5 Building blocks for high-speed analogue circuits

5.1 Executive summary

This chapter is about the use of electronic devices in elementary circuit blocks found in any micro- or millimetre-wave system – or in the analogue portions of fibre-optic communications systems. An introductory section describes linear two-ports on the basis of scattering parameters, discusses different gain definitions and treats important aspects of stability as well as noise in two-ports, with special emphasis on noise reduction by proper choice of generator impedance.

Following this, amplifiers, oscillators and mixers are described in sequence. In the amplifier section, small-signal parameters are used to derive fundamental properties of common topologies, from the simplest, one-transistor circuits to more complex gain cells, such as the cascode and differential amplifiers. Tuned amplifiers are covered, as well as broadband amplifier techniques, including distributed amplification. Finally, low-noise and power amplifiers are being treated, as well as non-linearities in amplifiers.

The oscillator section discusses how small-signal instability and non-linear gain compression effects combine to create stable sinusoidal oscillations. Important oscillator topologies and noise phenomena affecting the phase stability of oscillators are also covered.

Mixer circuits show how specifically designed non-linear circuits provide frequency-translating capabilities. Mixing principles are discussed first, followed by several mixer topologies using field effect and bipolar transistors.

5.2 Basic relations for two-port networks

5.2.1 Scattering parameter theory

Small signal equivalent circuits for semiconductor devices and circuits are usually represented in two-port form as shown in Figure 5.2. At low frequencies, two-port networks are represented by an impedance matrix, an admittance matrix, a hybrid matrix or a chain matrix. These matrices are described by Z, Y, *h* or ABCD parameters. Such representations are suitable at low frequencies where the parameters may be measured by placing short or open circuits at the input and output terminals of the two-port. At high (microwave) frequencies where there are travelling waves, short and open circuits

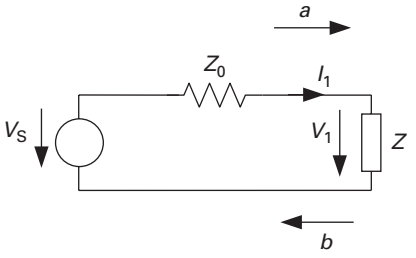

Fig. 5.1 Source and load circuit for the S-parameter discussion. a and b are normalised power waves.

cannot be precisely placed and the above-mentioned matrix representations cannot be accurately determined. The development of the vector network analyser made it possible to perform measurements of high frequency travelling wave circuits. Scattering parameters were introduced by Kurokawa [24]. He defined a set of *normalised power waves* a and b and introduced a normalising impedance Z_0:

$$a = \frac{V + Z_0 I}{2\sqrt{Z_0}} \tag{5.1}$$

$$b = \frac{V - Z_0 I}{2\sqrt{Z_0}}, \tag{5.2}$$

where V and I are the voltage and the current, respectively, at a load Z.

These normalised power waves are chosen in this way so that they relate to the power delivered to the load. Refer to Figure 5.1 where a source with a real source impedance Z_0 is connected to an arbitrary load $\underline{Z}$. The maximum power is delivered to the load if $\underline{Z} = Z_0^* = Z_0$. Then,

$$P_{Z,\max} = \frac{V_1^2}{Z_0} = \frac{V_S^2}{4Z_0} = |a|^2.$$

$|a|^2$ is hence the available power from a generator with source impedance Z_0. This also tells us that the unit of a (and b) is $\sqrt{W}$.

Now consider:

$$|a|^2 - |b|^2 = \Re\{V_1 I_1^*\}. \tag{5.3}$$

This is the power delivered to the load for arbitrary $\underline{Z}$. We can hence interpret $|a|^2$ as the power travelling towards the load, and $|b|^2$ as the power travelling from the load back to the generator, with the difference dissipated in the load. The ratio of b to a is the reflection coefficient:

$$\Gamma = \frac{b}{a} = \frac{V_1 - Z_0 I_1}{V_1 + Z_0 I_1} = \frac{\underline{Z} - Z_0}{\underline{Z} + Z_0}, \tag{5.4}$$

because $\underline{Z} = V_1 / I_1$.

