a contributes to the loss in energy, this load can be expressed by an equivalent resistance R_b. This resistance will be accurately determined for a few diode characteristics in the following calculations, but its value follows to a first approximation from the simplification that all energy supplied by the source is dissipated in resistor R, so that:

$$\frac{A^2}{2R_b} = \frac{V_g{}^2}{R}$$

As we always attempt to make $V_g \approx A$ in a rectifying circuit, it follows:

$$R_b \approx \tfrac{1}{2}R$$

This result is only valid for circuits where only the d.c. voltage V_g is applied across resistor R. For example, Fig. 40-2 shows a circuit where not only this voltage, but also the a.c. voltage is applied across R. The load resistance will thus equal the parallel value of $\frac{1}{2}R$ and R, therefore $\frac{1}{3}R$.

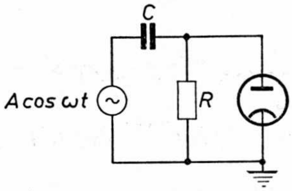

Fig. 40-2

We shall now calculate for a few diode characteristics the ratio V_g/A of the d.c. voltage to the amplitude of the a.c. voltage as well as the load resistance R_b, as functions of the signal amplitude.

Let us first assume that the diode possesses an exponential relation between voltage and current. We then find to a good approximation for the

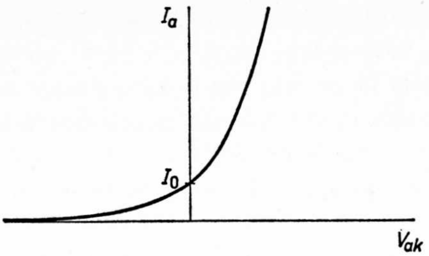

Fig. 40-3

valve diode: $I_a = I_0 e^{aV_{ak}}$ (Fig. 40-3). a usually has a value of approx. 10 volt^{-1}. With $v_s = A \cos \omega t$ and neglecting the ripple voltage impressed on V_g (which is usually small), i.e. $V_{ak} = v_s - V_g$, we have:

$$i_a = I_0 \, e^{-aV_g} \, e^{aA \cos \omega t}$$

We have for the expansion of $e^{aA \cos \omega t}$

$$e^{aA \cos \omega t} = J_0(\mathrm{j}aA) + 2 \sum_{n=1}^{\infty} (-\mathrm{j})^n J_n(\mathrm{j}aA) \cos n\omega t$$

where $J_k = k$-th order Bessel function, so that:

$$i_a = I_0 \, e^{-aV_g} \left(J_0(\mathrm{j}aA) + 2 \sum_{n=1}^{\infty} (-\mathrm{j})^n J_n(\mathrm{j}aA) \cos n\omega t \right)$$

This yields for the d.c. component i_{a0} and the amplitude of the fundamental current i_{a1}:

$$i_{a0} = I_0 \, e^{-aV_g} \, J_0(\mathrm{j}aA)$$

$$i_{a1} = -2\mathrm{j} \, I_0 \, e^{-aV_g} \, J_1(\mathrm{j}aA)$$

We also have: $V_g = i_{a0} R$ and $R_b = \dfrac{A}{i_{a1}}$

Therefore: $$V_g \, e^{aV_g} = I_0 R \, J_0(\mathrm{j}aA) \qquad (40.1)$$

and $$\frac{R_b}{R} = \frac{1}{2} \frac{A}{V_g} \frac{J_0(\mathrm{j}aA)}{-\mathrm{j}J_1(\mathrm{j}aA)} \qquad (40.2)$$

Fig. 40-4 gives $J_0(\mathrm{j}\alpha A)$ and $-\mathrm{j}J_1(\mathrm{j}\alpha A)$, as well as their ratio $J_0/-\mathrm{j}J_1$ as a function of αA. It follows that for $\alpha A > 5$, both functions will equal $\mathrm{e}^{\alpha A}$ (except for a constant factor $\approx \mathrm{e}^2$) and the ratio will approximately equal unity. This means that in (40.1) and (40.2) at the usual values of $\alpha I_0 R$ ($10^2 < \alpha I_0 R < 10^5$), the relative difference between A and V_g becomes small, and R_b approximates to $\frac{1}{2}R$.

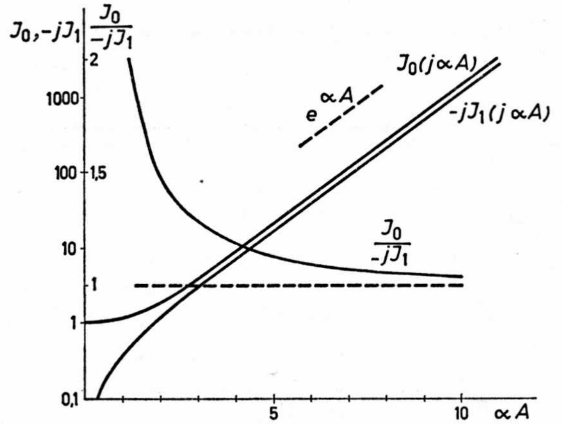

Fig. 40-4

We can also write for (40.1): $\mathrm{e}^{(\alpha V_g + \ln \alpha V_g)} = \alpha I_0 R J_0(\mathrm{j}\alpha A)$, which can be approximated with the above values of $\alpha I_0 R$ and $J_0(\mathrm{j}\alpha A) > 1$ for all values of αA to:

$$\alpha V_g = \ln \alpha I_0 R + \ln J_0(\mathrm{j}\alpha A)$$

This relation between V_g and A is represented for a few values of $I_0 R$ in the upper curves of Fig. 40-5, whilst the lower curves show the relation for R_b. A value of 10 volt^{-1} has been assumed here for α. At a signal voltage of a few volts, the deviation from a linear relation between V_g and A will be quite small and R_b has almost reached the value $\frac{1}{2}R$.

Example: Fig. 40-6 shows a rectifier circuit with valve diode, used as an a.c. voltage probe for valve voltmeters. The exponential relation between current and voltage can be assumed as a sufficiently accurate approximation of the characteristic of double-diode EAA91 at small currents. This relation is shown on the right-hand side of the figure for both a new and a very old specimen. In both cases, $\alpha \approx 10$ volt^{-1}, but the value for current I_0 and also for the product $I_0 R$ differ by about a factor 200. We find in these extreme cases for $I_0 R$: 5×10^4 and 250 respectively. However, Fig. 40-5 shows that this only produces a small difference in the ratio V_g/A at signal voltages exceeding approx. 10 volts.

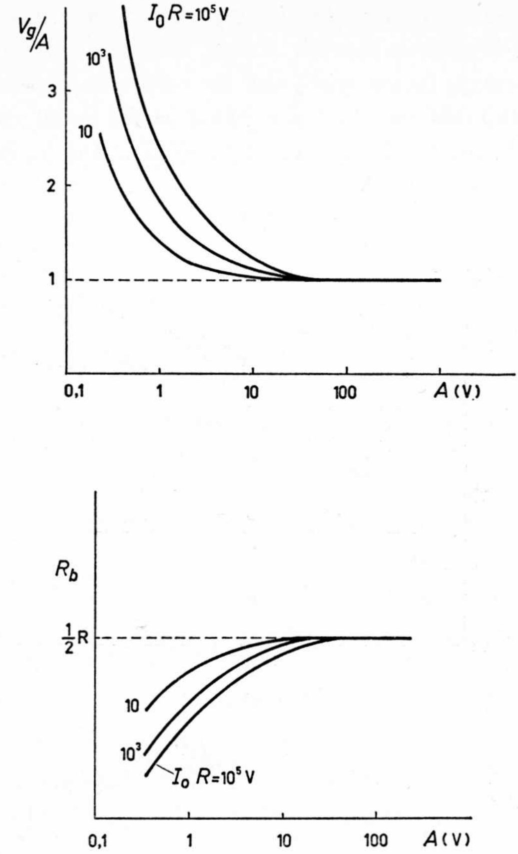

Fig. 40-5

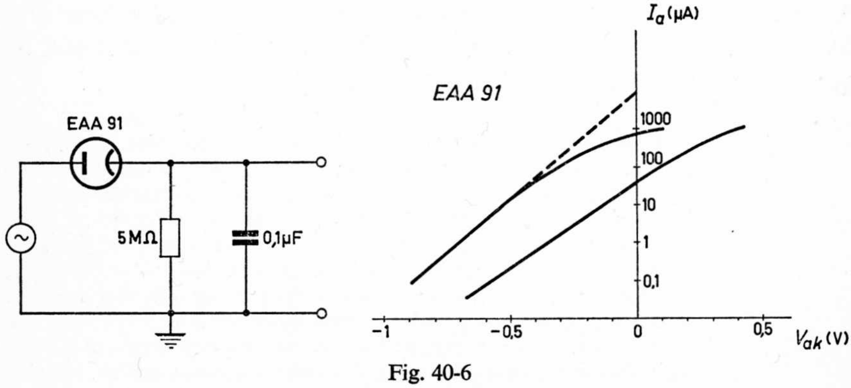

Fig. 40-6

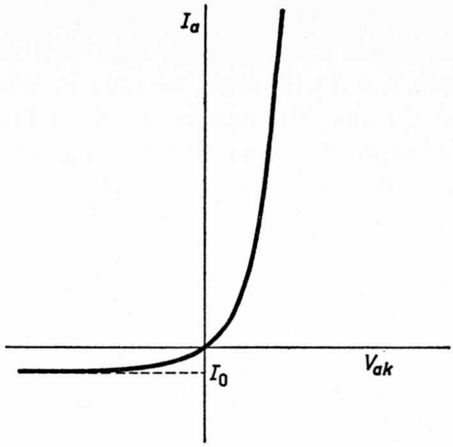

Fig. 40-7

Let us now consider the diode characteristic $I_a = I_0(e^{aV_{ak}} - 1)$ of Fig. 40-7, which is applicable to many semiconductor diodes and where $a = 40$ volt^{-1}. Compared to the previous case, this only means a change in i_{a0}:

$$i_{a0} = I_0 \left\{ e^{-aV_g} J_0(\mathrm{j}aA) - 1 \right\}$$

Equation (40.1) can now be written:

$$(V_g + I_0R)\, e^{aV_g} = I_0R\, J_0(\mathrm{j}aA)$$

I_0 is approx. 10^{-7} A for normal germanium diodes and approx. 10^{-10} A for silicon diodes, so that, with the usual values of $R\,(\approx 10^6$ ohms), I_0R will attain at most the same order of magnitude as A and V_g.

Fig. 40-8 gives the ratio V_g/A as a function of A for a few possible values of I_0R, where $a = 40$ volt^{-1}. It follows that this product should preferably be made

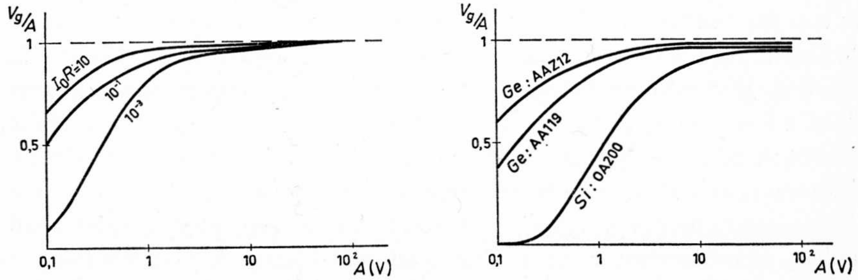

Fig. 40-8

larger than unity. In order to limit the output impedance R of the circuit to a reasonable value, I_0 should not be smaller than 10^{-5} A and germanium diodes should be used. For $R=10^6$ ohms, the ratio V_g/A has a value of 0.95 at a signal voltage of 1 volt. The right-hand side of Fig. 40-8 shows the measured results on two Ge-diodes and one Si-diode. R was here 10^6 ohms. For the Si-diode the product I_0R proves to be considerably smaller than 10^{-3}.

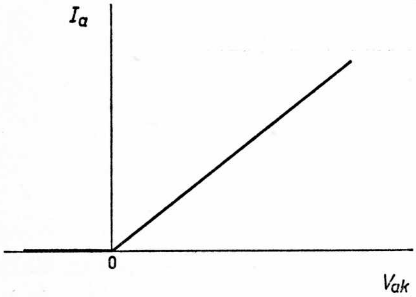

Fig. 40-9

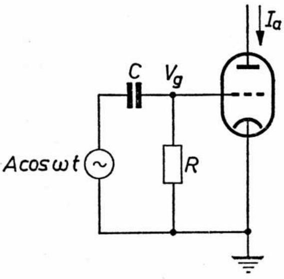

Fig. 40-10

Now assuming the ideal diode characteristic of Fig. 40-9, the ratio of V_g to A will be linear and the difference between these two values will decrease with increasing R and slope $S=\Delta I_a/\Delta V_{ak}$ of the conducting diode. It appears from a simple calculation, where the function A cos ωt for the relevant small value of ωt can be approximated to $A\{1-\frac{1}{2}(\omega t)^2\}$, that the relative difference $(A-V_g)/A\approx 3(SR)^{-2/3}$, and that the product SR must be greater than 5000 to obtain a deviation of less than 1 per cent.

In order to keep the loading of the signal source small the resistor R in Fig. 40-1 should be made as large as possible. However, the disadvantage is that the rectifier circuit has a large output impedance and can hardly be loaded at all. The use of a triode gives considerable improvement (Fig. 40-10) when the grid-cathode voltage V_g is directly converted into an anode current I_a of a few milliamps. We call this grid detection. However, because of the curvature of the I_a-V_g characteristic, the deviation from a linear relation between I_a and V_g is fairly large, which renders this method less suitable for measurement purposes. The same applies to the corresponding transistorized circuit, where moreover the maximum allowed voltage is relatively low.

Even with large values of R the loading on the signal source is still fre-

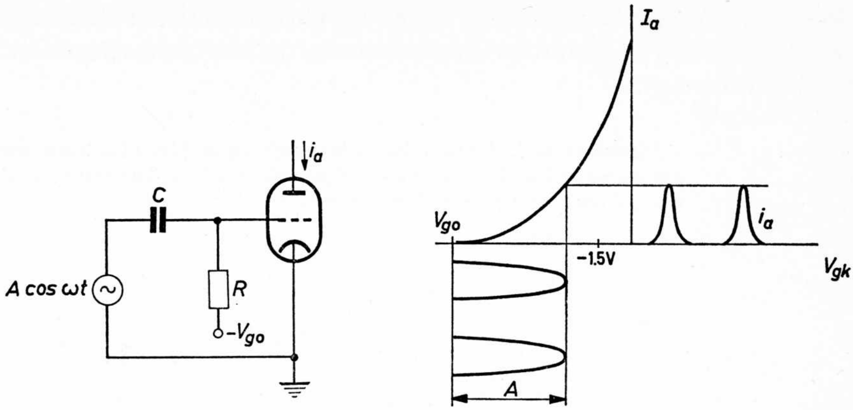

Fig. 40-11

quently a nuisance. This can be avoided by using a triode (Fig. 40-11) where the negative grid bias voltage has been chosen so that no current will pass through the valve in the absence of an a.c. signal, whilst the signal voltage is limited to such an amplitude that the grid voltage remains at least 1.5 volt negative. The grid current is then negligible. However, in this case too the strong non-linear relation between grid voltage and anode current imposes a restriction, whilst the sensitivity depends greatly on the bias adjustment. A much better circuit can be obtained by ensuring a linear relation between I_a and V_g by means of a large cathode resistance. We can then take the cathode as the output terminal and obtain the circuit of Fig. 40-12. This corresponds entirely to the diode detection circuit; the only difference being that the charge supplied to the capacitor is no longer derived from the signal source, but from the anode circuit. Because of this inherent property of not loading the signal source, this network is known as the "infinite impedance

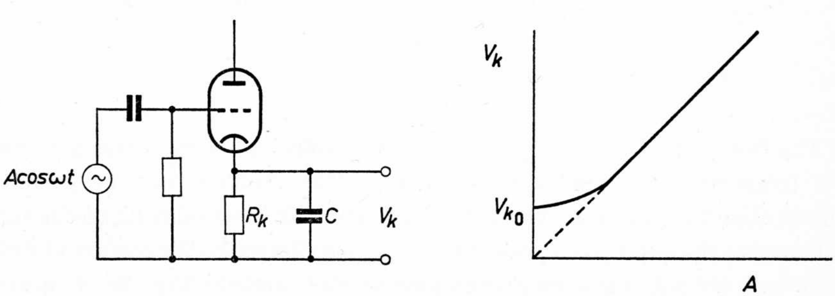

Fig. 40-12

detector". Some load on the signal source will obviously exist in the case of the corresponding transistorized circuit because of the base current. However, this is rather small.

Measured results for the relation between V_e and A for a Ge-transistor are shown in Fig. 40-13 with $R_s = 0$ and $R_s = 1\ M\Omega$. The input resistance of the circuit proves to be a few $M\Omega$.

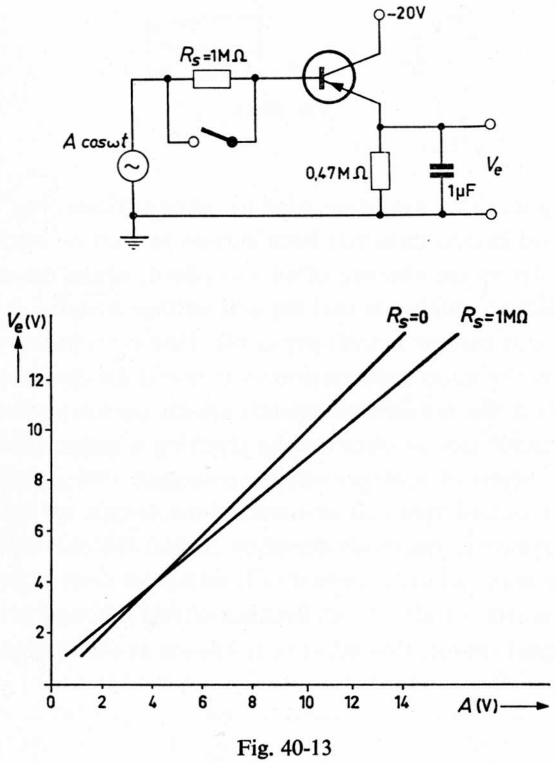

Fig. 40-13

The bias voltage V_{k0} of Fig. 40-12 is many volts and is also dependent on the temperature and the supply voltages. This voltage can be largely compensated by using a balanced circuit. The only remaining disadvantage is then that the latter has a poor response to small signals. We can avoid this drawback by only measuring variations of large signals. Fig. 40-14 shows the circuit for valves.

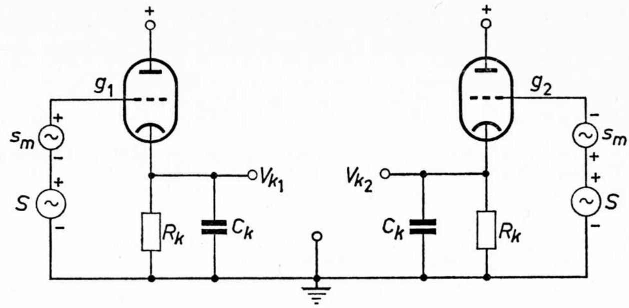

Fig. 40-14

We apply the sum and the difference of the measurement signal s_m and a much larger auxiliary signal S to the left-hand and right-hand grids respectively. S has here a constant amplitude. With equal frequency and phase for S and s_m the relation between the output voltage and the amplitude of s_m will have a very good linearity, particularly if the two triode circuits have been designed as symmetrically as possible. When changing the phase between s_m and S, the output voltage will also change, and at a phase difference of 180° it will be equal but opposite to the signal for zero phase difference. The detection circuit is thus phase-sensitive. If the frequencies of S and s_m are not equal, the amplitudes of the sum and difference signals will vary at the difference frequency. If the capacitances C_k have such a value that the detector can follow these changes, V_{k1} and V_{k2} will also vary at this frequency. By offering these signals to a low-pass RC-filter (Fig. 40-15) we can achieve the effect that only a narrow frequency band on both sides of frequency ω_s of the auxiliary signal will be passed. We have thus obtained

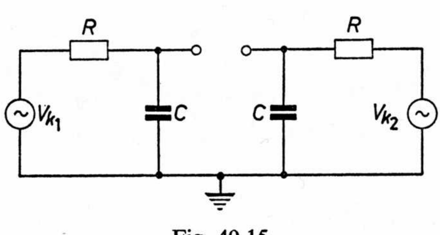

Fig. 40-15

a band-pass filter which follows the frequency of the auxiliary or synchronization signal. Such a detection circuit is called a "synchronous detector", so that the circuit of Fig. 40-14 with the addition of the RC-filters of Fig. 40-15 can be called a phase-sensitive synchronous detector.

The above can be expressed as follows in mathematical form for the circuit of Fig. 40-14:

When $$S = S_0 \cos \omega_s t = \mathrm{Re}\left\{ S_0 e^{j\omega_s t} \right\}$$

and $$s_m = s_{m0} \cos\left\{(\omega_s + \Delta\omega)t + \varphi\right\} = \mathrm{Re}\left[s_{m0}\, e^{j\{(\omega_s + \Delta\omega)t + \varphi\}}\right]$$

we have:

$$v_{g1} = \mathrm{Re}\left[\left\{ S_0 + s_{m0} e^{j(\Delta\omega \cdot t + \varphi)}\right\} \cdot e^{j\omega_s t}\right]$$

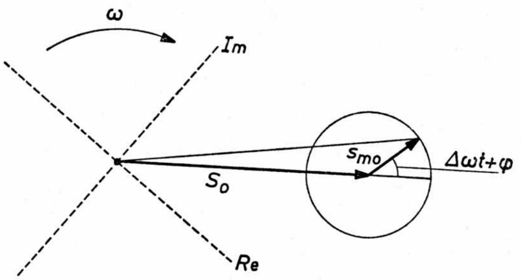

Fig. 40-16

Assuming an ideal operation of the detection circuit and that $\omega \gg \Delta\omega$ the cathode voltage v_{k1} will be increased for each cycle of the carrier-wave frequency to a value which corresponds to the amplitude of $S_0 + s_{m0} e^{j(\Delta\omega\, t + \varphi)}$ (Fig. 40-16). If the time constant $C_k R_k$ is not too large, the cathode voltage will be able to follow the slow fluctuations in the amplitude. We have therefore:

$$v_{k1} = \sqrt{S_0^2 + 2\, S_0\, s_{m0} \cos\,(\Delta\omega \cdot t + \varphi) + s_{m0}^2}$$

Similarly we find:

$$v_{k2} = \sqrt{S_0^2 - 2\, S_0\, s_{m0} \cos(\Delta\omega \cdot t + \varphi) + s_{m0}^2}$$

For values of $s_{m0} \ll S_0$, these waveforms can be approximated to

$$S_0 \sqrt{1 \pm 2 \frac{s_{m0}}{S_0} \cos(\Delta\omega \cdot t + \varphi)} \approx S_0 \pm s_{m0} \cos(\Delta\omega \cdot t + \varphi)$$

so that the difference voltage $v_{k1} - v_{k2}$ has the value $2 s_{m0} \cos\,(\Delta\omega . t + \varphi)$.