We now expand the normalised power wave concept to a two-port. The incident waves at each port are again designated as a and the reflected waves are designated as b, while the subscript denotes the port where the power waves are measured. For the two-port, the normalised power waves are a_1, b_1, a_2 and b_2, as shown in Figure 5.2. The

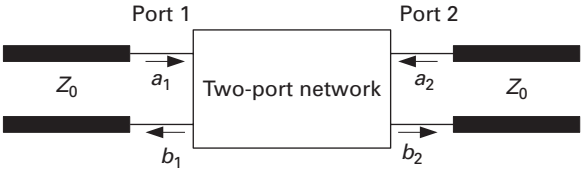

Fig. 5.2 Two-port network embedded in a transmission line with characteristic impedance Z_0, with incident waves a_1, a_2 and reflected waves b_1, b_2.

scattering parameters are the coefficients of linear equations relating the reflected waves b to the incident waves a:

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{5.5}$$

$$b_2 = S_{21}a_1 + S_{22}a_2. \tag{5.6}$$

The scattering parameters can therefore be expressed as the ratio of two power waves, provided that all the ports are terminated in a non-reflective fashion ($a = 0$ at all other ports). For a two-port,

$$S_{11} = \frac{b_1}{a_1}|_{a_2=0} \tag{5.7}$$

$$S_{12} = \frac{b_1}{a_2}|_{a_1=0} \tag{5.8}$$

$$S_{21} = \frac{b_2}{a_1}|_{a_2=0} \tag{5.9}$$

$$S_{22} = \frac{b_2}{a_2}|_{a_2=0}. \tag{5.10}$$

These relations can be written in matrix form as follows:

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \tag{5.11}$$

Let Z_0 be the characteristic impedance of the transmission lines connected to ports 1 and 2 of the two-port. If port 2 is terminated by Z_0, there is no reflection at the load and hence the wave incident at port 2, a_2, is zero. Similarly, if port 1 is terminated by Z_0 and the stimulus is fed to port 2, a_1, is zero. If port 1 is designated to be the input and port 2 the output, then S_{11} is the input reflection coefficient with Z_0 the output, S_{22} is the output reflection coefficient with the input terminated by Z_0, S_{21} is the forward transmission coefficient with Z_0 as the output load, and S_{12} is the reverse transmission coefficient with Z_0 at the input.

In general, the scattering parameters are complex. The polar form of the scattering parameter is useful in many applications:

$$S = \mid S \mid \exp^{\mathrm{j}\theta}, \tag{5.12}$$

where $\mid S \mid$ is the magnitude of S and θ is the phase.

Properties of scattering parameters

The following properties of the scattering parameters are important in two-port network applications. Subsequently, it is assumed that the transmission line is lossless with negligible attenuation, such that the line's complex propagation constant $\gamma = \alpha + \jmath\beta \approx \jmath\beta$.

(i) *Reciprocity*. Passive networks are reciprocal (unless they contain non-reciprocal components like isolators or circulators), and the S-parameters satisfy

$$S_{jk} = S_{kj}. \tag{5.13}$$

This property can be written in general form as

$$[S] = [S^T]. \tag{5.14}$$

It states that the matrix is equal to its transpose denoted by $[S^T]$.

(ii) *Lossless networks.* An important property of lossless networks is that the product of the transposed complex conjugate scattering matrix and the scattering matrix is equal to the unitary matrix.

$$[S][S^T]^* = [I], \tag{5.15}$$

where

$$[I] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.16}$$

defines the unitary matrix.

(iii) *Lossy networks.* In lossy networks, the network itself dissipates power, hence

$$\sum |\, a_k \,|^2 \; > \sum |\, b_k \,|^2 \,. \tag{5.17}$$

The scattering matrix satisfies the property

$$[I] - [S][S^T]^* > 0. \tag{5.18}$$

(iv) *Reference planes.* Measurements can be made at different planes along the transmission lines connected to the two-port network; this changes the results due to the signals' travelling wave nature. The reference plane is the position where the actual measurements are made. If the positions of the ports are shifted by electrical distances $\beta\ell$ away from the reference planes, the S-parameters in these shifted planes can be related to the initial S-parameters in the reference plane.

If the S-parameters were measured originally at the planes $z_1 = 0$ and $z_2 = 0$ and if the reference planes are now at $z_1 = \ell_1$ and $z_2 = \ell_2$ as in Figure 5.3, the resulting S-matrix is given by

$$\begin{bmatrix} S'_{11} & S'_{12} \\ S'_{21} & S'_{22} \end{bmatrix} = \begin{bmatrix} S_{11}\exp^{-j2\theta_1} & S_{12}\exp^{-j(\theta_1+\theta_2)} \\ S_{21}\exp^{-j(\theta_1+\theta_2)} & S_{22}\exp^{-j2\theta_2} \end{bmatrix}, \tag{5.19}$$

where

$$\theta_1 = \beta\ell_1$$
$$\theta_2 = \beta\ell_2.$$

The expression (5.19) assumes that the transmission lines are lossless.