For signals of the same frequency as the synchronization signal, i.e. $\Delta\omega = 0$, the output signal becomes $2\, s_{m0} \cos \varphi$, i.e. proportional to the amplitude of the measurement signal and the cosine of phase angle φ between measurement and synchronization signal.

For signals with $\Delta\omega \neq 0$, voltage $v_{k1}-v_{k2}$ becomes an a.c. voltage with frequency $\Delta\omega$, which will be reduced by the RC-filter of Fig. 40-15 by a factor $\sqrt{1+\Delta\omega^2\tau^2}$. This frequency-dependence corresponds to that obtained from an LC-filter: $\sqrt{1+Q^2\beta^2}$. If the same sharp form were obtained without using synchronous detection, we would have: $Q\beta=\tau\Delta\omega$ and $\beta=2\Delta\omega/\omega_s$, which gives: $Q=\frac{1}{2}\omega_s\tau$. For example, with $\omega_s=5000$ and $\tau=1$ second, this gives for Q a value of 2500.

Compared to some other synchronous detectors, the circuit of Fig. 40-14 has the advantage that no frequency bands centred on harmonics of ω_s are passed. Disadvantages of this circuit are its extreme sensitivity to changes in the amplitude of the auxiliary signal and the deviation from linearity between input and output signals in the case where s_{m0} becomes of the same order of magnitude as S_0. The phase sensitivity is also a disadvantage in some measurements. We shall refer to the latter point when discussing demodulation, but we can mention here that small changes in the phase angle around zero have little effect. This is so because at $\varphi\approx 0$, we have $\cos\varphi\approx 1-\frac{1}{2}\varphi^2$, so that at, say, $\varphi=8°$ the sensitivity is only reduced by 1 per cent.

A very great improvement in linearity and independence of the amplitude of the synchronization signal is obtained by applying the principle of Fig. 40-17. Here the auxiliary signal S is not sinusoidal but square

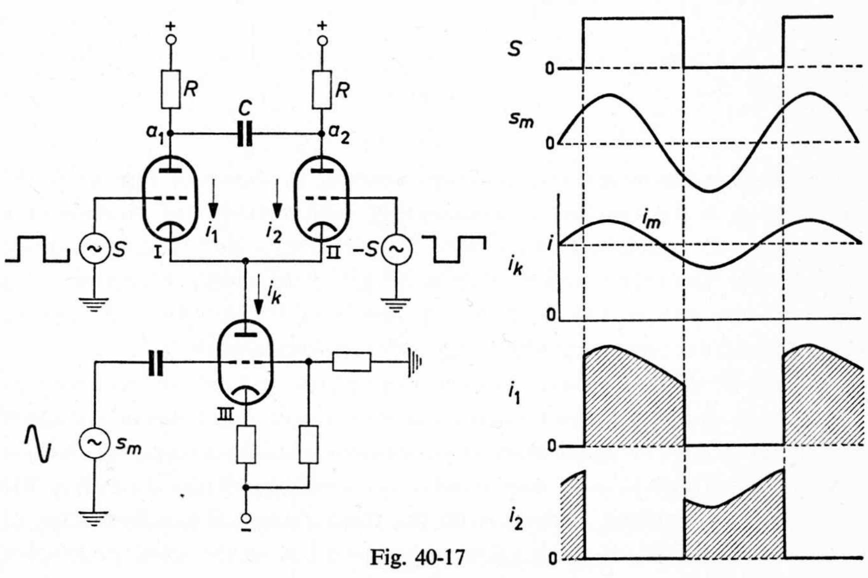

Fig. 40-17

and so large that one of valves I and II always carries the total current i_k. This current is supplied by valve III and consists of a d.c. component i and a component i_m which is proportional to the measurement signal s_m. For the currents passing through the upper valves we now obtain the situation as shown in the right-hand side of Fig. 40-17. It will be easily appreciated (as proved by the following calculation) that the difference between the mean currents $\bar{i}_1$ and $\bar{i}_2$, when S and s are of equal frequency, is proportional to the amplitude of s_m as well as to the cosine of the phase difference between s_m and the fundamental component of S. By inserting equal resistors in anode circuits I and II, and a capacitor between the anodes, the voltage difference between the anodes becomes proportional to what could be called the "progressive mean" of $i_1 - i_2$. At a sufficiently large value of C, an almost direct current will flow through the resistors, which equals the mean value of $i_1 - i_2$, and whereby the contribution of the signal current $\{i_1(\tau) - i_2(\tau)\}dt$ at instant τ will decrease exponentially with time. The time constant for this is 2 RC (Fig. 40-18).

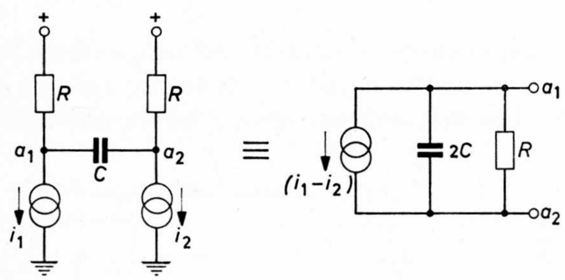

Fig. 40-18

By making the circuit in a balanced version, as shown in Fig. 40-19 the smoothing of the currents is considerably simplified. In the absence of a measurement signal, a current now flows through both common anode circuits for the entire period instead of half this period. Moreover, this method also allows one to diminish considerably the effect of possible deviations from symmetry which can occur in square wave S.

Fig. 40-20 shows a practical design of this type of synchronous detector. Paraphase amplifier valve I provides from a single square-wave voltage S of, say, 20 volts, two equal amplitude, opposite-phased voltages, which drive valves III + V and IV + VI respectively. The control voltages for valves VII and VIII are similarly derived from the measurement signal by means of valve II. The grids of these valves are connected to the same point of a

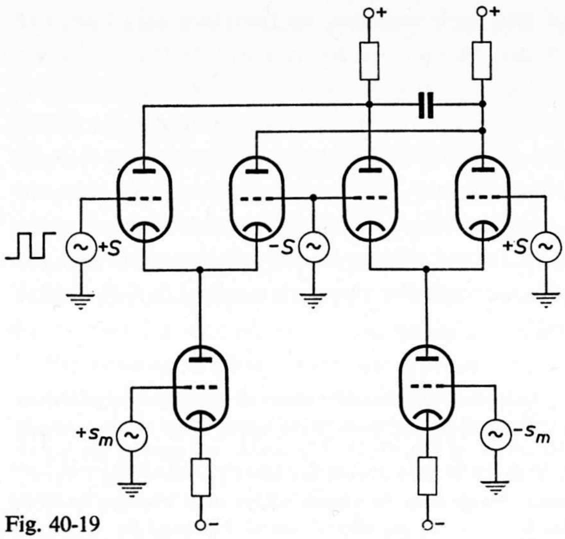

Fig. 40-19

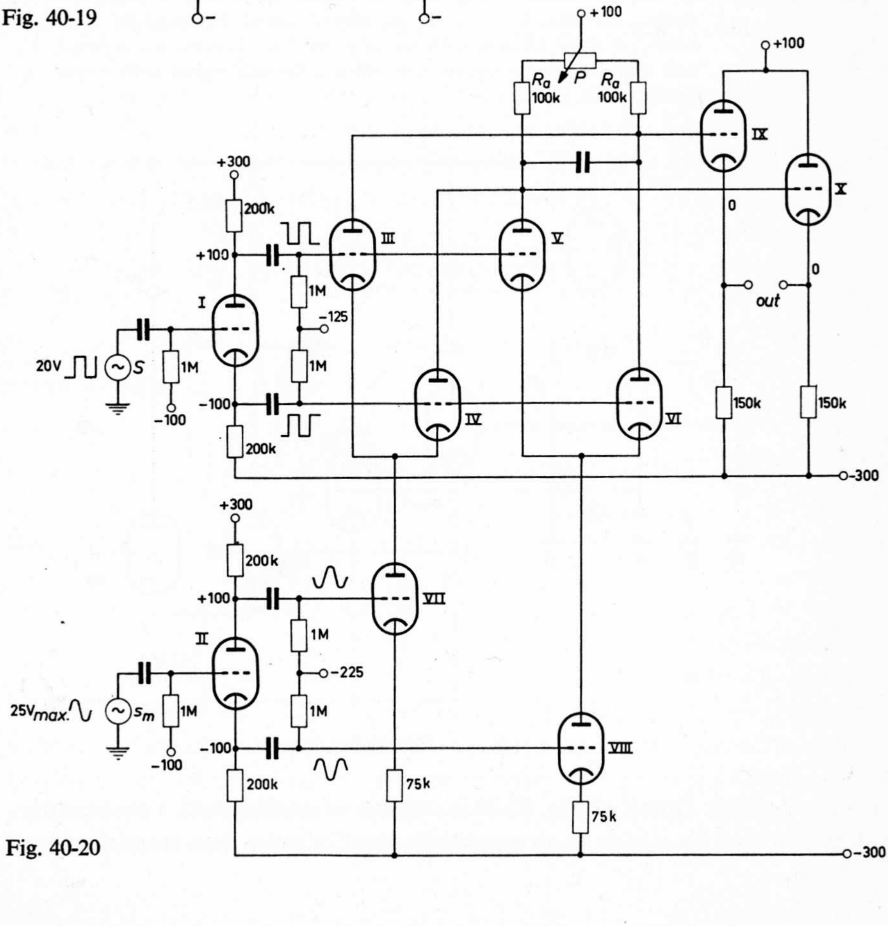

Fig. 40-20

voltage divider by means of grid leak resistors, so that they will have the same d.c. potential. Both signal currents and d.c. currents through both valves will show little difference if their cathode resistances are also made equal. The internal resistance of the output between the anodes of the upper valves is too large for feeding a meter directly. This can, however, be achieved with the two cathode followers IX and X, the grids having been directly connected to these anodes. Any imbalance in d.c. potential levels can be eliminated by potentiometer P. The advantage of this adjustment is that the signal current will always pass through the same resistance $2R_a+P$, sothat it does not affect the sensitivity.

If desired, the balanced output signal between the cathodes of valves IX and X can be changed into an unbalanced signal of the same accuracy by means of the circuit of Fig. 40-21. The anode voltages of the upper valves (III, IV, V and VI) have almost no effect on the signal current, so that the anode voltage may be varied within very wide limits. The output terminal U_1 is kept at almost earth potential by feedback amplifier stage IX and cathode follower XII. The output terminal U_2 will therefore carry almost the entire difference signal with respect to earth.

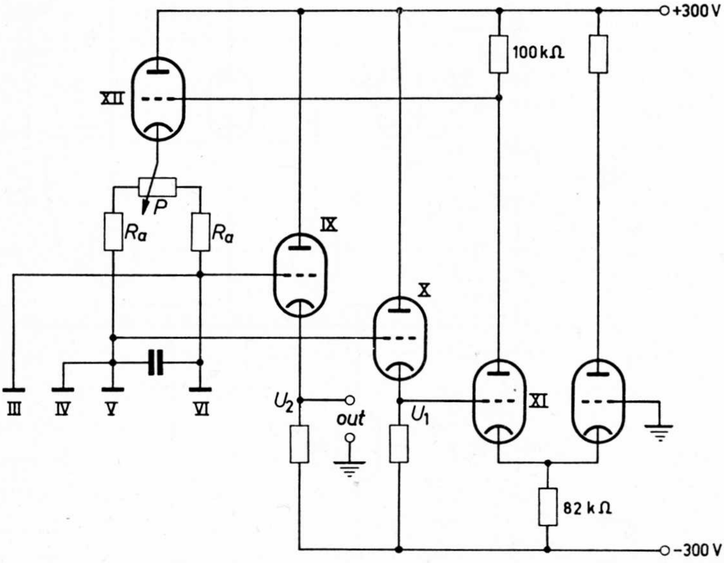

Fig. 40-21

The detector circuit of Fig. 40-20 is capable of working with a particularly high accuracy for signals in an amplitude range of more than three decades.

The deviation from linearity is less than 0.01 per cent of the "full scale" signal up to rather high measurement frequencies ($\approx 10^5$ c/s). Symmetrical design ensures minimum zero error and drift. This circuit can also be made with transistors without significant changes. It is true that a fraction of the signal current does not pass to the collectors of the upper transistors, and variations in this fraction may be a source of error. However, transistors with a high current amplification factor will reduce this trouble and a high accuracy can be reached. The symmetrically designed synchronous detector of Fig. 40-19, using either valves or transistors, satisfies nearly all requirements which may have to be met in the practice of electronic measurements.

Phase sensitivity has the advantage that the effect of possible out-of-phase interference signals is reduced, so that it is possible to attain the highest sensitivity by this method. It is a disadvantage for some measurements that odd harmonics also contribute to the output signal because of the square-wave form of the synchronization signal. This can be avoided (whilst maintaining the same high accuracy) by using an accurate multiplication circuit, as will be discussed in Section 42.

The above is derived mathematically from the following calculation: $s_m(t)$ and $-s_m(t)$ are the measurement signals on the grids of valves VII and VIII in Fig. 40-20; the signal currents in the anode circuits of the upper valves will be expressed by $cs_m(t)U(t)$, where c is a constant and $U(t)$ the "square-wave function" which has a value of $+1$ during one half-cycle and of -1 during the other half-cycle. Choosing a suitable zero point, we can write for $U(t)$:

$$U(t) = c_1(\cos \omega_0 t - \tfrac{1}{3}\cos 3\omega_0 t + \tfrac{1}{5}\cos 5\omega_0 t - \ldots)$$

where $\omega_0 = 2\pi/T$

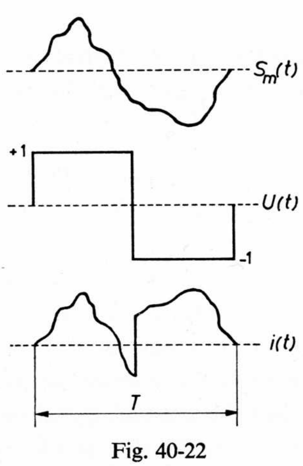

Fig. 40-22

If:
$$s_m(t) = \sum_{\omega} A(\omega)\cos(\omega t + \varphi_\omega)$$

this gives:

$$i(t) = c_2\left[\sum_{\omega} A(\omega)\cos(\omega t + \varphi_\omega)\cos\omega_0 t - \tfrac{1}{3}\sum_{\omega} A(\omega)\cos(\omega t + \varphi_\omega)\cos 3\omega_0 t + \tfrac{1}{5}\sum_{\omega}\ldots + \ldots\right]$$

or:
$$i(t) = \tfrac{1}{2}c_2\Big[\sum_{\omega} A(\omega)\big[\cos\{(\omega+\omega_0)t + \varphi_\omega\} + \cos\{(\omega-\omega_0)t+\varphi_\omega\}\big] - \tfrac{1}{3}\sum_{\omega} A(\omega)\big[\cos\{(\omega+3\omega_0)t+\varphi_\omega\} + \cos\{(\omega-3\omega_0)t+\varphi_\omega\}\big] + + \tfrac{1}{5}\sum_{\omega} A(\omega)[\ldots] + \ldots\Big]$$

The anode impedances have the form $Z_a = R/(1+j\omega\tau)$, where $\tau = 2RC$, which means a rapid drop in Z_a for frequencies above $\omega_g = 1/\tau$ (Fig. 40-23).

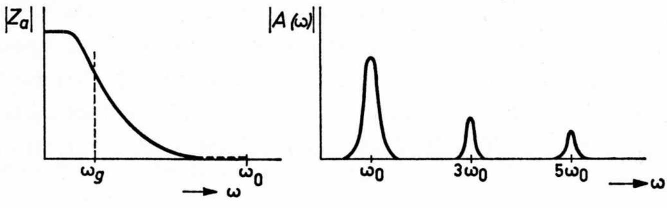

Fig. 40-23

The output voltage will thus almost exclusively contain components of $i(t)$ with frequencies below ω_g. If $\omega_g \ll \omega_0$ and therefore $\tau \gg T$ these will only be the components:

$$A(\omega)\cos\{(\omega-\omega_0)t+\varphi_\omega\},\ \tfrac{1}{3}A(\omega)\cos\{(\omega-3\omega_0)t+\varphi_\omega\},\ \tfrac{1}{5}A(\omega)\cos\{(\omega-5\omega_0)t+\varphi_\omega\},\ldots$$

where $\omega\approx\omega_0$, $\omega\approx3\omega_0$, $\omega\approx5\omega_0, \ldots$

This means that only narrow bands of $s_m(t)$ around frequencies ω_0, $3\omega_0$ $5\omega_0, \ldots$ are passed. The attenuation increases here in proportion to the, central frequency of the band. For signals which have a frequency equal to one of the central frequencies, the attenuation is determined by $\cos\varphi_\omega$. In particular when $\omega=\omega_0$, the output signal will be proportional to $A(\omega_0)\cos\varphi_{\omega_0}$.

By shifting the square wave one quarter of a cycle, we can measure $A(\omega_0) \sin \varphi_{\omega_0}$ instead of $A(\omega_0) \cos \varphi_{\omega_0}$. $A(\omega_0)$ is then obtained by measuring these signals simultaneously and adding them vectorially. In this way one obtains phase-insensitive synchronous detection. A circuit which permits the vectorial addition of two signals is discussed in the next section.

The two components $A(\omega_0) \cos \varphi_{\omega_0}$ and $A(\omega_0) \sin \varphi_{\omega_0}$ can also be used to determine a Nyquist diagram.

We should finally mention that also the bandwidth of a synchronous detector can be restricted more effectively by using more extensive circuits instead of a simple *RC*-filter. For example, an almost rectangular pass-band can be obtained instead of a bell-shaped one.

It is necessary for the accurate measurement of a signal's amplitude that the phase difference between this measurement signal and the synchronization signal is zero, or almost zero. The circuit of Fig. 40-24 can be used over a limited frequency band for changing the phase of a sinusoidal signal, without changing its amplitude. For a "paraphase amplifier" with equal anode and cathode resistors, the anode and cathode signals will be the same size but of opposite phase. If it is possible to neglect the load of the *RC*-circuit, we have, with $v_a = v_k \approx v_i$:

$$v_o = a v_i \frac{1 - j\omega RC}{1 + j\omega RC}$$

with a real and approximately equal to unity.

Any change in R only results in a phase change of the transfer function v_o/v_i.

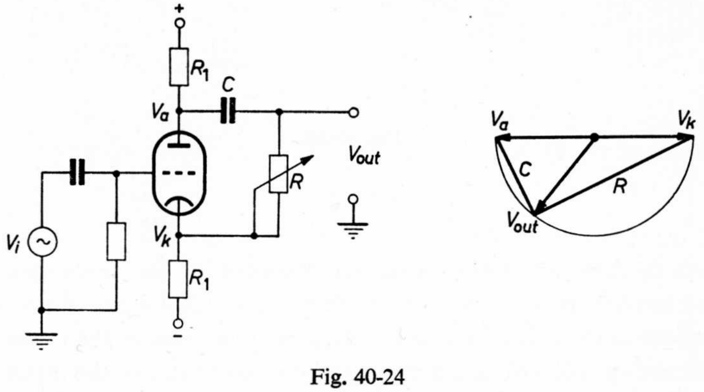

Fig. 40-24

Excellent linearity and small zero error and drift make the detector, designed according to the principle of Fig. 40-19, most suitable for use in measurement electronics. As will be shown by some examples, it is usually easy to obtain the synchronization signal. Even if no such signal is available it is still possible to retain some advantages of this type of detector circuit by using a "quasi-synchronous" detector. The square-wave voltage S which serves as synchronization signal is derived here from the measurement signal by amplifying the latter and subsequently clipping it. As long as the measurement signal is large compared to noise plus interference, the square wave is determined solely by the measurement signal. The same properties of linearity and suppression of interference signals are then obtained as with the synchronous detector. However, this circuit is no longer phase-sensitive: the phase difference between the fundamental of S and the measurement signal is constant and can be made zero. If the signal is of the same order of magnitude as the noise, the zero crossings will no longer be determined solely by the signal but also by the noise, thus increasing the latter's effect. As the signals become still smaller one eventually reaches the state where only the noise is measured. The relation between output voltage and amplitude of the measured signal will thus be as shown in Fig. 40-25. The quasi-

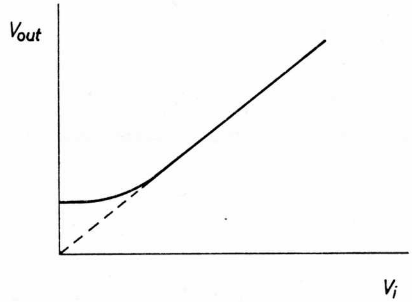

Fig. 40-25

synchronous detection is thus a full-wave version of the diode detector of Fig. 40-1 but has the great advantage of an almost ideal diode characteristic, i.e. excellent linearity and minimum disturbing voltages in the circuit itself. Its applicability is determined by the signal-to-noise ratio of the input signal, as well as by the permissibility of phase-insensitivity.

The importance of phase-sensitivity can clearly be seen when considering the measurement of the amplitude of a signal resulting from the demodulation of an amplitude-modulated signal. As shown in Section 36, the amplitude A

of the carrier wave $A\cos\omega_c t$ is proportional to the measurement signal v_s (Fig. 40-26) and the following two possible equations can apply:

$$v_c = A_0(1 + mv_s)\cos \omega_c t$$

and

$$v_c = A_0 m v_s \cos \omega_c t$$

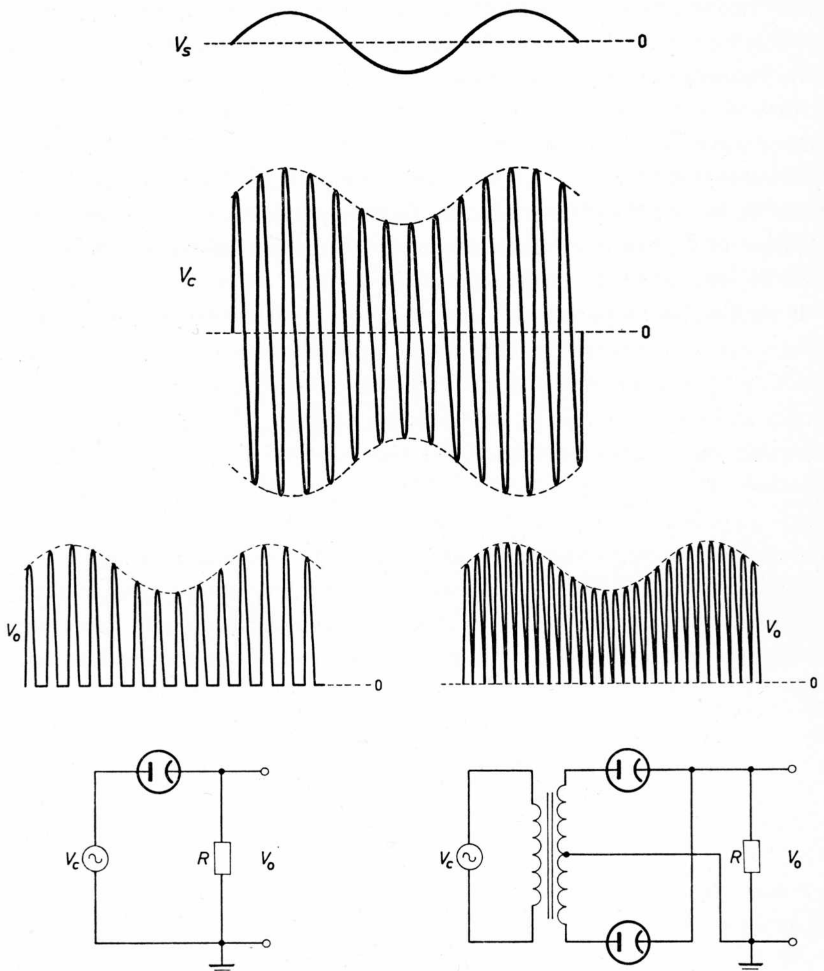

Fig. 40-26

A Fourier component of v_s with frequency ω_s gives rise in both cases to two components in the frequency spectrum of v_c with frequencies $\omega_c \pm \omega_s$. However, in the first case the carrier wave will be present in the absence of a measurement signal, but not so in the second case. If it is required to demodulate these signals, i.e. derive a signal proportional to v_s, the signals must be shifted in the frequency spectrum by an amount ω_c and this can be obtained by multiplying by the square-wave function $U(t)$ with period $T = 2\pi/\omega_c$. However, additional shifts by frequencies $3\omega_c$, $5\omega_c$, etc. will then occur, exactly analogous to what happened in the circuit of Fig. 40-19. In the following we shall assume, where necessary, that in that case, by filtering the modulated signal v_c, only multiplication by the component with frequency ω_c is of importance.

Assuming that the auxiliary signal has a phase angle φ with respect to the carrier wave of v_c, the multiplication factor will be, apart from a constant, $\cos(\omega_c t + \varphi)$. This gives for both cases a component which is indeed proportional to v_s, but the proportionality factor contains $\cos\varphi$. It is therefore a requisite of faithful reproduction that φ is constant, preferably differs little from 0°, and under no circumstances be 90°.

If we have an amplitude-modulated signal governed by the first equation, where $mv_s < 1$, practically automatic multiplication by the square-wave function $\{\frac{1}{2} + \frac{1}{2}U(t)\}$ will occur in the case of half-wave rectification with a diode, as long as $1 - mv_s$ is not too close to zero. The phase is then automatically zero (left-hand side of Fig. 40-26). Full-wave quasi-synchronous detection gives multiplication by exactly $U(t)$, whereby $\varphi = 0$. This also yields a correct result (right-hand side of Fig. 40-26).

A different situation arises when we have to operate on a signal of the second form, where pure multiplication of the measurement signal by the carrier wave occurs (Fig. 40-27). In this case, both methods with "automatic" auxiliary signal will give a wrong result because of the rectification of the

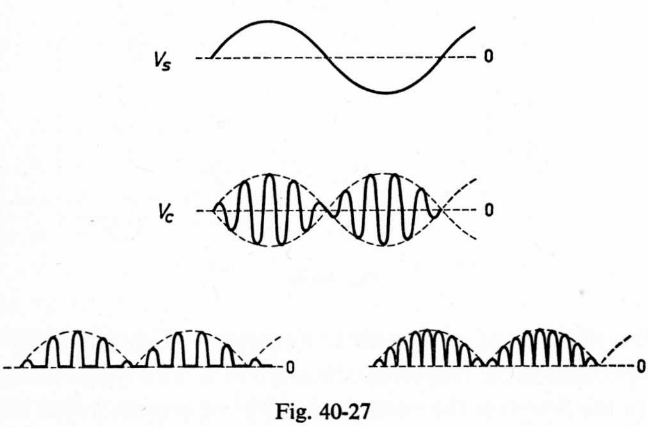

Fig. 40-27

measurement signal. This is usually not permissible, and a separate synchronization signal should be present, the phase of which must preferably equal that of the carrier wave.

In measurement electronics we chiefly meet amplitude-modulated signals of the second type, firstly as a direct result of physical measurements, but also because this method of modulation can be carried out very accurately (see Section 41). Apart from the fact that in electronic measurements it is often necessary to use a separate synchronization signal because of the poor signal-to-noise ratio, obtaining such a synchronization signal does not offer any difficulty.

The a.c. Wheatstone bridge can be mentioned as a simple example of a measurement where phase sensitivity may or may not be important. If we are only interested in the value of the change in resistance, the detection of the a.c. signal does not have to be phase-sensitive. On the other hand, if it is necessary to determine the sign of the change as well, a phase-sensitive detector will be required.

Fig. 40-28 is the block diagram of a measurement arrangement using synchronous detection. The disc has a number of slits and is used for the chopped illumination of a material of which, for example, it is required to determine the conductivity as a function of the wavelength of the incident light. By using a synchronous motor for driving the disc, the frequency of the illumination will be a multiple of the mains frequency and will usually not fluctuate by more than 1 per cent. A constant current I flows through the measurement resistance R so that a change in R is transposed into a change in voltage, which, after possible transformation and pre-amplification, is applied to a selective amplifier. The latter ensures a drastic limitation of the bandwidth. This is usually necessary because otherwise the noise and interference signals could overload the detector. To preclude the effect of fluctuations in the mains frequency on sensitivity, the transmission curve of the selective amplifier

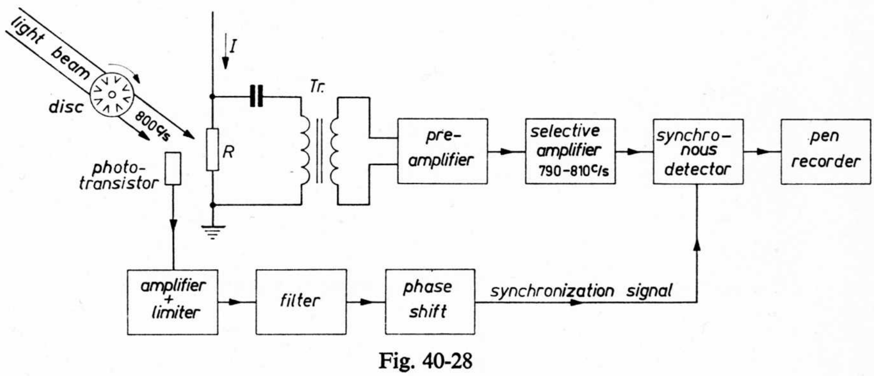

Fig. 40-28

should be flat over the corresponding range of the chopping frequency. This can be achieved by "staggered tuning" (Section 33). The synchronization signal is obtained by also illuminating a photo-transistor or photo-cell with the chopped light. After amplification, filtering and possible phase shift, the sinusoidal voltage is converted into a square-wave voltage which is applied to the synchronous detector.

This type of measurement circuit is very typical, particularly in the 5–10^4 c/s frequency range. With a good design, the sensitivity will be chiefly restricted by the noise of the source resistance. For example, if this source resistance is equivalent to a few ohms, signals of the order of magnitude of nanovolts can be measured using a bandwidth of 1 c/s or less.

Another often occurring measurement with non-sinusoidal periodic signals is the determination of the amplitudes of Fourier components. This can be done in the same way as for the synchronous detector by varying the frequency of the synchronization signal. This type of equipment is called a "wave analyzer".

For most non-sinusoidal signals, the waveform usually contains more important information than the amplitude. Therefore, after linear amplification it is usual to record the waveform on a CRT-tube or some other type of recording instrument (pen or galvanometer type). Quite often one wants to have a close look at a certain part of the signal. This can be achieved over a fairly wide frequency range by using a difference amplifier; the signal is applied to one input, whilst a d.c. voltage at the other input shifts the zero level of the amplifier (Fig. 40-29). This method is usually to be preferred to others because when the amplifier is well designed, overloading does not necessarily cause a recovery time. The upper part of Fig. 40-30 shows the recording of a rectangular voltage with a height of 10 volts, where the peaks

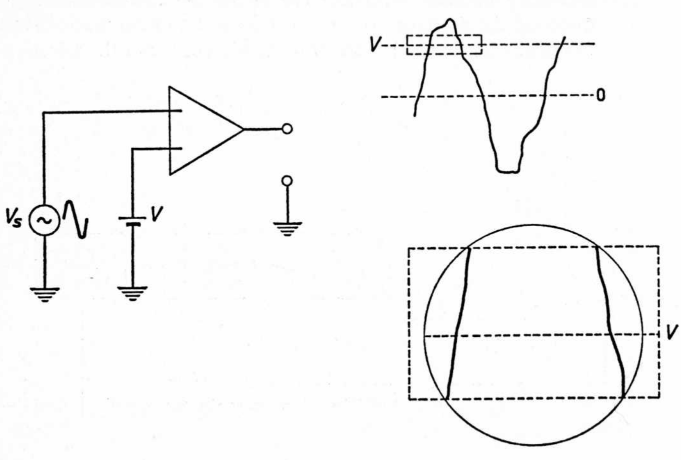

Fig. 40-29

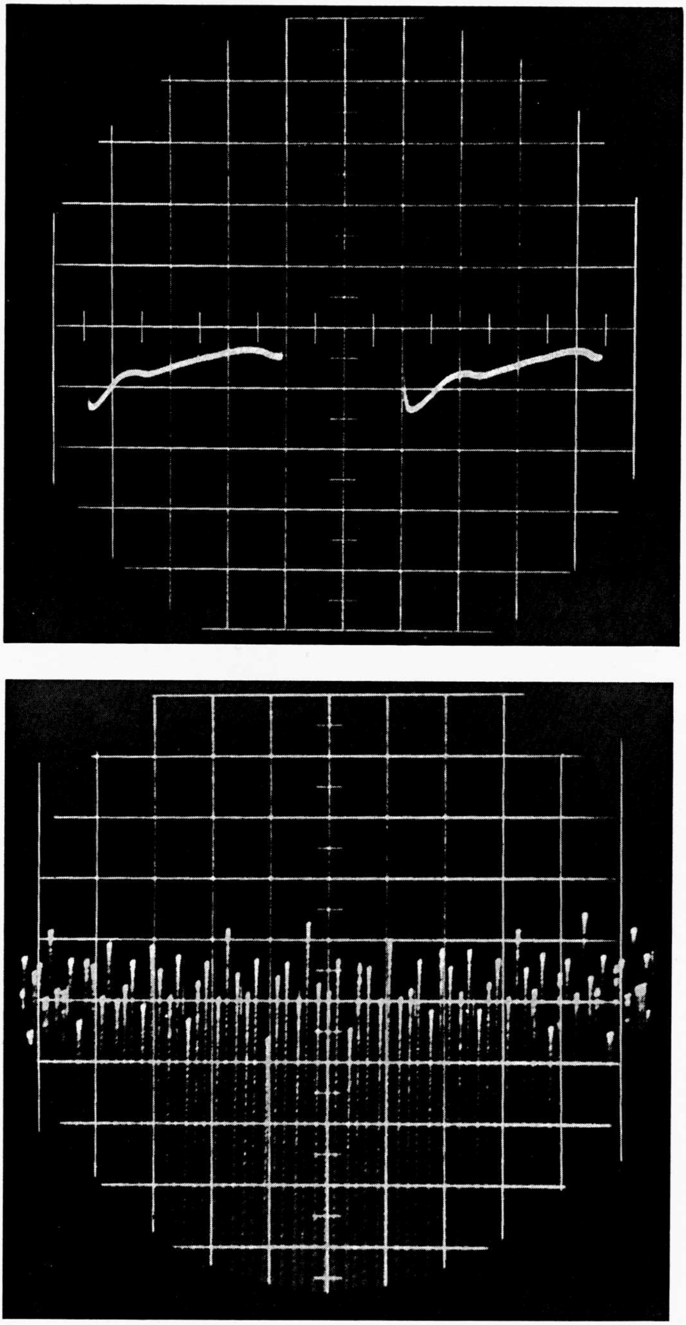

Fig. 40-30

can be observed with a sensitivity of approx. 2 millivolts per cm. This "microscopic" method is obviously also suitable for studying sinusoidal signals, especially the amplitude stability of an oscillator per cycle, lower part of Fig. 40-30, where the peaks of a sinusoidal voltage with an amplitude of 10 volts are recorded with the same sensitivity of 2 mV/cm. It is also clear that this type of compensation method permits the extremely accurate determination of the magnitude of a.c. signals.

With linear systems it is often required to measure the phase difference between two sinusoidal signals having the same frequency. The amplitudes of these signals may show large mutual differences and large changes. A direct, frequency-independent, reading of the phase difference is obviously preferred, and if possible on a linear scale.

Fig. 40-31 shows a basic circuit which achieves this with great accuracy for frequencies up to quite high values ($\approx 10^5$ c/s). By adequate amplification and the use of an instantaneous limiting circuit as discussed in Section 37, p. 323, measurement voltages S_1 and S_2 are converted into square-wave

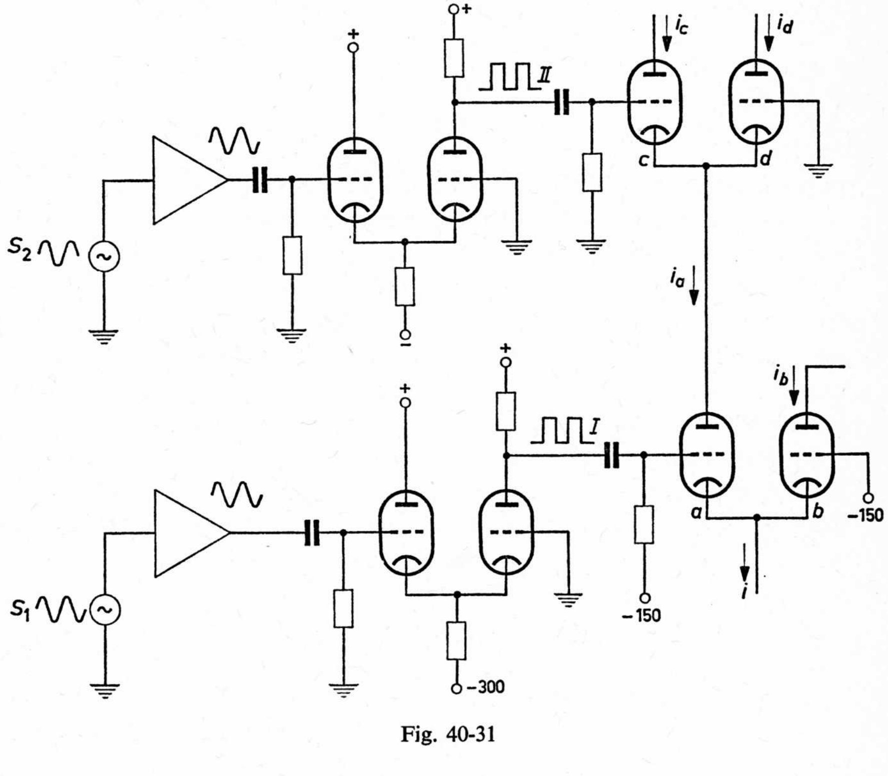

Fig. 40-31

voltages I and II which are mutually shifted by a time $\tau = \varphi/\omega$, where $\omega =$ signal frequency. By making these voltages a few tens of volts, they can be used to drive the current switches *a–b* and *c–d* as mentioned when synchronous detection was discussed. We thus obtain a current flow through the various valves as shown in Fig. 40-32. This figure shows that the width of the current

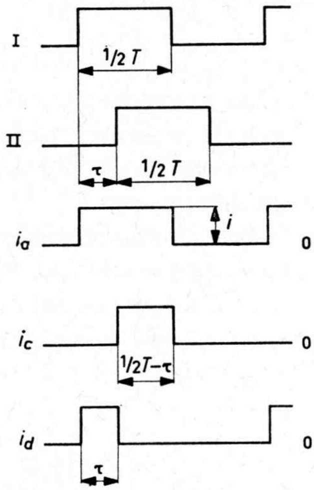

Fig. 40-32

pulses i_d equals the shift τ between I and II, and by giving the amplitude a constant value i, the value of these current pulses, averaged over a cycle, will be proportional to the phase difference φ. Since the width of i_c is half a period minus τ the difference between i_d and i_c can also be used as a measure of φ. In this case a refinement of the circuit is obtained by a symmetrical design (Fig. 40-33) where the mean current in both branches is exactly the same when the phase difference is zero. Averaging is once again achieved by inserting a capacitor between the anodes.

As long as the effect of the edges of the current pulses can be neglected, as is the case with frequencies up to a few tens of kc/s, the accuracy will be particularly great, and it is possible to measure phase differences to within 10^{-4} radians. There is, however, one proviso, namely that both input signals be distortion-free. If this is not so, the zeros and thus the pulsewidths will also be determined by the distortion.

It is sometimes considered a disadvantage of the circuit of Fig. 40-31 that it does not distinguish between φ and $-\varphi$, so that a phase difference of $(360° - \varphi)$ gives the same indication as a phase difference φ. This can be

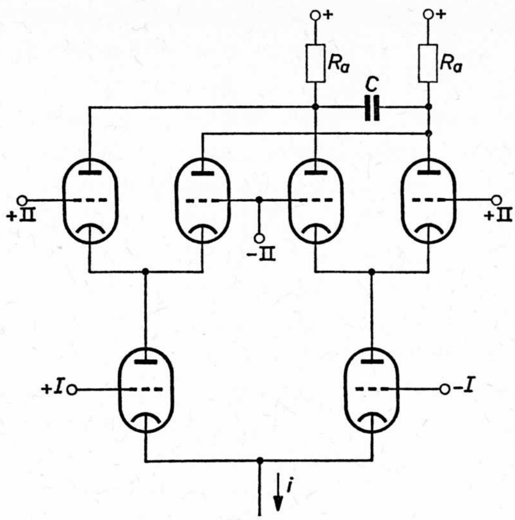

Fig. 40-33

avoided by applying signals I and II first to two flip-flop circuits and using the output signals from these for the actual control. Then fig. 40-34 instead of Fig. 40-32 applies to the various current waveforms in Fig. 40-31.

It is obvious that these circuits can also be designed without any difficulty when using transistors.

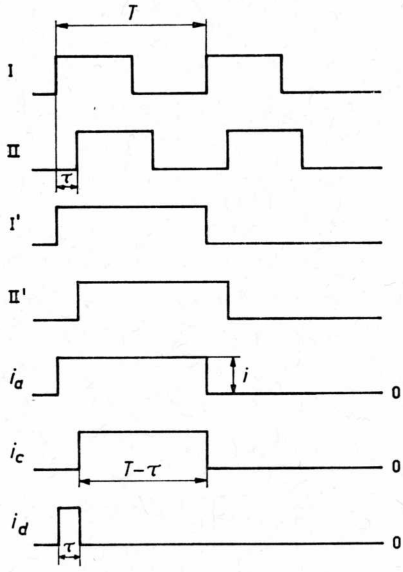

Fig. 40-34

41. Modulation and demodulation circuits

Apart from the discussion of the chopping technique for d.c. amplifiers, we have so far only treated the theoretical side of modulation and demodulation. We shall therefore now discuss practical circuitry. Again we shall not attempt to be complete, but shall devote special attention to accuracy, which is so important in measurement electronics.

It has been a long established custom in radio techniques for obtaining an amplitude-modulated signal to make use of the non-linear relation between the anode current and the control voltage of a valve. Fig. 41-1 shows the principle of these circuits. By a suitable selection of the value of capacitor C and grid leak resistor R, the control grid voltage will equal the sum of the h.f. carrier-wave signal $v_c \cos \omega_c t$ and the l.f. modulation signal $v_m \cos \omega_m t$. As is shown by the following calculation the required modulation is obtained by

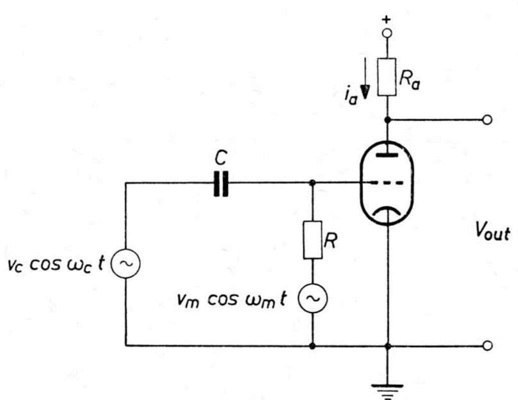

Fig. 41-1

the quadratic term of the power series which defines the anode current as a function of the control voltage. However, the higher-power terms in this series yield undesirable products which cannot be eliminated by filtering. On the other hand, this circuit satisfies the requirement that the carrier wave is still present even in the absence of the modulation signal. This makes it possible to use simple circuits for demodulation as the one shown in Fig. 40-1.