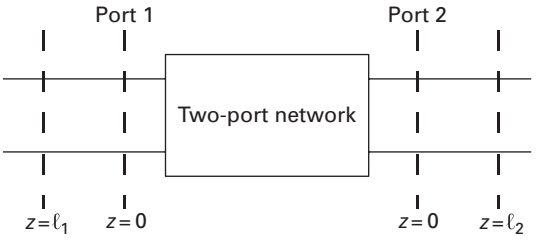

Fig. 5.3 Change of reference planes.

The scattering parameters can be converted to current–voltage parameters such as impedance ($[Z]$) parameters as well as admittance ($[Y]$) parameters. These conversions are given by Gonzalez [14].

5.2.2 The Smith chart

The analysis of two-port networks at microwave frequencies was tedious and time-consuming before speedy computation methods were available with computer-aided design software. A graphical aid to calculate various network properties such as impedances was developed by Smith [36, 37] and referred to as the *Smith chart*. The accuracy of results obtained from the Smith chart is quite adequate in most cases.

When a transmission line is terminated in an arbitrary impedance Z, there are reflections along the line, and the reflection is defined as the ratio of the voltage in the wave reflected from the terminating load to the voltage in the wave incident on the terminating load. In Equation (5.4), we had already defined the reflection coefficient Γ, which can be expressed as

$$\Gamma = \frac{Z - Z_0}{Z + Z_0}. \tag{5.20}$$

In most applications, it is convenient to use the value of the impedance normalised to the characteristic impedance or other reference impedance. The normalised value is given by

$$z = \frac{Z}{Z_0}. \tag{5.21}$$

Equation (5.20) can now be written as

$$\Gamma = \frac{z - 1}{z + 1}. \tag{5.22}$$

Both z and Γ are complex quantities and they are written in terms of their real and imaginary parts:

$$\Gamma = u + jv \tag{5.23}$$

$$z = r + jx. \tag{5.24}$$

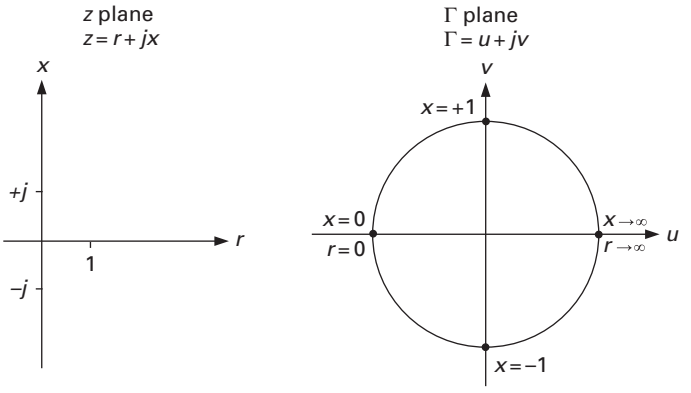

Fig. 5.4 Mapping between the z plane and the Γ plane.

It is often useful to express the reflection coefficient in polar coordinates.

$$\Gamma = |\Gamma| e^{j\theta}, \tag{5.25}$$

where $|\Gamma|$ is the magnitude, and θ is the phase of the reflection coefficient.

It follows that z and Γ are defined in two complex planes. Equation (5.22) gives the relationship between points in the two complex planes. The relationship is known as *mapping*. Equation (5.22) is a bilinear transformation where orthogonal lines in the z plane map into orthogonally intersecting circles in the Γ plane. Furthermore, it is a conformal mapping whereby the angle between the two line segments is maintained in mapping between the z and Γ planes. It is to be noted that a straight line is simply a degenerate circle. Figure 5.4 shows the correspondence between points in the z and Γ planes. Expanding Equation (5.22) we have

$$u + jv = \frac{r + jx - 1}{r + jx + 1}. \tag{5.26}$$

Equating real and imaginary parts on both sides,

$$u = \frac{r^2 - 1 + x^2}{(r+1)^2 + x^2} \tag{5.27}$$

$$v = \frac{2x}{(r+1)^2 + x^2}. \tag{5.28}$$

By eliminating x from Equations (5.27) and (5.28), we have

$$\left(u - \frac{r}{r+1}\right)^2 + v^2 = \left(\frac{1}{r+1}\right)^2. \tag{5.29}$$