If we take for the valve characteristic the relation:

$$i_a = Sv_s + \beta v_s^2 + \gamma v_s^3 + \ldots$$

where $v_s = v_g + v_a/\mu$, we find for the circuit of Fig. 41-1 (neglecting the effect of the anode voltage on the current): $v_s = v_c \cos \omega_c t + v_m \cos \omega_m t$, so that we can write for the anode current:

$$i_a = \underline{Sv_c \cos \omega_c t} + Sv_m \cos \omega_m t + \frac{\beta v_c^2}{2}(1 + \cos 2\omega_c t) +$$

$$+ \frac{\beta v_m^2}{2}(1 + \cos 2\omega_m t) + \underline{\beta v_c v_m \{\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t\}}$$

$$+ \frac{\gamma v_c^3}{4}(\cos 3\omega_c t + \underline{3 \cos\omega_c t)} + \frac{\gamma v_m^3}{4}(\cos 3\omega_m t + 3 \cos \omega_m t) +$$

$$+ \tfrac{3}{2}\gamma v_c^2 v_m \{\cos \omega_m t + \tfrac{1}{2}\cos(2\omega_c + \omega_m)t + \tfrac{1}{2}\cos(2\omega_c - \omega_m)t\} +$$

$$\underline{+ \tfrac{3}{2}\gamma v_c v_m^2 \{\cos \omega_c t + \tfrac{1}{2}\cos(\omega_c + 2\omega_m)t + \tfrac{1}{2}\cos(\omega_c - 2\omega_m)t\}} + \ldots$$

If ω_c is very much larger than ω_m and if only frequencies are passed in the immediate neighbourhood of ω_c, i.e. ω_c, $\omega_c \pm \omega_m$, $\omega_c \pm 2\omega_m$ etc., only the underlined components will appear in v_{out}:

$$v_{out} = -SR_a v_c \Bigg[\left(1 + \frac{3\gamma v_c^2}{4S} + \frac{3\gamma v_m^2}{2S}\right) \cos \omega_c t +$$

$$\frac{\beta v_m}{S}\{\cos(\omega_c + \omega_m)t + \cos(\omega_c - \omega_m)t\} +$$

$$\frac{3\gamma v_m^2}{4S}\{\cos(\omega_c + 2\omega_m)t + \cos(\omega_c - 2\omega_m)t\} + \ldots\Bigg]$$

which illustrates the introduction and relative effect of the unwanted signals.

With the balanced design of Fig. 41-2, we have for one valve:

$$v_s = v_c \cos \omega_c t + v_m \cos \omega_m t$$

and for the other:

$$v_s = -v_c \cos \omega_c t + v_m \cos \omega_m t$$

whilst the output signal is proportional to $i_1 - i_2$. Although many unwanted terms will be eliminated when the characteristics of both valves are exactly the same, the third term in the above expression will remain.

In radio and television receivers we have a special kind of modulation, or if one wants demodulation, known as "mixing". By this is meant that the carrier-wave frequency of the received signal is shifted to a constant

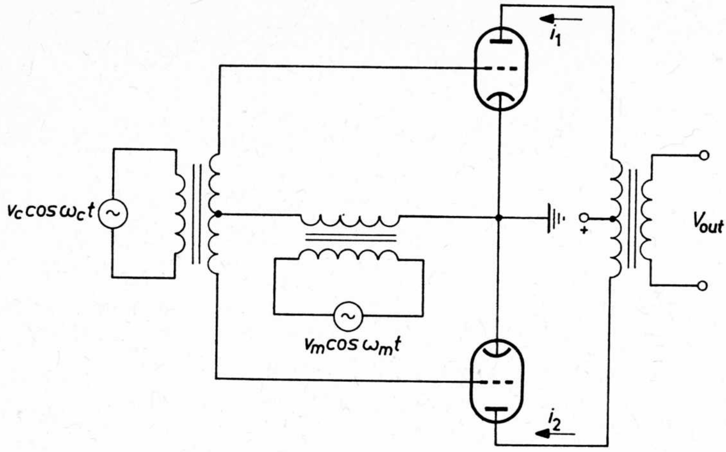

Fig. 41-2

frequency, so that the selective part of the receiver only needs to operate at this frequency. Valves known as "mixers" are used for this purpose, i.e. multi-grid valves where the auxiliary signal and the received signal are applied to different grids. Once again, the operation is based on the non-linear behaviour of a valve's characteristic and is subjected to the same restrictions. As in this case there is usually not a difference of several magnitudes between the carrier and modulation frequencies, various other unwanted products will occur.

Greater accuracy is theoretically possible when the non-linear characteristic of the modulating element is better defined than a valve's characteristic. Transistors are better in this respect, but a far greater gain in accuracy can be obtained by using valves or transistors as switches.

We have considered the possibilities and shortcomings of the various devices for their use as switches when discussing d.c. amplifiers. Some of the difficulties mentioned there will disappear when only a.c. voltages are used for modulation, especially when the signals are large. For example, the use of diodes is illustrated in the circuit of Fig. 41-3, where we assume that the diodes behave as ideal switches and the square-wave voltages are large with respect to the l.f. modulation signal. With the balanced design of Fig. 41-4, known as a "ring modulator", a pure multiplication of the two signals and therefore complete suppression of the carrier-wave signal occurs in principle. This type of circuit is often used in telecommunications.

The use of transformers should be avoided wherever possible in measurement electronics. A modulation circuit with valves, which satisfies this

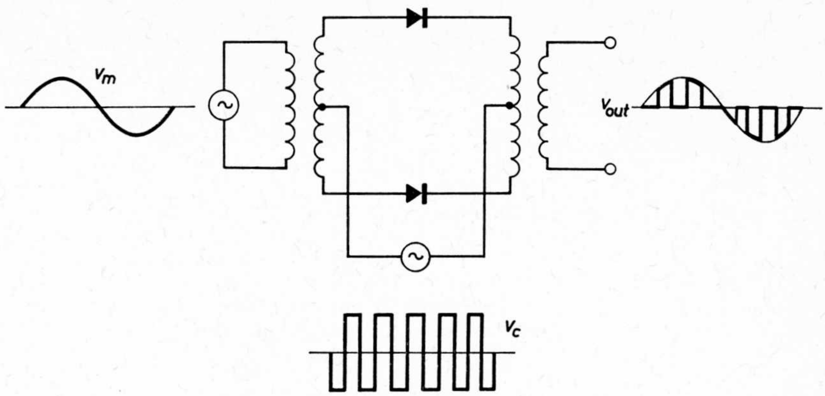

Fig. 41-3

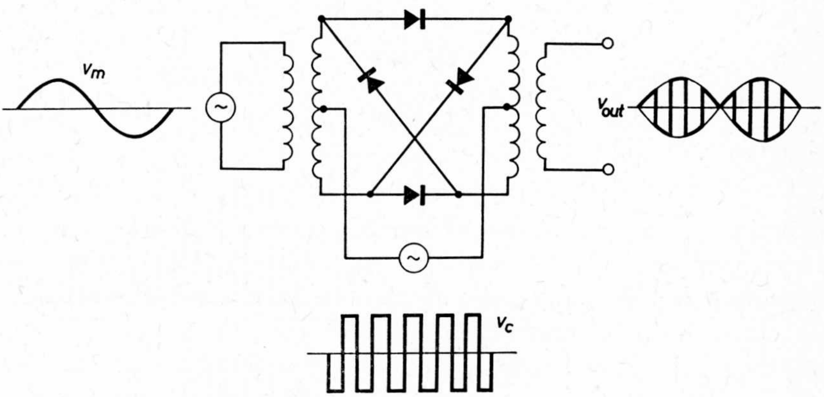

Fig. 41-4

requirement and also gives a very great accuracy, is shown in Fig. 41-5. Its operation shows great similarity with the detection circuit of Fig. 40-17. The upper valves are made to conduct in turn at the same frequency as that of the carrier-wave. Currents i_1 and i_2, passing through these valves, are varied with the modulation signal by means of a linear current source. These currents, as well as their difference, will appear as indicated in Fig. 41-6. By inserting a parallel resonant circuit between the anodes, which is tuned to the carrier frequency, the voltage $v_{a1}-v_{a2}$ will exclusively contain frequencies in a band around this central frequency. In this way, a signal which is accurately modulated in amplitude is obtained without suppression of the carrier wave. By also applying the modulation signal v_m in balance

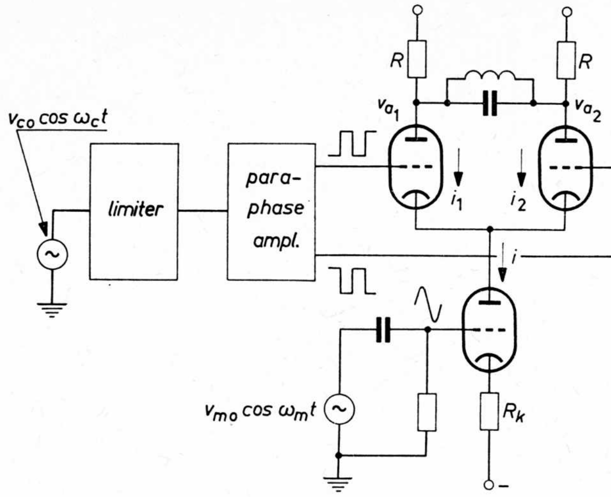

Fig. 41-5

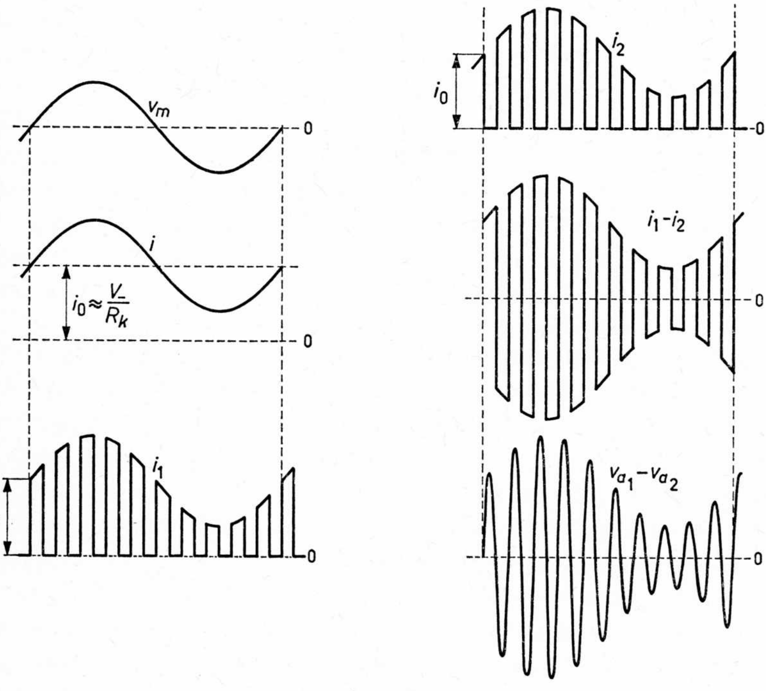

Fig. 41-6

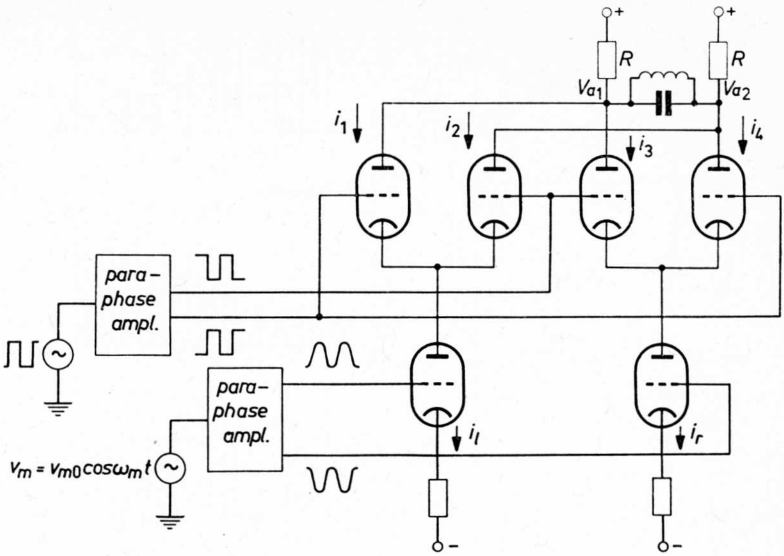

Fig. 41-7

(Fig. 41-7) we obtain a signal with carrier-wave suppression, as shown by the current waveforms of Fig. 41-8.

The shortcomings of the valve when used as a switch are compensated in these circuits by modulation with a current source. When transistors are used, in addition to the circuits corresponding to Figs 41-5 and -7, we also have the possibility of modulation by voltage control. Fig. 41-9 gives an example of a circuit without carrier-wave suppression. The modulation voltage supplied by emitter follower III is connected alternately to *A* and *B* by means of transistor switches I and II. If we make the base currents of these transistors sufficiently large in the conducting state, the collector-emitter voltages can be restricted to less than 0.1 volt. The behaviour of the voltages at points *A* and *B* is illustrated by i_1 and i_2 in Fig. 41-6 and the difference voltage v_{AB} corresponds to $i_1 - i_2$. Fig. 41-10 shows the balanced design, where the carrier wave is suppressed. To ensure a high degree of symmetry, the modulation signals are equalized by potentiometer P_1 and the d.c. voltage reference levels by P_2. The voltage waveforms correspond in this case to the currents in Fig. 41-8.

With regard to modulation circuits we should note in the first place that the amplitude of the modulated signal depends very little on the difference in the width of the positive and negative halves of v_c. It follows from Fourier

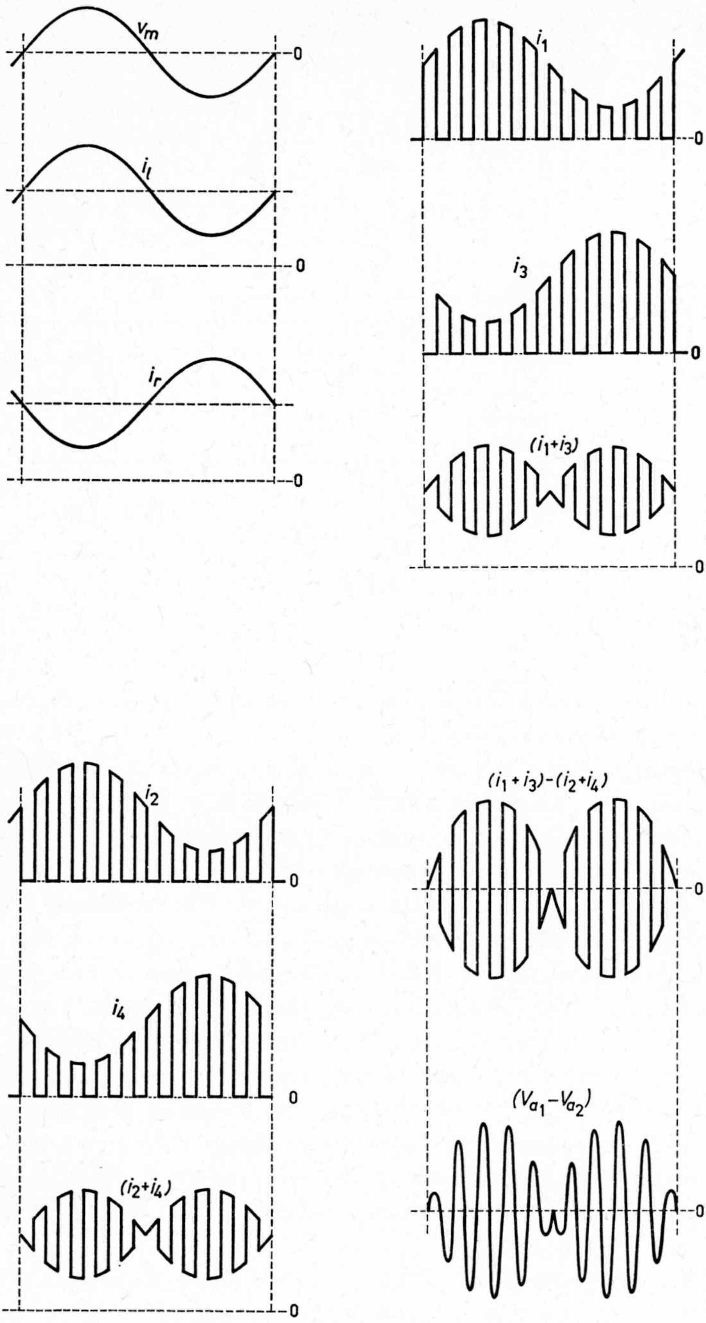

Fig. 41-8

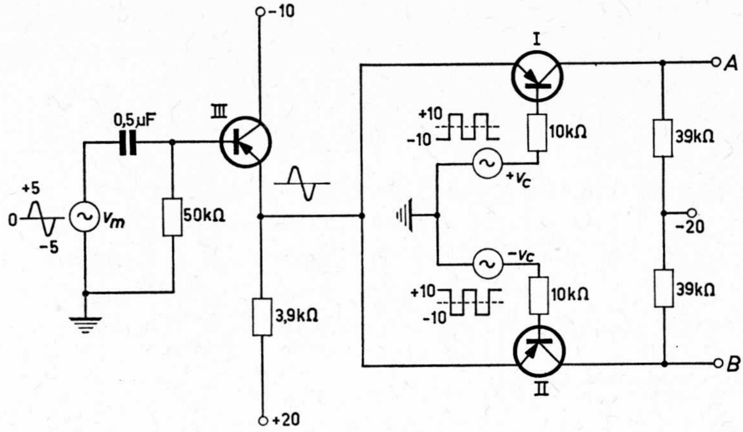

Fig. 41-9

analysis that just at the zero-crossing the contribution to the integrand which determines the amplitude of the fundamental, is very small. It is obvious that the circuits discussed above are suitable for amplitude modulation, but not for multiplying two sinusoidal voltages. The amplitude of the output signal is namely only proportional to the amplitude of the mo-

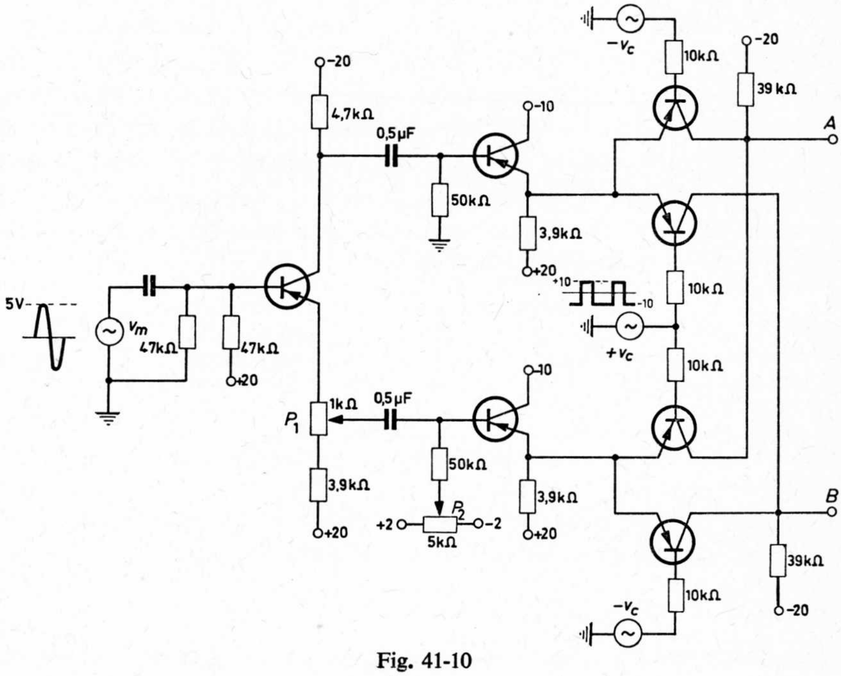

Fig. 41-10

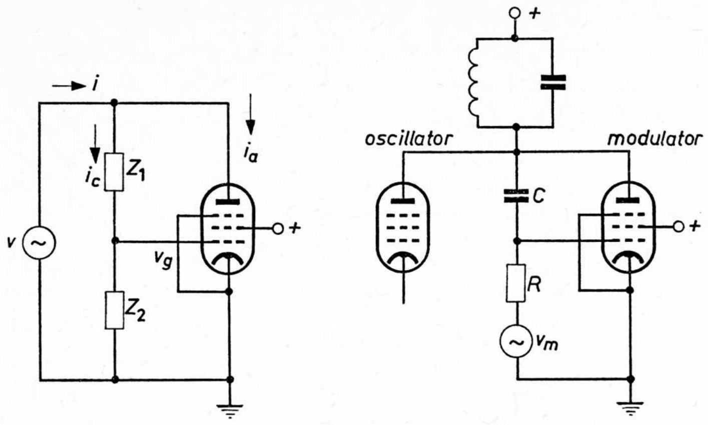

Fig. 41-11

dulation signal and independent of that of the carrier wave. In the next section we shall discuss a circuit which allows one to obtain the product of two voltages with great accuracy, whether they are sinusoidal or not.

Demodulation of amplitude-modulated signals is identical to the measurement of amplitudes, which was discussed in the previous section.

Radio technique makes use of circuits containing a "reactance valve" for obtaining a signal which is frequency-modulated (left-hand side of Fig. 41-11). This type of circuit enables one to produce an apparent impedance, the value of which varies with the modulation signal in an approximately linear manner. By making this impedance capacitative or inductive and effectively connecting it parallel to an *LC*-circuit, it is possible to vary the resonant frequency of this circuit and hence also make the frequency of an oscillator vary with the modulation signal.

For

$$i_a = Sv_g, \quad v_g = \frac{Z_2}{Z_1 + Z_2} v, \quad i_c = \frac{v}{Z_1 + Z_2}$$

we have

$$i = i_a + i_c = \frac{v}{Z_1 + Z_2} (1 + SZ_2)$$

When we make Z_1 a capacitor and Z_2 a resistor of such value that $1/\omega C \gg R$ and $SR \gg 1$, the above expression can be approximated by:

$$i = \mathrm{j}\omega CSRv$$

which means that the circuit has an apparent capacitance of SRC. If we use a valve that preferrably has a quadratic $I_a - V_g$ characteristic, so that S changes linearly with V_g, we can vary the operating conditions of the valve in accordance with the modulation signal so that a varying capacitance is obtained. Fig. 41-11, right-hand side, gives the principle of a practical circuit for doing this.

This type of circuit has some disadvantages which generally prohibit its use in measurement electronics:

1. The frequency variations depend on the valve or transistor characteristic, and are therefore not very well defined;
2. The maximum relative frequency variation is of the order of a few per cent;
3. The apparent impedance is not purely reactive, but has a real part that varies with S as well. The damping of the resonant circuit is thus also varied, and hence the amplitude of the oscillation.

Frequency modulation is equally possible with inductors of which the inductance can be varied by changing the permeability of the core material, varactors (capacitors where the capacitance is a function of the applied voltage) and condensor microphones. However, similar disadvantages are inherent in these methods. The first and third disadvantage can be remedied by using feedback and amplitude limitation, but the disadvantage of the small maximum frequency variation remains.

To keep the amplitude of an f.m. signal, obtained with the circuit of Fig. 41-11, constant, we usually utilize a diode limiter (left-hand side of Fig. 41-12). We have seen in Section 40 that the load of such a circuit is about $\frac{1}{2}R$ at constant amplitude. In the equilibrium state the energy corresponding with this, as well as the losses in the resonant circuit, are provided by the oscillator valve. If the amplitude of the oscillator increases, the charge injected into capacitor C will cause an instantaneous additional load on the resonant circuit, which will tend to counteract the increase in amplitude. When the amplitude decreases, the diode will conduct over a shorter time and the resonant circuit will be less damped, with the result that this change is also counterbalanced.

The amount of stabilization resulting in this way is often insufficient for measurement purposes. We can then use the method proposed by Mayo and Head (right-hand side of Fig. 41-11). This method consists of inserting in series with the diode a parallel resonant circuit, which is tuned to the third harmonic of the oscillator frequency. This circuit can therefore only be used when the relative frequency deviation is small. With a normal diode circuit, the change in damping at a certain change in amplitude will be relatively small because the diode will only conduct for a small fraction of the cycle. Because the diode current consists of small sharp pulses, the amplitude of the third harmonic is quite large. By now inserting a resonant circuit which has a high resistance just at this harmonic, this current component will be much smaller and even be reduced to zero under ideal conditions. However, this implies that diode current i_d must change its waveform so that this

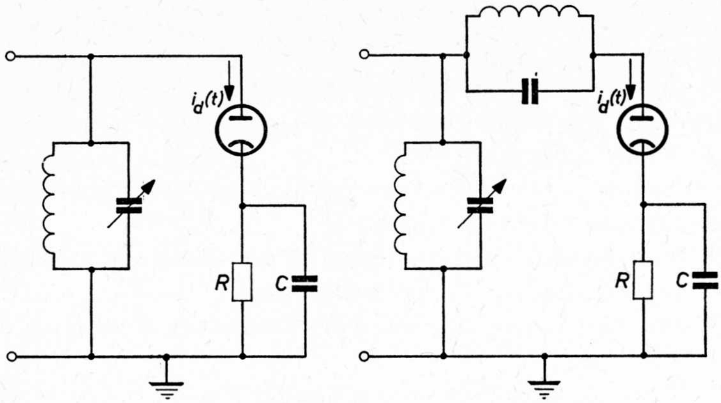

Fig. 41-12

requirement is satisfied. The amplitude of the third harmonic in $i_d(t)$ is proportional to the integral

$$\int_0^T i_d(t) \cos 3\frac{2\pi t}{T} \mathrm{d}t$$

where T is the "instantaneous" cycle of the signal. It follows from Fig. 41-13 that, if this integral is to be reduced to zero or a very small value, it will be necessary for the diode current to continue to flow over a much larger fraction of the cycle, i.e. almost a third of a cycle. In other words, the use of the resonant circuit adds so much third harmonic to the sinusoidal voltage that the voltage across the diode circuit has become almost rectangular. Any change in the amplitude will cause a much greater change in damping than that caused by the original diode circuit. The improvement in amplitude stability thus achieved is therefore very great, namely 100 times or more.

The circuit discussed in Section 39 for generating a sawtooth voltage waveform (Fig. 39-27) can also be used to generate a very accurate f.m. signal, which can be varied over a wide frequency range (Fig. 41-14). Here use is made of a Schmitt trigger (I-II). Dependent on the state of this trigger, the capacitance C will be charged with current i_1 or discharged with difference current $i_2 - i_1$, where $i_2 > i_1$. Period T is determined by the values of these currents, the value of capacitance C and the accurately fixed difference between the stable levels of the trigger circuit:

$$T_{\text{charge}} = \frac{C\Delta V}{i_1}, \quad T_{\text{disch}} = \frac{C\Delta V}{i_2 - i_1}. \text{ Therefore } T = C\Delta V\left(\frac{1}{i_1} + \frac{1}{i_2 - i_1}\right)$$

and the frequency is thus varied by variation of one or both currents with a signal.

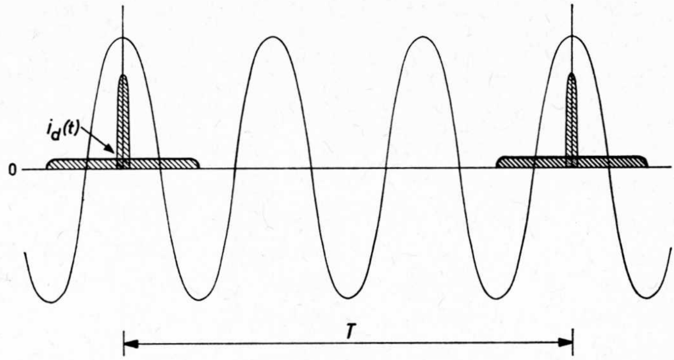

Fig. 41-13

In order that filtering can easily produce a sinusoidal voltage it is advantageous to make the charging and discharging times of equal length, which produces an equilateral triangular waveform. This is the case in Fig. 41-14 when $i_2 = 2i_1$, which requires adjustment and is also difficult to maintain if

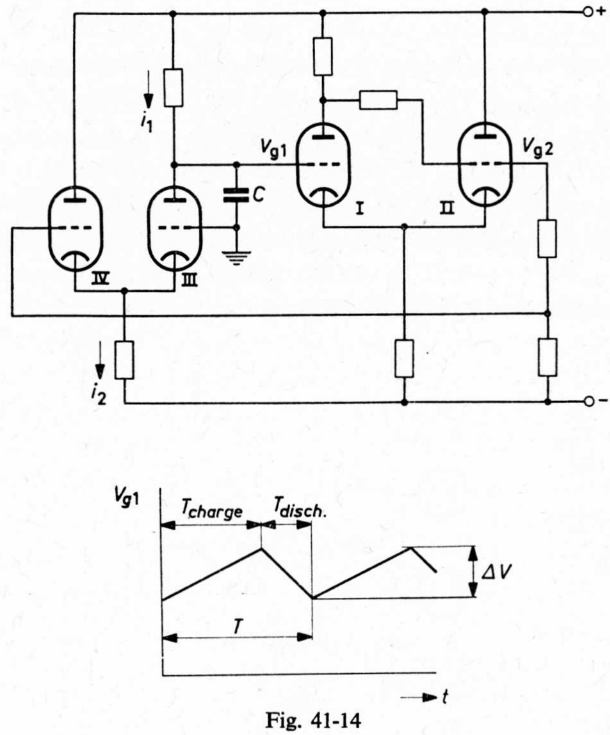

Fig. 41-14

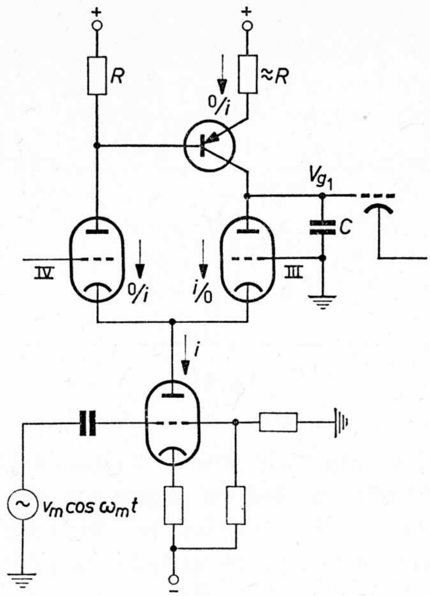

Fig. 41-15

both must vary with the signal. The circuit of Fig. 41-15 is much more accurate for this purpose. This allows for making the charging and discharging currents equal, while both can be varied linearly with the modulating signal. When valve III conducts, the discharging current will equal *i*. Because no current passes through valve IV, the base potential of the transistor will be V_+: the transistor will pass no current and the charging current of the capacitor will be zero. When valve III is cut off, current *i* will pass through valve IV, and the base potential will be *iR* volts negative with respect to V_+. If we now

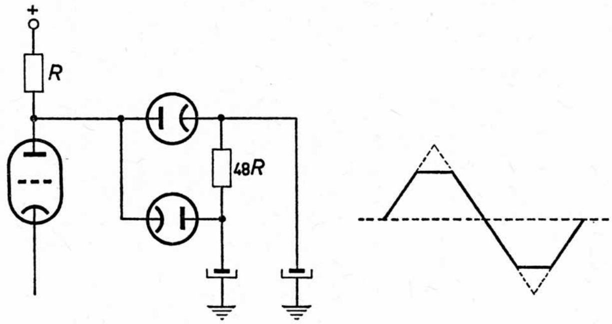

Fig. 41-16

make the emitter resistance of the transistor approximately equal to R, the collector current can be made exactly equal to i and the capacitor will then be charged with this current i. The period is in this case $2C\Delta V/i$, so that the frequency of the triangular voltage is proportional to i. Variation of i with modulation signal $v_m \cos\omega_m t$ can be carried out in this manner. As mentioned before, this circuit allows a very large change in relative frequency. For further improvements of this circuit the reader is referred to the literature mentioned in Section 44.

To convert a triangular voltage with equal charge and discharge times into a sinusoidal voltage it is necessary to suppress the higher harmonics; these are: third 11%, fifth 4%. Since the frequency may vary over many decades with this circuit it is not possible to use resonant circuits. The third harmonic can be fully suppressed by adequate design of the limiting circuit of Fig. 37-14. This is the case if the triangular voltage is clipped by 1/6 of the amplitude in both directions. Fig. 41-16 shows how this would be achieved with ideal diodes. If resistors are also inserted in series with the diodes, the cut-off will be less abrupt and a good approximation of the sinusoidal form will be achieved. If so desired, the effect of the other harmonics can be reduced by using a number of these circuits.

For the demodulation of an f.m. signal with limited frequency deviation, we think in the first place of an inductor. When we have a current $i(t)$ which is proportional to an f.m. signal, we can write:

$$i(t) = i_0 \sin \int_0^t \omega_{\text{inst}}\, dt$$

This will produce a voltage across an inductor L (as we have seen in Section 36):

$$v(t) = L\frac{di(t)}{dt} = Li_0\, \omega_{\text{inst}} \cos \int_0^t \omega_{\text{inst}}\, dt$$

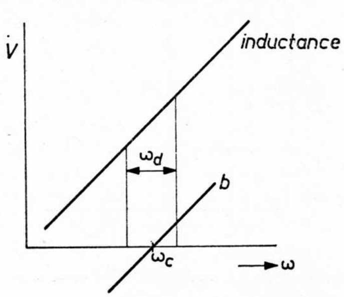

Fig. 41-17

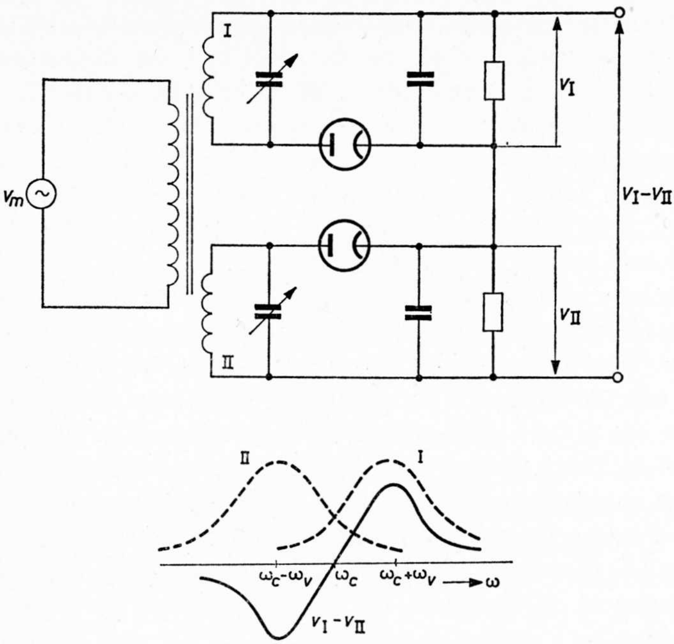

Fig. 41-18

This voltage, after rectification, will produce a voltage proportional to the amplitude and therefore also to ω_{inst}. The disadvantage of this method is that at a frequency deviation ω_d which is small compared to the central frequency ω_c, the required signal will also be small with respect to the zero signal (Fig. 41-17). An impedance with the shape of curve *b* is therefore much more suitable for the detection of an f.m. signal. This can be obtained by means of the circuit of Fig. 41-18. Resonant circuits I and II are tuned

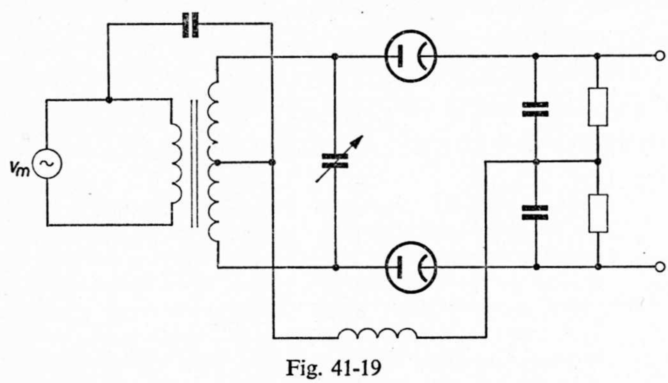

Fig. 41-19

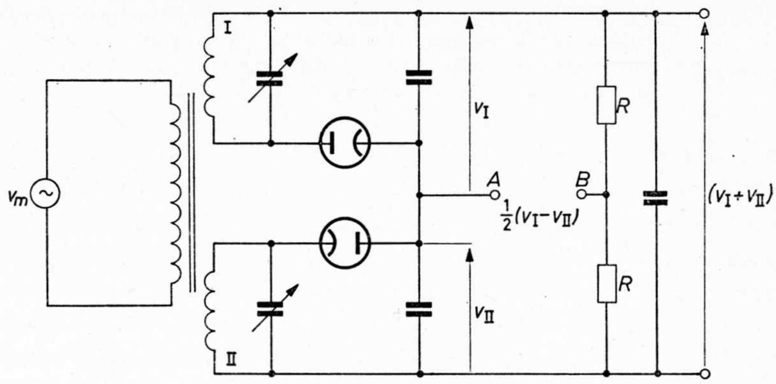

Fig. 41-20

to frequencies $\omega_c + \omega_v$ and $\omega_c - \omega_v$ respectively, where ω_v is of the same order of magnitude as ω_d, the frequency deviation. The difference between the rectified signals v_{I} and v_{II} has a frequency dependence as shown in the lower part of Fig. 41-18.

The "Foster-Seely" discriminator (Fig. 41-19) is a variation of the above circuit, where curves I and II are obtained with only one resonant circuit by providing additional coupling with the primary winding.

If one of the diodes in the circuit of Fig. 41-18 is reversed (Fig. 41-20), the output voltage will be $v_{\mathrm{I}} + v_{\mathrm{II}}$. This sum has the same constant value at a constant amplitude of the f.m. signal, for all frequencies in the range under consideration. By inserting a large capacitor across the output (analogous with what proved to be valid on the left-hand side of Fig. 41-12) the damping of the resonant circuits will be increased with increasing amplitude of the f.m. signal and the circuit will therefore stabilize the amplitude. The required difference voltage $\frac{1}{2}(v_{\mathrm{I}} - v_{\mathrm{II}})$ will be found between points *A* and *B*. This circuit, called "ratio detector" thus combines detection of the f.m. signal with amplitude limitation. As one of the diodes is reversed the resistors across the capacitors in Fig. 41-18 can be omitted and the diodes now carry the charging and discharging currents at point A.

We should note here that the sensitivity of this circuit is greater than would appear from a calculation based on the assumption that the two rectifying circuits operate independently. The mean d.c. current passing through the two diodes must be of equal value. Since with short current pulses, the ratio of d.c. to a.c. current components is constant (1:2), the a.c. current components of the diode currents must also be the same. However, this means that the resonant circuit which produces the largest of the two signals v_{I} and v_{II} is relatively less damped, so that the difference between the two signals is larger than when the requirement for the diode current would not apply. Due to the d.c. current coupling we have thus also a coupling between a.c. current components, which, theoretically, may even have different fundamental frequencies.

Detection of an f.m. signal can also be carried out very accurately with the phasemeter we mentioned in the previous section. If a signal current is directed through a parallel resonant circuit which is tuned to the central frequency ω_c, there will be a phase difference between this current and the voltage across the resonant circuit which, for frequencies in the proximity of ω_c, is proportional to the difference frequency $\omega-\omega_c$. The phasemeter converts this difference to a proportional voltage. This voltage is obtained by converting the sinusoidal waveforms to rectangular waveforms and measuring the overlap of these voltages. For this purpose, special multi-grid valves have been designed, where the anode current depends in such a way on the potentials applied to two grids that it will measure the overlap and consequently also the phase. However, the accuracy is insufficient for most measurement purposes.

Finally, a voltage which is proportional to the signal frequency can also be obtained by charging a capacitor with a constant charge per cycle. If the capacitor can discharge through a resistor, the voltage across the capacitor is proportional to the "progressive mean" of the frequency. The RC-product determines the time of averaging. The monostable trigger as discussed in Section 39 will give a constant charge per cycle if the pulsewidth is shorter than the smallest cycle of the f.m. signal, and if the circuit is triggered,

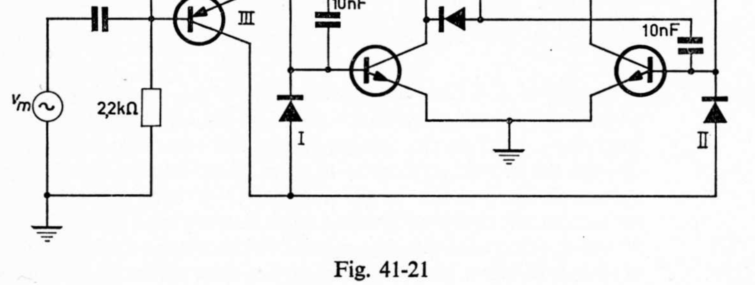

Fig. 41-21

for example, when the waveform crosses the reference level in the positive direction.

Apart from amplitude and frequency modulation, other methods are applied in electronic measurements, such as pulsewidth and pulseheight modulation. For example, the circuit of Fig. 41-14 can also be used for pulsewidth modulation, i.e. the width of the pulses initiated by the Schmitt trigger equals the period of the triangular voltage, and can obviously be made proportional to the modulating signal.

Fig. 41-21 gives an example of a simple pulsewidth modulator. Here the diodes I and II of the multivibrator circuit enable the speed at which the base voltage of the non-conducting transistor changes, to be determined by the current produced by current source III. Current source III can be controlled by the modulation signal.

We shall refer to these methods in the next section when discussing various operational circuits.

42. Mathematical operations

In simple measurements a signal, after amplification and possible modulation and detection, is often displayed directly on a pen recorder, oscilloscope or simple meter. However, with more complicated measuring techniques it is often necessary to subject the signal to one or more operations which could be called mathematical. For example, if it is known that a certain physical measurement leads to a voltage that is to a first approximation inversely proportional to the temperature, we may wish to know what deviations from this relation occur over a specific temperature range. It is obviously possible to plot the measured voltage directly against the temperature and to determine any deviations from the hyperbolic relation. The measurement will become much more elegant, however, by deriving a voltage which is the reciprocal of the measured voltage, and plotting the former against temperature. For this we need a circuit where the output is inversely proportional to the input.

Since we are able to begin with a sufficiently large signal with these operational circuits there is no need to be restricted by inherent limitations as discussed in the previous sections. It is not surprising therefore that the flexibility and remarkable capabilities of electronics are most clearly shown in the field of these operational circuits. It is of course not possible to discuss exhaustively the many circuits which exist today, but it is possible and indeed sensible to see how far the same basic principles underlie them. We shall also discuss some of the most important circuits in more detail. Regarding the latter, we shall devote special attention to some lesser known methods which allow one to achieve a high degree of accuracy from relatively simple circuits.

Amongst the most important methods is primarily the digital method, of which we can state here that it allows one an almost unlimited number of operations, which are all based on the same fundamental precept: the "yes-no" or "0-1" decision. The flip-flop can be used for this as "building block" and the decision can be carried out with theoretically unlimited accuracy. However, it is necessary that the input signal is coded in digital form, and the relation between input and output signal is then only apparent for discrete values. The largest area of application of the digital computer will doubtlessly be in accounting but also industrial processes have a need for the great possibilities of programmed control and evaluation and here again digital techniques will be increasingly preferred. There are, however, many relatively simple operations in industrial plants and one sometimes gets the impression that fashion dictates the use of digital circuitry for these operations as well. The truth is that simpler and cheaper techniques can be applied to advantage in many cases.

Of these latter techniques it is especially the analogue computing technique which is making an important contribution to processing circuitry because of its great flexibility in applications involving linear processes. As will be explained, this method which uses the "operational" amplifier as building block, is particularly direct for most linear processes, and extremely cheap when compared to digital techniques. However, the solutions obtained with this method often show a certain lack of elegance, especially in the case of non-linear processes, implying that the solution arrived at is not the best possible one. Optimum results are however indispensable in pure measurement techniques as used in laboratories and process control. It makes sense therefore to attempt to find the best possible solution for a specific operation.

As there is much literature on circuitry for either of the building block methods, we shall restrict ourselves to a brief discussion of these techniques, and pay special attention to some of the other methods.