By eliminating r from Equations (5.27) and (5.28), we obtain

$$(u-1)^2 + \left(v - \frac{1}{x}\right)^2 = \left(\frac{1}{x}\right)^2. \tag{5.30}$$

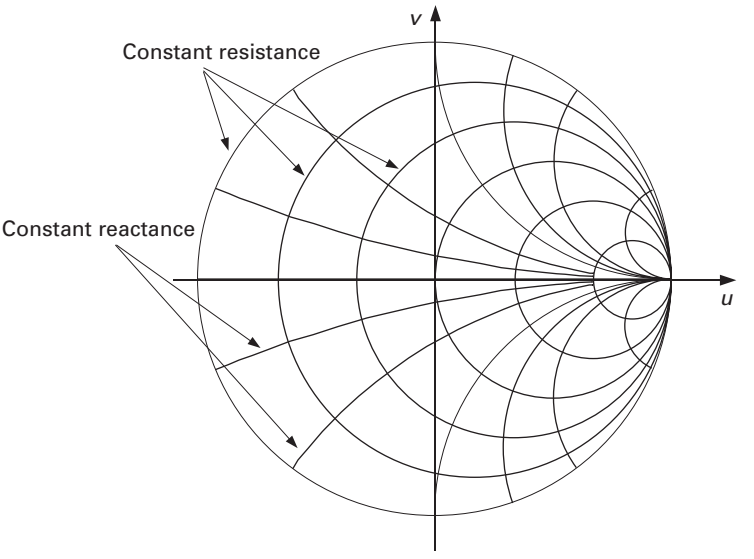

Fig. 5.5 Constant resistance and reactance circles in the Γ plane.

Equation (5.29) is the equation of a family of circles with their centres at $u = r/r + 1$, $v = 0$ and radii equal to $1/r + 1$, while Equation (5.30) is the equation of a family of circles with their centres at $u = 1$, $v = 1/x$ and radii equal to $1/x$. Equation (5.29) represents constant resistance circles; each value of $r \geq 1$ represents a circle. Equation (5.30) represents constant reactance circles which are plotted for all values of z when $\mathrm{Re}(z) \geq 0$. Both constant resistance and constant reactance circles are shown in Figure 5.5. A typical Smith chart representation for practical use is shown in Figure 5.6.

The Smith chart can also be used as an admittance chart. The constant resistance circles become the constant conductance circles and the constant reactance circles become the constant susceptance circles. The bilinear transformation in this case is

$$\Gamma = \frac{1 - y}{1 + y}. \tag{5.31}$$

The impedance and admittance representations of the Smith chart are symmetric with respect to the origin of the Smith chart. Because in typical impedance-matching problems, both impedance and admittance representations are needed, they are frequently plotted in the same diagram (see Figure 5.7).

5.2.3 Impedance matching

To maximise gain in an amplifier, its input and output impedances must be chosen to be the complex conjugate of the generator and load impedances, respectively. To achieve this, impedance-matching networks are connected at the input and the output of the amplifier, which convert the true generator and load impedances (frequently 50 Ω) to the necessary values. Figure 5.8 shows the block diagram of the complete network.