For the linear operation on signals, i.e. addition, subtraction, integration and differentiation, we first consider the method which utilizes operational amplifiers. This method is consistently followed in analogue computers. It is based on the fact that the relation between input and output can be solely determined by passive components if we apply strong feedback. Fig. 42-1 shows the basic circuit. A is basically an amplifier with infinite negative amplification over the entire frequency band. With feedback, the input voltage v_i must be zero. The relation between input and output signals is then:

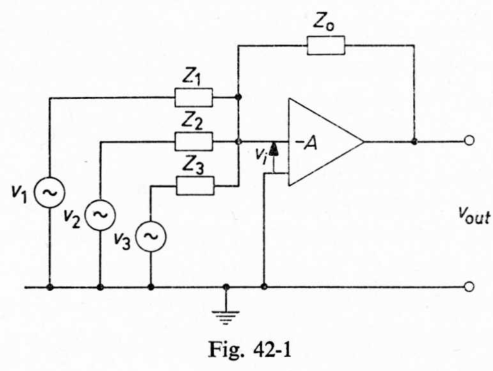

Fig. 42-1

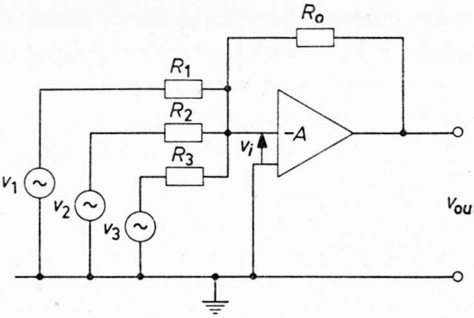

Fig. 42-2

$$-\frac{v_{\text{out}}}{Z_o} = \frac{v_1}{Z_1} + \frac{v_2}{Z_2} + \ldots + \frac{v_n}{Z_n}$$

Replacing all impedances by resistances (Fig. 42-2):

$$v_{\text{out}} = \alpha v_1 + \beta v_2 + \ldots + \eta v_n$$

where $\alpha, \beta \ldots \eta$ are negative numbers.
Replacing Z_o by a capacitor with a value C, Z_1 by resistor R and the remaining impedances infinite (Fig. 42-3), we have with $A = \infty$:

$$v_{\text{out}} = -\frac{1}{RCp} v_1 \quad \text{or} \quad v_{\text{out}} = -\frac{1}{RC}\int_0^t v_1 \mathrm{d}t$$

The output voltage is then proportional to the integral of the input voltage. Similarly, we obtain in principle when $Z_o = R$ and $Z_1 = 1/Cp$:

$$v_{\text{out}} = -RC\frac{\mathrm{d}v_1}{\mathrm{d}t}$$

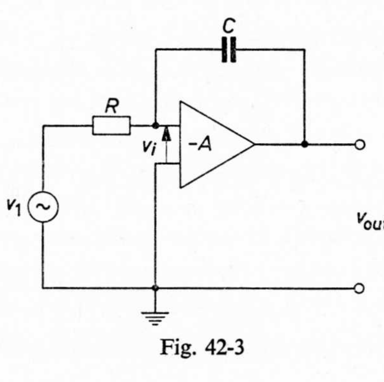

Fig. 42-3

The fact that the d.c. amplifier (usually of the Goldberg type) is not ideal, compels us to take into account with this type of feedback system not only the stability problems we have mentioned before, but also the fact that the operations can only be carried out with limited accuracy. For example, stability criteria prevent the use of this circuit as differentiator over a large frequency range. Regarding the lack of accuracy, apart from the finite amplification, the chief problems encountered are the limited bandwidth, the zero error, the input current and possible additional phase shifts.
This may be illustrated by a few calculations.

We have for the addition circuit of Fig. 42-2 for finite amplification:

$$-v_{\text{out}}\left[1+\frac{1}{A}\left(1+\frac{R_o}{R_1}+\frac{R_o}{R_2}+\ldots\frac{R_o}{R_n}\right)\right]=v_1\frac{R_o}{R_1}+v_2\frac{R_o}{R_2}+\ldots+v_n\frac{R_o}{R_n}$$

At given values of resistance, a relative change $\Delta A/A$ in the finite amplification therefore produces a relative change in the coefficient of value $\Delta A/A \cdot \varepsilon$, where:

$$\varepsilon=\frac{1}{A}\left(1+\frac{R_o}{R_1}+\frac{R_o}{R_2}+\ldots+\frac{R_o}{R_n}\right)$$

We find for the integrator circuit of Fig. 42-3:

$$v_{\text{out}}=-Av_i \quad \text{and} \quad \frac{v_1-v_i}{R}=(v_i-v_{\text{out}})pC$$

so that:

$$\left\{\left(1+\frac{1}{A}\right)p\tau+\frac{1}{A}\right\}v_{\text{out}}=-v_1 \qquad (42.1)$$

where $\tau=RC$. For a perfect integrator we have $v_{\text{out}}=v_1/p\tau$, so that the operator $\{(1+1/A)+1/Ap\tau\}$ is the disturbing factor; its effect depends on the waveform.

For example, if v_1 is a step function of value V, v_{out} should equal $-Vt/\tau$, where t = integration time. However, we find by Heaviside's calculation method from equation (42.1):

$$v_{out}=-AV\left\{1-e^{-t/(A+1)\tau}\right\}$$

for which can be written in good approximation:

$$v_{\text{out}} = -\frac{A}{A+1} V\left\{\frac{t}{\tau} - \frac{t^2}{2(A+1)\tau^2}\right\}$$

The effect of the finite amplification is therefore expressed by the factor:

$$\frac{A}{A+1}\left\{1 - \frac{1}{2(A+1)}\,\frac{t}{\tau}\right\}$$

If the signal is sinusoidal, both amplitude and phase will be affected, and further an interfering switch-on phenomenon occurs. With $v_1 = V e^{j\omega t}$ the particular solution of differential equation (42.1) becomes:

$$v_{\text{out}} = -V\,e^{j\omega t}\left\{\left(1 + \frac{1}{A}\right) j\omega\tau + \frac{1}{A}\right\}^{-1}$$

which can be written:

$$-V\frac{A}{A+1}\,\frac{1}{\sqrt{\omega^2\tau^2 + (A+1)^{-2}}}\,e^{j(\omega t - \pi/2)}\,e^{j\varphi}$$

where $\qquad \varphi = \text{arc tg}\,\dfrac{1}{\omega\tau(A+1)}.$

We thus find with $v_1 = V \sin \omega t$:

$$v_{\text{out}} = -V\frac{A}{A+1}\,\frac{1}{\sqrt{\omega^2\tau^2 + (A+1)^{-2}}}\cos(\omega t + \varphi)$$

A solution of the homogeneous differential equation $\{(1 + 1/A)p\tau + 1/A\}v_{\text{out}} = 0$ is

$$v_{\text{out}} = k\,e^{-t/(A+1)\tau}$$

and consideration of the initial conditions that the system is quiescent for $t = 0$, gives the complete solution:

$$v_{\text{out}} = -V\frac{A}{A+1}\,\frac{1}{\sqrt{\omega^2\tau^2 + (A+1)^{-2}}}\left\{\cos(\omega t + \varphi) - e^{-t/(A+1)\tau}\cos\varphi\right\}$$

where $\cos\varphi = \omega\tau/\sqrt{\omega^2\tau^2 + (A+1)^{-2}}$. In the idealized case the solution would have been

$$v_{\text{out}} = -V\frac{1}{\omega\tau}\cos\omega t$$

Similar calculations can be undertaken for the effect of the limited bandwidth and phase shifts. They again show that these effects are dependent on the specific input waveform.

The effect of the zero drift and the input current is also important. When considering the integration circuit of Fig. 42-4, where the zero error is allowed for by the equivalent input voltage v_{dr} of the amplifier and the input current by current source i_g, we see that:

$$\frac{v_1 - v_i}{R} = (v_i - v_{\text{out}})Cp - i_g, \text{ or } v_1 = v_i(1 + \tau p) - i_g R - \tau p v_{\text{out}}$$

and also $v_{\text{out}} = - A(v_i + v_{dr})$, so that

$$-v_{\text{out}}\{ 1/A + \tau p(1 + 1/A) \} = v_1 + v_{dr}(1 + \tau p) + i_g R$$

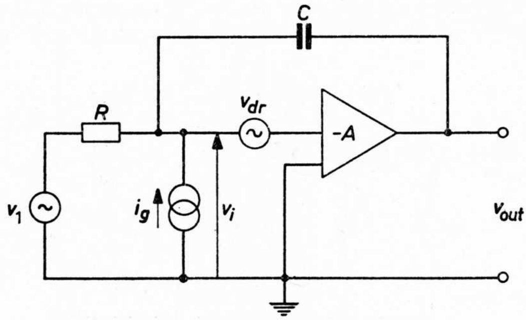

Fig. 42-4

This proves that, compared to the input signal v_1, not only the drift voltage itself, but also its derivative is important, and that the effect of the input current is proportional to resistance R.

A good valve amplifier allows the measurement of integrals up to 100 μVs per minute at rather high values of R. We should remember in this respect that the integral of the drift voltage will increase as the square of the time because a drift voltage behaves linearly with time to a first approximation. Over very short measurement periods (of the order of seconds) the derivative of the drift voltage also plays a part. A typical value, corresponding to the 100 μVs per minute measurement is: 1 μVs per second.

This value may not be small enough for some measurements. We must then use chopper-stabilized amplifiers. With transistorized circuits the limitation is defined by the extent to which the base current can be compensated. The impedance level is then obviously much lower.

It is also theoretically possible to replace one or more of the impedances in the circuit of Fig. 42-1 by components which have a non-linear relation between potential and current. As it is then still more difficult to satisfy the stability conditions, these circuits are only used in a few incidental cases. To sum up, we can say that in measurement techniques, the integration and the summation circuit are primarily applied.

As mentioned when discussing the principle of "adding what is lacking", this principle can often be used to advantage instead of feedback. As example we discussed an addition circuit (Fig. 27-6) which has greater accuracy, stability and response than the circuit of Fig. 42-2.

With regard to linear operation, we should once more emphasize the usefulness of the difference amplifier; not only because of its excellent linear processing of difference measurements, but also because of the fact that it is a simple operational amplifier where one of the input grids can be used

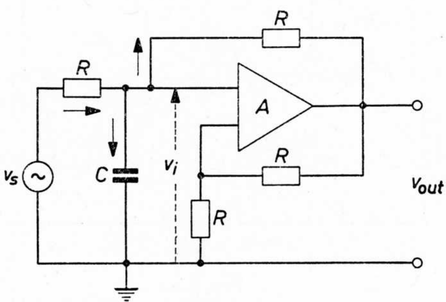

Fig. 42-5

for feedback. An example of this is the integrator circuit shown in Fig. 42-5. Its principle was discussed in Section 26 as an example of impedance transformation; it serves to compensate the effect of voltage v_i across the capacitor, which interferes with perfect integration. By feeding the output back to the lower input grid, the output voltage will be $2v_i$, provided the difference input voltage has been sufficiently amplified. We thus have:

$$\frac{(v_s - v_i)}{R} = \frac{(v_i - 2v_i)}{R} + v_i Cp \quad \text{or} \quad v_{\text{out}} = 2v_i = \frac{2v_s}{RCp}$$

which indicates perfect integration. With this circuit we only have to take into account (in the normal field of operation of the amplifier) the limitations imposed by zero-point drift and input current. By replacing capacitor C by an inductor, a differentiating circuit is obtained. However, there is little need for this, because the simple circuit shown on the left-hand side of Fig. 42-6 (corresponding to Fig. 12-5) is excellent for this purpose. If necessary, the internal resistance of the cathode follower can be further reduced by applying simple feedback by means of a transistor (right-hand side of Fig. 42-6). The internal resistance of the cathode was 0.1 Ω for the values indicated, which is 3000 times smaller than with the left-hand circuit!

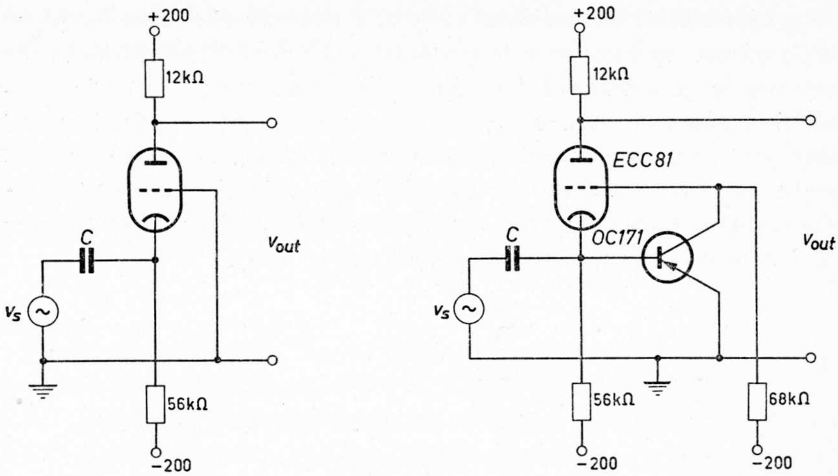

Fig. 42-6

As noted in Section 12, this type of circuit is also excellent as an addition circuit (Fig. 12-7).

While the number of linear operations is very limited, the collection of non-linear operations is theoretically infinite. It is therefore impossible to discuss all circuits developed over the years which must be considered as belonging to this group. Perhaps the most interesting procedure is to discuss those circuits which are the basis of a large number of operations.

For instance, one method utilizes the facility of producing impedances which depend on frequency in a certain manner. For example, the impedances discussed in Section 33, whose amplitudes above a certain minimum frequency $\omega_{\min}$ can be made proportional to ω^{α}, provided $-1 \leqslant \alpha \leqslant 1$.

If a frequency-modulated current $i(t) = i_0 \sin \int_0^t \omega_{\text{inst}}\, dt$ is fed through such an impedance, where ω_{inst} is proportional to the signal v_i, the amplitude of the voltage across this impedance will be proportional to ω^{α}, and therefore also to v_i^{α}, provided the conditions of Carson and Fry (Section 36) are satisfied. We thus obtain after detection, the operation:

$$v_{\text{out}} = c(v_{\text{in}})^{\alpha}$$

where

$$-1 \leqslant \alpha \leqslant +1$$

The limiting case $\alpha = -1$ is obtained by taking for the impedance a capacitor. If transfer impedances are used instead of ordinary impedances, values of α outside the indicated range are also possible. Since the phase shift produced by these impedances is the same for all frequencies in the frequency range considered, i.e. $\frac{1}{2}\pi\alpha$, and since all components of a non-sinusoidal voltage lie in the frequency range considered for the fundamental component (the range extends theoretically from ω_{min} to ∞) we can also use the triangular waveforms produced by the f.m. circuit of Fig. 41-14. The rectangular waveform occurring during the switching of the Schmitt trigger can then be used for synchronous detection (Fig. 42-7).

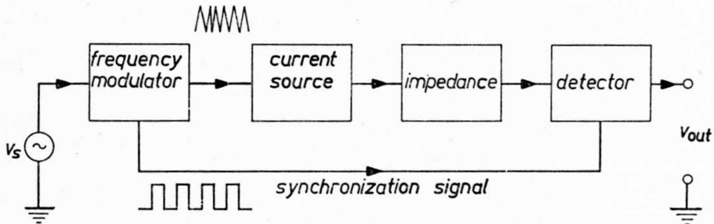

Fig. 42-7

Another method (Fig. 42-8) uses the same basic circuit as the f.m. circuit of Fig. 41-14. Since the voltage across capacitor C must lie between the two fixed levels of the Schmitt trigger, the mean voltage across the capacitor will be constant If B_1 conducts for a time T_{disch} and the period is T the mean discharge $i_2\,T_{disch}$ is equal to the injected charge i_1T, and thus $T_{disch}/T = i_1/i_2$.

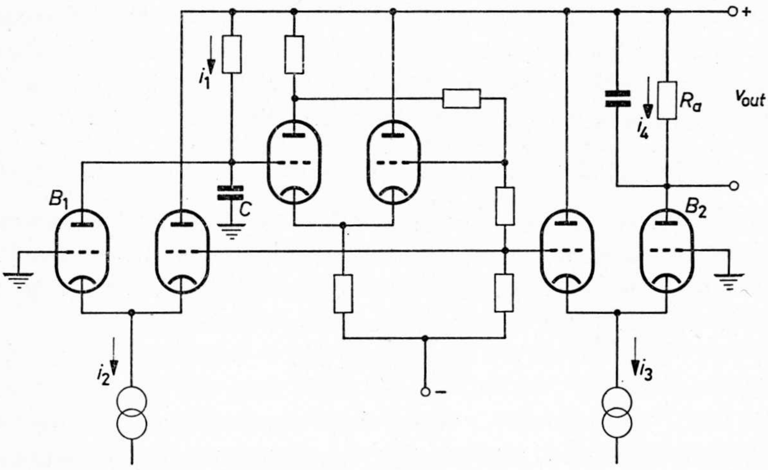

Fig. 42-8

If valve B_2 passes current i_3 for the same time $T_{\rm disch}$, we find for the mean current i_4 through this valve:

$$i_4 = i_3 \frac{T_{\rm disch}}{T} = \frac{i_3 i_1}{i_2}$$

and for the output voltage:

$$v_{\rm out} = - R_a \frac{i_3 i_1}{i_2}$$

Before considering the various applications, we should mention the remarkably high accuracy obtained with the above-mentioned relations. This is due to the fact that only a number of comparatively simple invariable factors is made use of, e.g. as long as the Schmitt trigger operates, the voltage across the capacitor must retain the same mean value; the time during which B_1 conducts is then accurately fixed, whilst the simultaneous opening and closing of B_1 and B_2 does not pose any difficulty because of the large square-wave control voltages. Accuracy is only limited by the extent to which the condition that a valve is open also means that no current passes through that valve. It appears that the emission from filament to anode is the chief offender

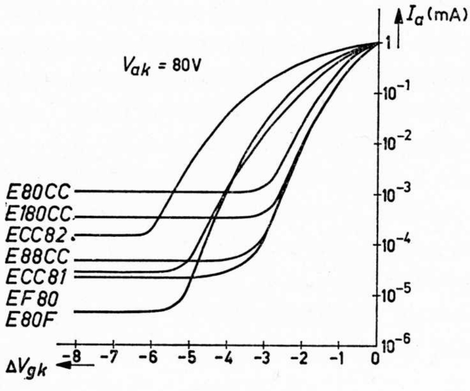

Fig. 42-9

in this respect. Fig. 42-9 shows the result of the measured currents for a few valve types. Dependent on the type, these currents prove to lie between approx. 1 μA and 1 nA. Carrying out the same analysis with transistors, accuracy is determined by the variations in the ratio of base to collector current, as well as by the leakage currents of the collectors. However, here

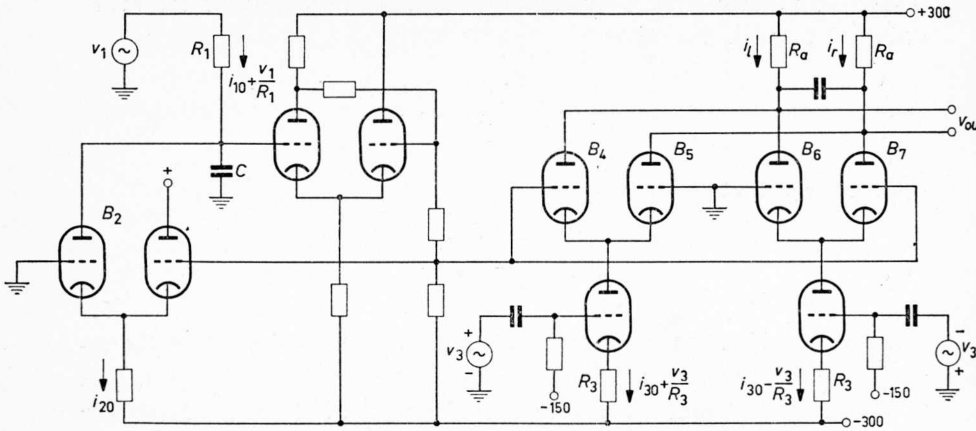

Fig. 42-10

too it is possible to design a circuit which is adapted to the specific properties of the transistor used as a switch. Circuits with field-effect transistors can be designed in a similar way to those with valves.

When current i_1 is kept constant (Fig. 42-8) the output voltage will be proportional to the quotient of currents i_3 and i_2, or $v_{\text{out}}=Ci_3/i_2$, which means that this network is capable of carrying out the division of two signals. When current i_3 is also kept constant, the output voltage will be inversely proportional to i_2. This type of circuit can be applied to convert the time intervals between irregularly occurring pulses (e.g. heart beats) into a corresponding instantaneous frequency.

When current i_2 is kept constant, the output voltage will be proportional to the product of currents i_1 and i_3, and we have obtained a multiplication circuit. This circuit can, of course, only be used for producing products or quotients with a positive value of i_3. By using the balanced design of Fig. 42-10, a "four-quadrant" multiplier can be obtained, i.e. a circuit which, for both polarities of the two signals v_1 and v_s, gives the product:

$$v_{\text{out}} = c\, v_1 v_3$$

When $v_1=v_3$, a pure squaring circuit is obtained.

Using the same symbols (0 indicating a quiescent current) we have for the left part of Fig. 42-10:

$$i_{20}T_{\text{disch}} = \left(i_{10} + \frac{v_1}{R_1}\right)T \quad \text{or} \quad T_{\text{disch}} = \frac{\left(i_{10} + \dfrac{v_1}{R_1}\right)}{i_{20}}T$$

where T_{disch} is the time that valve B_2 conducts. Valves B_5 and B_6 conduct for the same time, so that we can write for currents i_l and i_r in the common anode circuits:

$$i_l T = \left(i_{30} + \frac{v_3}{R_3}\right)(T - T_{disch}) + \left(i_{30} - \frac{v_3}{R_3}\right) T_{disch}$$

and

$$i_r T = \left(i_{30} + \frac{v_3}{R_3}\right) T_{disch} + \left(i_{30} - \frac{v_3}{R_3}\right)(T - T_{disch})$$

Therefore:

$$i_r - i_l = \frac{2v_3}{R_3}\left(\frac{2T_{disch}}{T} - 1\right) = \frac{2v_3}{R_3}\left(\frac{2i_{10}}{i_{20}} + \frac{2v_1}{i_{20}R_1} - 1\right)$$

so that, for $i_{20} = 2i_{10}$, the output voltage $v_o = R_a(i_r - i_l)$ becomes proportional to $v_1 v_3$, where both polarities are allowed for the voltages.

The division circuit can, of course, also be designed as a balanced circuit.

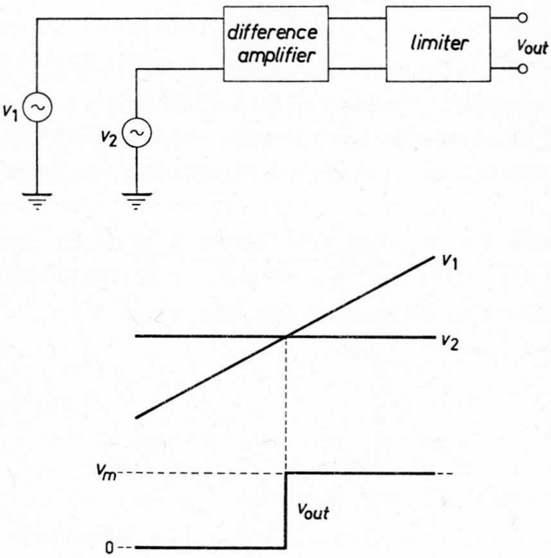

Fig. 42-11

The "inverse-function generator" is used in a third method which offers very many facilities. If we have a good difference amplifier (Fig. 42-11) followed by an instantaneous limiter (in practice usually the last stage of the difference amplifier itself) the following relations can be obtained for the output signal of this combination:

$$v_{out} = 0 \text{ for } v_1 < v_2 \quad \text{and} \quad v_{out} = v_m \text{ for } v_1 > v_2$$

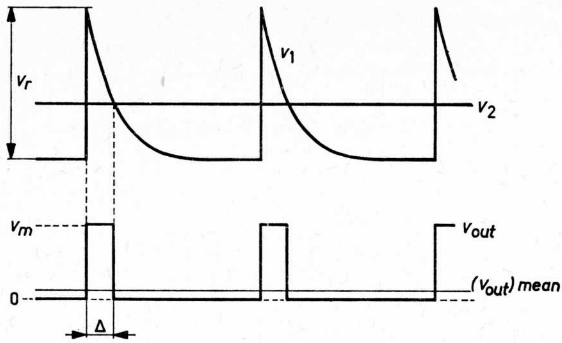

Fig. 42-12

independent of the level of v_1 and v_2. If v_1 is a periodic signal, e.g. a periodically repeated exponential function $v_1 = v_r\,e^{-\alpha t}$, where the constant voltage v_r is larger than $(v_2)_{max}$, we find for the time Δ that v_1 is larger than v_2 (Fig. 42-12):

$$\Delta = -\frac{1}{\alpha}\ln\frac{v_2}{v_r}$$

so that Δ and hence the mean output voltage (v_{out}) will vary in proportion to ln v_2. More generally we can say that when v_1 is a periodic signal which depends on time and this dependence is not of a complicated multi-valued nature, the mean output voltage will depend on v_2 according to the inverse function.

For example, if v_1 is a sinusoidal voltage $v_r \sin \omega t$, the output voltage will be proportional to arc sin (v_2/v_r), whilst if $v_1 = ct^2$, the output voltage will be proportional to $\sqrt{v_2}$ (top part of Fig. 42-13).

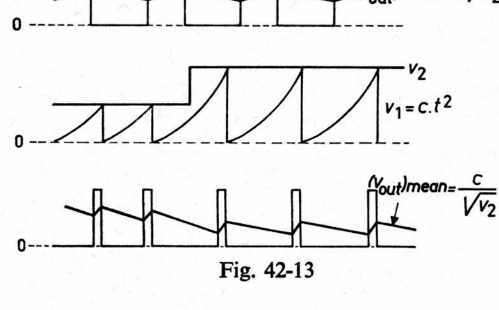

Fig. 42-13

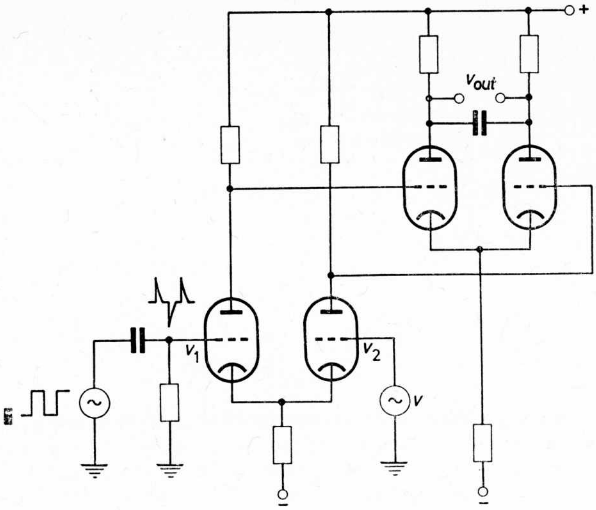

Fig. 42-14

A variation occurs when the frequency of the scanning signal v_1 is also determined by v_2, as indicated for the quadratic function in the lower part of Fig. 42-13. If the pulse produced per cycle has a fixed height and width, the output signal will be proportional to the reciprocal value of the inverse function: $(v_{out})_{\text{mean}} = c/\sqrt{v_2}$.

It is, of course, also possible to make the pulse height of the former circuit, and both pulse height and width of the latter circuit dependent on other signals, so that a great many computing circuits can be realized in this manner. Fig. 42-14 gives an example of a simple logarithmic voltmeter based on the principle of the inverse-function generator, where only moderate precautions regarding zero-point drift of the amplifier are necessary to operate without difficulty over a factor of 1000 (millivolts to volts). As can be easily calculated, we have in this case:

$$v_{out} = c \frac{v}{|v|} \ln|v|.$$

As a special application of the inverse-function generator we should mention the generator of sinusoidal voltages with periods of hours. By applying a sinusoidal voltage with amplitude v_r and a relatively high frequency to one input of the inverse function generator (Fig. 42-15), and to the other input a voltage v_2, the output voltage will be proportional to arc sin v_2/v_r. If we now take for v_2 a sinusoidal voltage with the same amplitude as v_r, but of a much lower frequency, v_{out} will become a triangular voltage, whose height will be determined by the

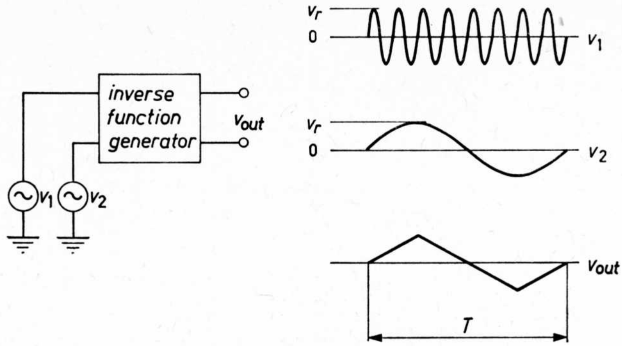

Fig. 42-15

limiting circuit of the inverse-function generator, and where the period will be equal to that of v_2. We thus achieve a sine-triangle transformation.

The reversed process is more important in practice and is obtained by using feedback (Fig. 42-16). The feedback to the input (2) of the inverse-function generator compels the output (3) to follow a triangular voltage. When the latter has the correct amplitude, a sinusoidal voltage will appear at (2) which has the amplitude of the auxiliary voltage across (1) and the cycle of the triangular voltage. The voltage on the output of the inverse-function generator ensures that the triangle generator produces a voltage of the correct value. Since it is easy to ensure an accurately linear charging and discharging of capacitors for very long periods (hours), it is possible to produce very pure sinusoidal voltages with cycles of many hours in this way. The advantage here is that these voltages do not exhibit any switch-on phenomena and can start with any required phase. The maximum period of an oscillator constructed according to this principle proved to be approx. 4 hours, whilst the deviation from the sinusoidal form was less than 1 per cent.

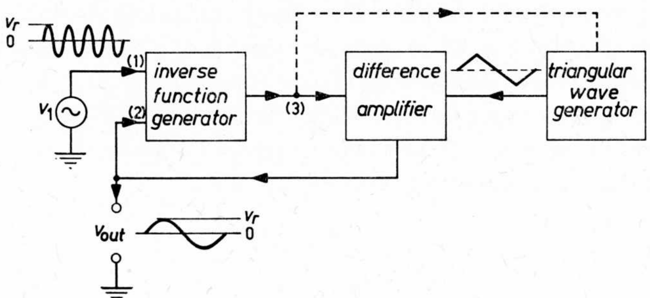

Fig. 42-16

Should we attempt to produce a sinusoidal voltage with a period of several hours in the normal way, i.e. by satisfying the differential equation $d^2V/dt^2 + \omega^2 V = 0$ for this voltage, the duration of the switch-on phenomena (apart from the problem of amplitude stabilization) would amount to 5 to 10 cycles. In other words, it would take us a whole day!

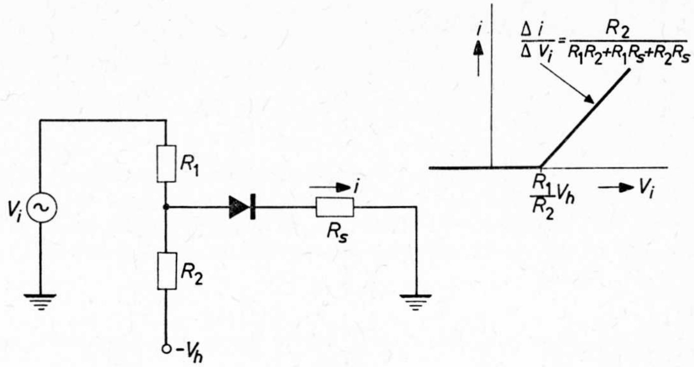

Fig. 42-17

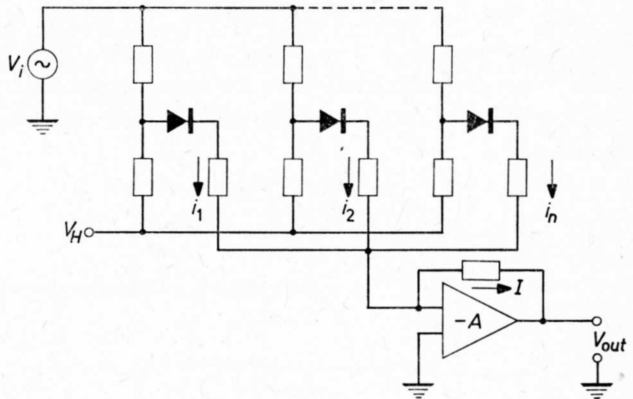

Fig. 42-18

A frequently used method for the production of non-linear relations between potentials and currents utilizes a combination of resistors and diodes, as illustrated in Figs 42-17—19. Assuming an ideal diode in Fig. 42-17, the relation between current i and input voltage v_i will be as indicated. The point on the abscissa is determined by the ratio of R_1 to R_2 and the value of the auxiliary voltage V_h, whilst the slope of the characteristic is determined by the three resistors R_1, R_2 and R_s. By changing the connections of the diode and/or reversing the polarity of the auxiliary voltage, we can obtain arbitrary line segments. Summation of the currents i of a number of similar circuits (e.g. as in Fig. 42-18 by means of an operational amplifier) gives an accurate approximation of the non-linear relation between the total current I (i.e. the output voltage v_{out} and the input voltage v_i (Fig. 42-19).) Many variations of these methods are obviously possible, such as using transistors instead of diodes as switches.

The linear approximation of a curve $y = f(x)$ is based on approximating this function by the first two terms of the Taylor's series:

$$f(x + h) = f(x) + hf'(x) + \frac{h^2 f''(x)}{2!} + \ldots \frac{h^k}{k!} f^{(k)}(x) + \ldots$$

If $f''(x)$ is not zero, the resultant error will be of the order $\frac{1}{2}h^2 f''(x)$. If all derivatives $f''(x) \ldots f^{(k-1)}(x)$ are zero, the error will be of the order of magnitude $h^k f^{(k)}(x)/k!$. If it is required that the deviation nowhere exceeds δ, the permitted values of h can be derived from this series expansion.

Some measurement apparatus make use of linear potentiometers to which resistors can be connected at mutually equal distances (Fig. 42-20). In the special case that all resistors have the same value, the voltage will behave exponentially as a function of the slider position. This is used in some pen recorders with logarithmic indication. There are also potentiometers where the resistance between the slider and one of the contacts is made to represent a certain function (ln, sin, cos). Accurate working specimens of this type are very expensive.

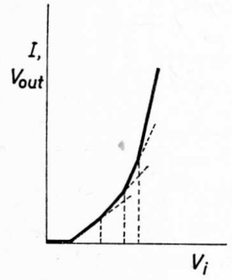

Fig. 42-19

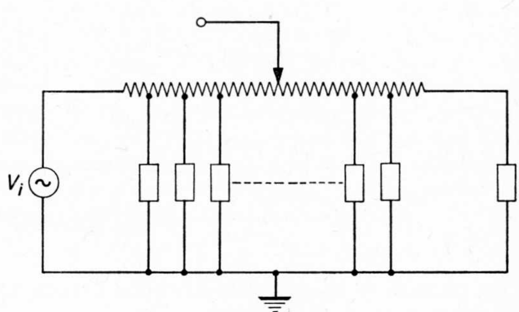

Fig. 42-20

Amplitude modulation also offers the possibility of mathematical operations, such as the vectorial addition of two values. For example, when signal A is modulated on carrier wave $\cos \omega t$ and signal B on $\sin \omega t$, the sum signal will be:

$$A \cos \omega t + B \sin \omega t = \sqrt{A_2 + B^2}\,(\cos\varphi \cos \omega t + \sin\varphi \sin \omega t) =$$

$$\sqrt{A^2 + B^2} \cos(\omega t - \varphi)$$

where $\quad \cos \varphi = \dfrac{A}{\sqrt{A^2 + B^2}}$ and $\sin \varphi = \dfrac{B}{\sqrt{A^2 + B^2}}$

If the signal is demodulated, the amplitude will be proportional to $\sqrt{A^2+B^2}$, and the phase difference φ between the sum signal and $A \cos \omega t$ will be proportional to arc tan B/A.

The value $\sqrt{A^2+B^2}$ can, for example, be used to make the spot intensity of an oscilloscope proportional to the writing speed, so that the brightness remains constant. The writing speed is $\sqrt{v_H{}^2+v_V{}^2}$, where v_H and v_V represent the horizontal and vertical writing speeds; these are obtained by differentiation of the deflection voltage. Instead of multiplying by sin ωt and cos ωt, it is simpler and more accurate to do this with square waveforms with a time interval of $\frac{1}{4}$ period and by filtering out the fundamental after summation.

These square waves can be obtained by starting from a square-wave voltage with double the frequency required to control two flip-flop circuits. In order to ensure that the two pulses are mutually shifted by exactly $\frac{1}{4}$ period, we must use the combination of multivibrator and flip-flops, as mentioned in Section 39 and shown in Fig. 42-21.

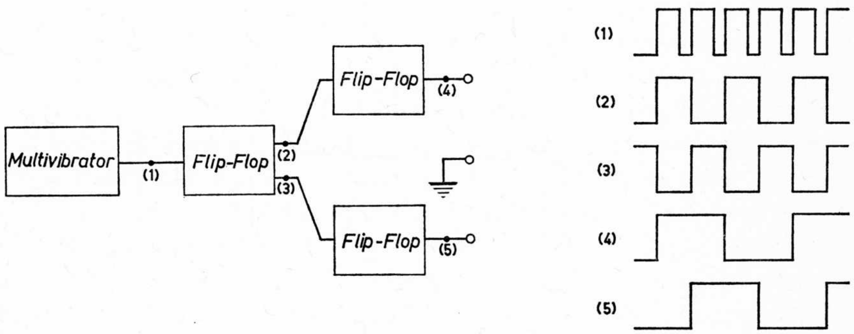

Fig. 42-21

Apart from the methods discussed above, which are based on principles which in themselves are operations, there are many operational circuits which can be traced back to the properties of the components used. A simple example of the latter is the quadratic relation between the applied voltage and the heat generated by a resistor, which may be used to determine the effective values of currents and potentials. Generally speaking, the accuracy of this class of operations is relatively small; on the other hand these methods are rather simple and cheap.

A method often applied in measurement electronics, and which belongs to this group, has been mentioned in Section 32, namely that the relation between grid current and grid voltage is logarithmic over many decades (10^{-12}–10^{-4}A) for a number of valve types, so that the output voltage is proportional to the logarithm of the input current. Among other applications this is used for the measurement of the current in an ionization chamber, where the valve is sometimes built-in to minimize the effect of the leakage currents.

In a similar way we can make use of the exponential relation between the base voltage and the collector current of a transistor.

43. Accuracy

We mentioned in the previous section that digital circuits allow signal processing with infinite accuracy. One could therefore say that in principle the accuracy of signal processing does not offer any problems. By only applying digital operations, the accuracy will remain the same as for the primary signal. However, as we have mentioned in the introduction, this is too idealistic a presentation for several reasons.

Firstly, the actual measurement signal is frequently not electrical and must first be converted to an electrical one. This is, however, almost never done in digital form, and is always accompanied by a certain inaccuracy. Furthermore, these electrical signals usually need amplification before applying operations to them, which introduces further inaccuracy. Another point is that digital solutions are too expensive for the majority of operations desirable in measurement electronics, so that this part of the circuit is also designed in analogue form. We can therefore say that the final accuracy of the electrical part of signal processing is determined by the accuracy of the linear amplification and of the non-digital processing circuits. We shall therefore conclude this book by discussing a few relevant points which have not been mentioned or only barely referred to in the previous sections.

Several causes can be blamed for the inaccuracy of measurement results and signal processing. In the first place we have the continual presence of noise in electrical equipment. Inaccuracy produced by noise obviously plays a greater part when the signals in the circuit are small. When large signals are being processed other factors are usually more important and overshadow the noise effect. If noise is important for final accuracy, the circuit must be designed so that it is reduced to a level approaching the theoretical limit. Since noise sources and their values are known, their contribution can be calculated for each configuration. This allows us to select the most advantageous from several possible configurations. Nevertheless, noise will always impose a natural limit on accuracy.

An example of an entirely different form of inaccuracy is met when a transfer system is assumed to be linear while in fact it is not. The resulting error is here not limiting since it can be fully determined and compensated for, sothat we have complete control over it.

Most sources of inaccuracy in electronic equipment are based on a combination of inherent processes and a neglect of necessary precautions. We do not mean the latter in any way to be derogatory. For example, most components are sensitive to changes in temperature, but it would be extremely cumbersome as well as expensive to place all electronic instruments in thermostatically controlled enclosures for this reason alone. The more so because we still have to contend with natural limitations, which will nearly always play a dominating part under normal conditions when suitable circuits and components with a sufficiently low temperature coefficient are selected.

Apart from noise, ageing phenomena in particular set natural limits on accuracy. An example is the change in resistor values by recrystallization of the material. Most precision resistors of good manufacture have been "aged" artificially by keeping them at a high temperature for some considerable time.

Changes also occur in transistors because of the diffusion of impurities in the crystal and reactions at its surface. A modern way of preventing the latter is by coating the surface with a protective layer of oxide. Diffusion can further be retarded by avoiding high temperatures in the transistor.

The greatest changes of this nature occur in valves, which is not surprising in view of the high cathode temperatures. Relevant factors are evaporation of material from the cathode surface, "poisoning" of this surface by foreign atoms from the residual gases, and precipitation of barium on the grids. The only direct precautions which can be taken against the ageing process of valves consist in ensuring constant conditions and compensating their effects as much as possible by using two reasonably identical valves.

In the case of equipment supplied from the mains, the mains voltage is another important factor. As mentioned when discussing supplies, deviations from the nominal value up to 10 per cent in both directions often occur. It pays therefore to use stabilized supplies for most precision equipment, which can be derived from the mains supply as described in Section 29.

When the effect of the above-mentioned and similar interference sources has been reduced in a suitable manner to a permissible value, the valves and transistors will introduce the largest tolerances in electronic circuitry, especially when a new valve or transistor has been put in. One could think that the accuracy of the processes in which these active components are used, will be determined by these tolerances, or even that the accuracy cannot be better than the largest tolerance occuring. However, the discussion on feedback has already shown that this reasoning is incorrect. Employing feedback in an amplifier can be arranged so that the amplification is mainly determined by the ratio of two impedances, i.e. by components which may have (and in general will have) considerably better tolerances than the active components used.

This implies that there exist principles which make it possible to reduce the effects of certain tolerances. Two questions now present themselves:

1. Apart from feedback, are there any other principles with this same property, and which are they?
2. Is it possible to reduce the effect of all tolerances by taking suitable precautions? Because in the case of feedback, the tolerances of the impedances used for feedback retain their full weight.

In answer to the first question we can say yes, there are other principles that can be used to reduce the effect of certain tolerances. Some of these have been discussed, e.g. the principle of "adding what is lacking", but we have certainly not covered all possibilities. Indeed we may wonder if it would be possible to summarize the possibilities; it would at the very least be difficult to guarantee that any list would be complete for we lack a logical way of arriving at principles of this nature. Of course, we can find some of the fundamental reasons by tracing the basis of the accuracy of known accurate circuits. The authors have thus come to the conclusion that great accuracy in electronics is basically founded on very simple, almost trivial, truths. Let us give a few examples in practical circuits given in this book:

a. With a good triode, correctly adjusted, the grid current is negligible so that anode and cathode currents equal each other with great accuracy;
b. If, after a certain time, the voltage across a capacitor resumes its original value, this indicates that the charges flowing into and out of the capacitor during that time interval equal each other.
c. We shall see that the fact that the loop gain with an oscillator deviates extremely little from unity, can be used also for designing accurate circuits.

A complete survey of these types of principles would undoubtedly be most useful to electronic engineers, but the fact that such a survey does not exist should be a stimulating thought for the devotees of this science. It is also consoling to know that simplicity leads to equally good, if not better, results than complexity.

This simplicity refers to the principle and not to the number of components used, although the two often go together. A circuit can be made more elegant by using a smaller number of components, but only when this does not interfere with the properties of the whole system. Economy in the use of components – particularly in the case of expensive or unreliable ones – is of course a sound principle. However, especially in the "professional" sector, economizing on components will often be belied by a lower system reliability.

It is often possible to achieve a great gain in quality by simple changes. A striking example is shown in Fig. 43-1 where it is intended to give the same value to the grid biases of two triodes, amounting to approximately half the available supply voltage (difference amplifier, synchronous detector). If a

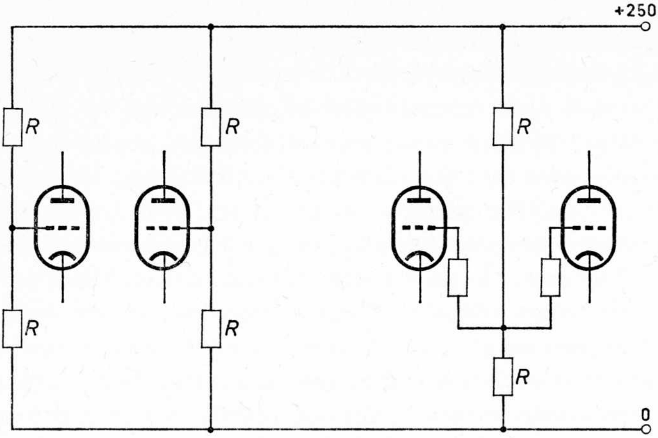

Fig. 43-1

maximum difference of 1 mV between the grid voltages is to be guaranteed (left-hand side of Fig. 43-1), a tolerance not exceeding 10^{-5} is permitted for the resistors. Moreover, the current noise of the resistors, which may be quite considerable, also occurs as difference voltage. The right-hand side circuit has the same components. The grid biases are here always the same, independent of any deviations in the resistors, and the current noise of the resistors R does not appear as difference voltage between the grids.

Fig. 43-2 illustrates a second example. If it is required to produce a voltage which is accurately proportional to the difference between voltages v_1 and v_2, we could in the first place think of using the operational amplifier which we know to be an accurate processing circuit. We then obtain the circuit shown

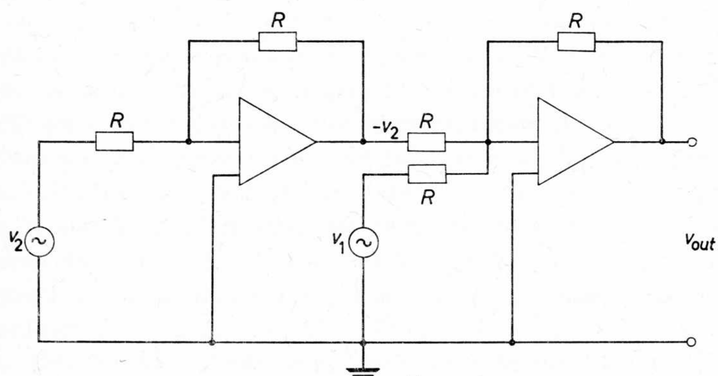

Fig. 43-2

in Fig. 43-2. We first derive a voltage $-v_2$ from v_2 and then add the former to v_1. Although these operations can be carried out quite accurately individually, the coefficients of v_1 and $-v_2$ may show a difference of the order of magnitude of the tolerances in the resistors. Even if the resistors are adjusted, we must take a possible deterioration in these resistors into account. This gives an accuracy of 0.1 per cent difference between the coefficients of v_1 and $-v_2$, which is about the best result obtainable. However, by using a good difference amplifier, which does not require any precision component, an accuracy in the difference voltage will be achieved without any difficulty which is one order of magnitude better. This will remain so even if there is a deterioration of many per cents in one or more of the components used.

The answer to the second question is also yes; this answer is given by visualizing circuits with the property that the tolerance in one of their electronic characteristics is smaller than the tolerance of each component used. The cathode follower (Fig. 43-3) is a simple example of this. "Amplification" here depends on the amplification factor μ, the internal resistance r_a and the cathode resistance R_k:

$$A = \frac{v_o}{v_i} = \frac{\mu R_k}{r_a + (\mu + 1)R_k}$$

We can now prescribe that a valve must be used for this cathode follower with an internal resistance r_{an} and an amplification factor μ_n, whilst the cathode resistor must be R_{kn}. However, components of exactly these values are not commercially available and we are therefore forced to permit deviations. Let us assume for the sake of simplicity that components are allowed if their values differ from the nominal values by a factor plus or minus λ; e.g. $\lambda = 1.05$ at a tolerance of 5 per cent. It will be easily seen that maximum amplification is reached when μ and R_k reach their largest values $\lambda\mu_n$ and λR_{kn}, and r_a its smallest value r_{an}/λ:

$$A_{\max} = \frac{\lambda^3 \mu_n R_{kn}}{r_{an} + \lambda^2(\lambda\mu_n + 1)R_{kn}}$$

Similarly:

$$A_{\min} = \frac{\mu_n R_{kn}}{\lambda^3 r_{an} + (\mu_n + \lambda)R_{kn}}$$

The tolerance factor λ_t for the amplification is therefore:

$$\lambda_t = \sqrt{\frac{A_{\max}}{A_{\min}}} = \sqrt{\frac{\lambda^3\{\lambda^3 r_{an} + (\mu_n + \lambda)R_{kn}\}}{r_{an} + \lambda^2(\lambda\mu_n + 1)R_{kn}}}$$

It is easy to conclude that λ_t will be smaller than λ if we ensure that $(\lambda^2+1)r_{an}<\lambda\mu_n R_{kn}$, which can be approximated in practice to $R_{kn}>2/S_n$, where S_n is the nominal transconductance. This is not a stringent requirement. We find for very large values of R_{kn} with respect to r_a:

$$\lambda_t - 1 \approx \frac{\lambda - 1}{\mu_n}$$

We thus obtain very small tolerances in amplification.

One could object that the above is not very convincing because the voltage amplification considered is smaller than unity. However, reduced tolerances can also be obtained with amplifications larger than unity. Fig. 43-4 illustrates an example following from Fig. 43-3 which allows amplifications up to 2 times.

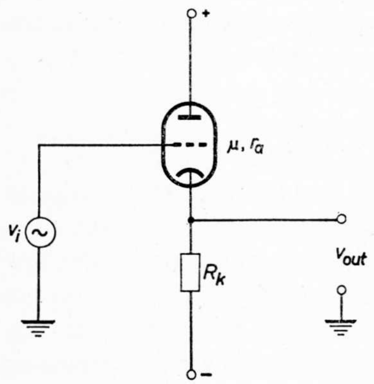

Fig. 43-3

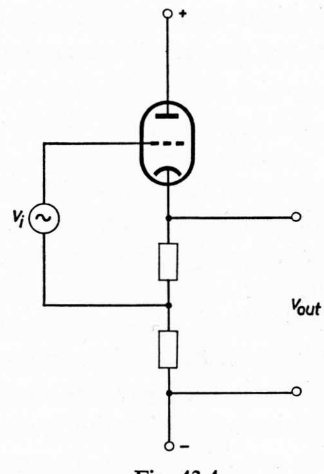

Fig. 43-4

Apart from the fact that this still represents a very mediocre amplification, this circuit is inelegant because the input voltage must be floating. We can avoid this by using a strongly fed-back multi-stage amplifier where feedback is achieved by means of a difference stage input. However, this system does not give more than 2× amplification, and that at the cost of a disproportionate number of components.

We shall now indicate by means of a final example how amplifiers with considerably higher amplification can be designed which possess the required property of a very small tolerance and which are easy to realize.

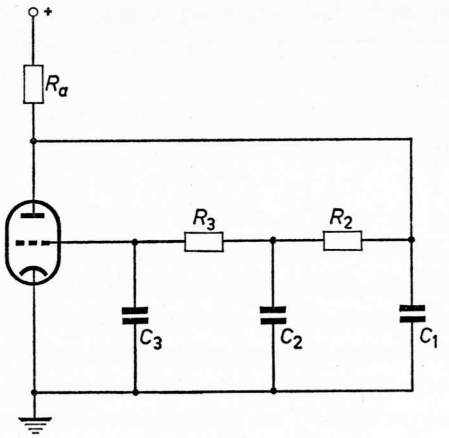

Fig. 43-5

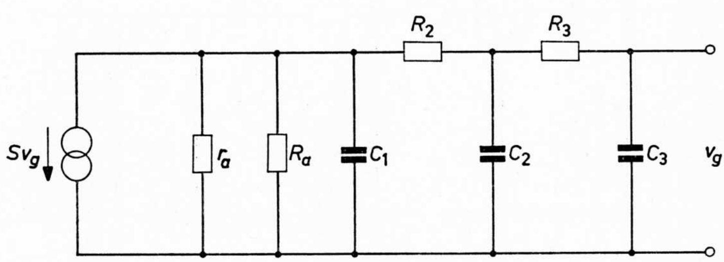

Fig. 43-6

We apply the principle for obtaining stabilized amplification, as mentioned above, where use is made of the fact that the loop gain of oscillating systems deviates from unity by an extremely small amount. The principle of this type of circuit is shown in Figs 43-5—8. In Fig. 43-5 the valve can be considered as being a current source with signal current Sv_g and internal resistance r_a. This circuit corresponds to that of Fig. 43-6 and therefore also to that of Fig. 43-7 where R_1 represents the parallel value of R_a and r_a. At constant frequency, the attenuation of the circuit depends on the values of the components. However, at the frequency for which the phase shift is π, i.e. the frequency at which oscillation will occur provided amplification is sufficient, this dependence will be very small if the most suitable values have been selected. This implies $R_3 \gg R_2 \gg R_1$ and $R_1C_1 \approx R_2C_2 \approx R_3C_3$. Entirely neglecting the mutual load of the three sections on each other and nominally putting $R_1C_1 = R_2C_2 = R_3C_3$ with a tolerance factor λ for all resistors and capacitors, it follows from a similar calculation as used for the cathode follower, that the

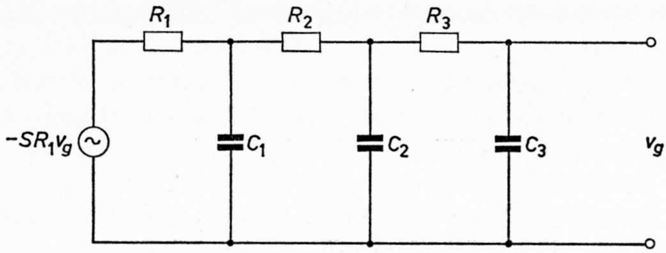

Fig. 43-7

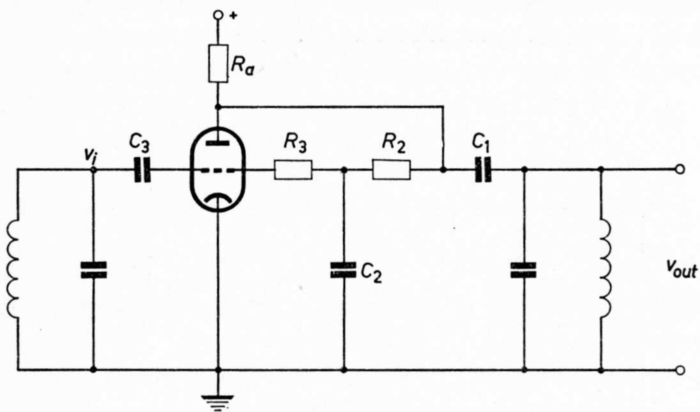

Fig. 43-8

nominal attenuation of the circuit is $4(\lambda^2+\lambda^{-2})$ and the tolerance factor $\frac{1}{2}(\lambda^2+\lambda^{-2})$. We thus find for a few practical values of λ:

$\lambda=1.02$; nominal attenuation	8.0062; tolerance	1.000777
$\lambda=1.05$;	8.0380;	1.004752
$\lambda=1.10$;	8.1458;	1.018223

If we now ensure by a suitable selection of the time constants and using a limiting device to adjust the slope of the valve, that an oscillation with small amplitude occurs at a low frequency, SR_1 will assume the value of the attenuation factor.

This situation will undergo little change when two resonant circuits tuned to a high frequency are inserted in the circuit (Fig. 43-8). At sufficiently high frequencies, the feedback to the grid will be zero because of C_2, and since R_2 was assumed to be large with respect to R_1, the circuit will operate as an amplifier at this frequency. The amplification is $v_{out}/v_i=-SR_1$. The tolerance will not be appreciably increased when suitable values are selected for the resonant circuits.

This circuit has a guaranteed amplification value at given tolerances of the components and in a given frequency range. This may be roughly compared with the guaranteed rejection factor of difference amplifiers at a given tolerance of the components.

These examples demonstrate that, apart from the operational circuits discussed in the previous section, it is also possible to achieve linear amplification with an accuracy which is not necessarily limited by the tolerances of the components. The fact that we are not in principle bound to given limitations makes it possible to invent ever new methods and thus achieve new progress in electronics.

44. Bibliography

Hardly any references have been mentioned in this book because the authors very much doubt the usefulness of the typically long list of papers.

The interested reader can just as easily make his own list by consulting the available literature but the real difficulty is to make a sensible selection of what to read. It seemed therefore better to give the titles of only a number of general books and periodicals which the authors or their collegues, thought of interest to instrument electronics. A number of papers on more specialized subjects are listed which either have been of importance in the development of the subject concerned or give a more detailed discussion.

This bibliography clearly does not pretend to be exhaustive, neither does it imply any opinion on the value of books and papers which are not mentioned.

Periodicals

Literature searches will be greatly facilitated by consulting *Electrical Engineering Abstracts* published by the Institute of Electrical Engineers, London.

There is a great number of specialized or more general periodicals which regularly publish papers on instrumental electronics. We mention the following:

Proceedings of the Institute of Electrical and Electronic Engineers (Proc. I.E.E.E.), U.S.A.

Transactions of the I.E.E.E.:

Sections: Instrumentation and Measurement
Circuit Theory
Bio-medical Electronics

Electronics, U.S.A.

Electronic Engineering, U.K.

Electronic Equipment Engineering, U.S.A.

Electronic Instrument Design, U.S.A.

Electrical Design News, U.S.A.

Electronic Design, U.S.A.

Elektronik, Germany

Instrument Review, U.K.

Internationale Elektronische Rundschau, Germany

The Radio and Electronic Engineer, U.K.

Archiv für Technisches Messen, Germany

Onde Electrique, France

Review of Scientific Instruments, U.S.A.

Journal of Scientific Instruments, U.K.
Solid State Design, U.S.A.

Periodicals published by manufacturers of electronic components and equipment include:

A.E.G. Mitteilungen, Germany
Bell System Technical Journal, U.S.A.
Brüel & Kjaer Review, Denmark
Electronic Measuring and Microwave Notes, Philips, Netherlands
Electronic Application Bulletin, Philips, Netherlands
General Radio Experimenter, U.S.A.
Hewlett Packard Journal, U.S.A.
Instrument Engineer, George Kent, U.K.
Honeywell Instrumentation, U.S.A.
Marconi Instrumentation, U.K.
Mullard Technical Communications, U.K.
Philips Technical Review, Netherlands
RCA Review, U.S.A.
Rohde und Schwarz-Kurzinformation, Germany
Science and Industry, Philips, Netherlands
Siemens Zeitschrift, Germany

Books which treat in detail the general subjects discussed in the present book:

Mathematics

Very comprehensive books discussing in depth practically all the mathematical techniques and functions one is likely to encounter in electronics are:

ANGOT, *Compléments de mathématiques à l'usage des ingénieurs de l'électrotechnique et des télécommunications*, Edition de la Revue Optique, Paris.

KORN and KORN, *Mathematical Tools for Scientists and Engineers*, McGraw Hill, New York.

SOKOLNIKOFF and REDHEFFER, *Mathematics of Physics and modern Engineering*, McGraw Hill, New York.

A less rigorous book is:

WARREN, *Mathematics applied to Electrical Engineering*, Chapman & Hall, London.

A lucid book treating the "daily routine" calculation methods in electronics is:

HEAD, *Mathematical Techniques in Electronics and Engineering Analysis*, Iliffe Books, London.

General Electrical Engineering

Very comprehensive books:

STRATTON, *Electromagnetic theory*, McGraw Hill, New York.

BECKER und SAUTER, *Theorie der Elektrizität* I, Teubner Verlag, Leipzig.

POHL, *Elektrizitätslehre*, Springer Verlag, Berlin.

Simpler, but with a lucid explanation of M-K-S system of units:

CORNELIUS, *Electrical Theory on the Giorgi-system*, Cleaver-Hume Press Ltd. London.

Principles of Radio Valves

Most books give a more or less deep study of the conducting mechanism in valves. A detailed treatment is given in:

SPANGENBERG, *Vacuum Tubes*, McGraw Hill, New York.

Principles of Semiconductors and Transistors

The following papers and books give a simple and easily understandable description of the conducting mechanism in these components:

VAN VESSEM, *The theory and construction of germanium diodes*, Philips Tech. Rev. **16,** 213, 1954.

STIELTJES and TUMMERS, *Simple theory of the junction transistor*, Philips Tech. Rev. **17,** 233, 1955.

STIELTJES AND TUMMERS, *Behaviour of the transistor at high current densities*, Philips Tech. Rev. **18,** 61, 1956.

V.D. ZIEL, *Solid State Physical Electronics*, Prentice Hall, New York.

SEVIN, *Field-effect Transistors*, McGraw Hill, New York.

LINDMAYER and WRIGLEY, *Fundamentals of Semiconductor Devices*, van Nostrand, Princeton.

General electronics

Shortly after the second world war, twenty-seven books on the development of radar were written by members of the staff of the Massachusetts Institute of Technology. Some of these are still extremely interesting, in particular:

VALLEY and WALLMAN, *Vacuum Tube Amplifiers*, McGraw Hill, New York.

CHANCE *et al., Waveforms*, McGraw Hill, New York.

Other books on general electronics:

SEELY, *Electron Tube Circuits*, McGraw Hill, New York.

ZEPLER and PUNNETT, *Electronic Circuit Techniques*, Blackie and Son, London.

Measurement Technique

TERMAN and PETTIT, *Electronic Measurements*, McGraw Hill, New York.

WIND, *Handbook of Electronic Measurement* I and II, Polytechnic Institute of Brooklyn, Brooklyn.

GOLDING and WIDDIS, *Electrical Measurements and Measuring Instruments*, Pitman, London.

REMICH *et al.*, *Electronic Precision Measurements and Experiments*, Prentice Hall, New York.

Handbooks of interest to the electronic engineer:

TERMAN, *Radio Engineers Handbook*, McGraw Hill, New York.

Reference Data for Radio-engineers, I.T.T., New York.

Philips Electron Tube and Semiconductor Handbook.

Philips Taschenbuch für die elektronische Messtechnik.

Components

HENNEY and WALSH, *Electronic Component Handbook*, McGraw Hill, New York.

Linear Circuit Theory

GUILLEMIN, *Introductory Circuit Theory.*

ibid, Communications Networks.

ibid, Synthesis of Passive Networks.

all published by Wiley & Sons, New York.

Feedback (Section 22)

BLACK, U.S.P. 1,686,792; see BODE, *Feedback, the History of an Idea.*

BLACK, *Stabilized Feedback Amplifiers*, Bell System Technical Journal, **13,** 1, 1934.

BODE, *Network Analysis and Feedback Amplifier Design*, van Nostrand, New York.

*BODE, *Feedback, the History of an Idea*, Proc. Symp. Active Networks and Feedback Systems 1960, Polytechnic Press, Brooklyn.

TRUXAL, *Control System Synthesis*, McGraw Hill, New York.

GILLE, PELEGRIN and DECAULNE, *Feedback Control Systems*, McGraw Hill, New York.

Distortion (Section 23)

RODRIGUES DE MIRANDA and ZAALBERG VAN ZELST, *New Developments in Output-transformerless Amplifiers*, Journal of the Audio Engineering Society, **6,** 244, 1958.

"Adding what is lacking" and Parallel Amplifier Feedback (Section 27)

BROCKWAY MCMILLAN, U.S. Patent 2,748,201 of 21-9-51; see BODE, *Feedback, the History of an Idea.*

ZAALBERG VAN ZELST, Dutch Patent 87339 of 19-4-51; *Stabilized Amplifiers,* Philips Tech. Rev. **9,** 25, 1947.

HOROWICZ, *Plant-adaptive systems vs ordinary Feedback Systems,* I.R.E. Transactions on Automatic Control, Vol. AC-7, 1962 No. 1, p.48 and No. 5, p. 119.

Balanced and Difference Amplifiers (Sections 19 and 28)

SCHMITT, *Journal of Scientific Instruments,* **15,** 136, 1938; *Review of Scientific Instruments* **12,** 548, 1941.

PARNUM, *Transmission Factor of Differential Amplifiers,* Wireless Engineering **27,** 125, 1950.

ANDREW, *Differential Amplifier Design,* Wireless Engineering, **33,** 13, 1955.

KLEIN, *Rejection Factor of Differential Amplifiers,* Philips Research Reports, **10,** 241, 1955.

KLEIN and ZAALBERG VAN ZELST, *General considerations on difference amplifiers* Philips Tech. Rev. **22,** 345, 1960; *Circuits for difference amplifiers* I and II, Philips Tech. Rev. **23,** 142 and 173, 1961; *Difference amplifiers with an rejection factor greater than one million,* Philips Tech. Rev., **24,** 275, 1962.

MIDDLEBROOK, *Differential Amplifiers,* Wiley & Sons, New York.

Supplies (Section 29)

PATCHETT, *Automatic Voltage Regulators and Stabilizers,* Pitman, London.

KLEIN and ZAALBERG VAN ZELST, *Combinations of valves and transistors in a stabilized* 2000 *V power supply,* Philips Tech. Rev., **25,** 181, 1963.

Noise (Section 31)

V.D. ZIEL, *Noise,* Prentice Hall, New York.

V.D. ZIEL and BECKING, *Theory of Junction Diode and Junction Transistor Noise,* Proc. I.R.E., **46,** 589, 1958.

Input considerations (Section 32)

BECKING, GROENDIJK and KNOL, *The Noise Factor of Four-terminal Networks,* Philips Research Reports, **10,** 345, 1955.

H.F. Amplifiers (Section 34)

MEINKE und GRUNDLACH, *Taschenbuch der Hochfrequenttechnik,* Springer, Berlin.

MILLMAN and TAUB, *Pulse and Digital Circuits*, McGraw Hill, New York.

D.C. Amplifiers (Section 35)

GOLDBERG, *Stabilization of Wide-band D.C. Amplifiers for Zero and Gain*, R.C.A. Review **11,** 296, 1950.

LANDSBERG, *General Purpose Drift-free D.C. Amplifier*, Philips Research Reports, **11,** 161, 409, 1956.

The Transistor as Chopper (Section 35)

BRIGHT and KRUPER, *Electronics*, 135, 1955.

BRIGHT, *Communications and Electronics*, A.I.E.E., 111, 1955.

EBERS and MOLL, Proc. I.R.E., **42,** 1761, 1954.

Vibrating-Reed Electrometer (Section 35)

VAN NIE and ZAALBERG VAN ZELST, *A vibrating capacitor driven by a high-frequency electronic field*, Philips Tech. Rev. **25**, 95, 1963-64.

MULLER, *The measurement of currents and voltages with the vibrating reed-electrometer*, Thesis University of Amsterdam, 1951 (in Dutch).

Bandwidth and Modulation (Section 36)

LOUIS CUCCIA, *Harmonics, Sidebands and Transients in Communication Engineering*, McGraw Hill, New York.

Oscillators (Sections 37 and 42)

THOMAS, *Theory and Design of Valve Oscillators for Radio and other Frequencies*, Chapman and Hall, London.

KLEIN and ZAALBERG VAN ZELST, *A low-frequency oscillator with very low distortion under non-linear loading*, Philips Tech. Rev., **25**, 22, 1963-64.

KLEIN and DEN HERTOG, *A Sine-wave Generator with Periods of Hours*, Electronic Engineering, **31,** 320, 1959.

Stability criteria (Section 38)
(see also Feedback)

*HURWITZ, *Über die Bedingungen unter welchen eine Gleichung nur Wurzeln mit negativen reellen Teilen besitzt*, Mathematische Annalen, **46,** 273, 1895.

NYQUIST, *Regeneration Theory*, Bell System Technical Journal, **11,** 126, 1932.

Relaxation Circuits (Section 39)

REICH, *Functional Circuits and Oscillators*, van Nostrand, Princeton.

Mathematical Operations (Section 42)

KORN and KORN, *Electronic Analog Computers*, McGraw Hill, New York.

Accuracy (Section 43)

ZAALBERG VAN ZELST, *Constant amplification in spite of changeability of the circuit elements*, Philips Tech. Rev., **9**, 309, 1947.

* Papers marked with an asterisk appear in the recently published book: BELLMAN and KALABA, *Selected Papers on Mathematical Trends in Control Theory*, Dover Publications, New York.

Index