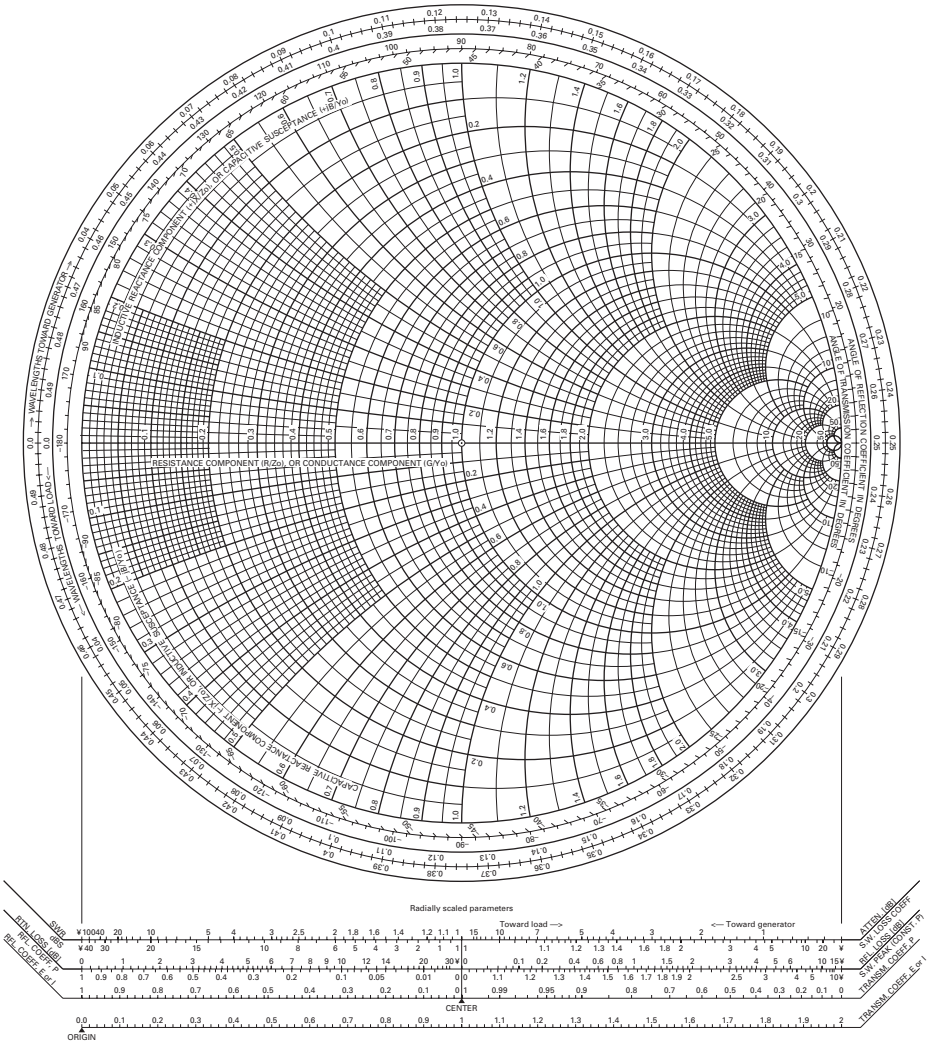

Fig. 5.6 Practical Smith chart.

Aside from the complex conjugate match, also referred to as *power match*, impedance transformation may be necessary to achieve minimum noise figure, or maximum output power. Detailed discussions of these techniques (with diverging goals) are given later in this chapter.

5.2.4 Power gains for amplifier design

The power gain in microwave circuits is expressed in terms of the scattering parameters for convenience in performing calculations using network analyser measurements.

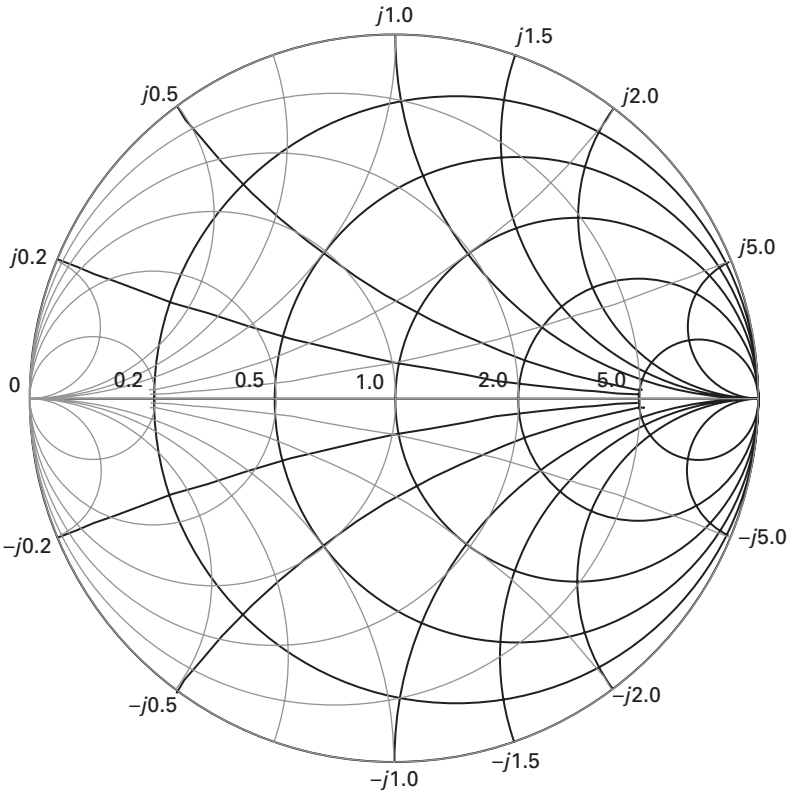

Fig. 5.7 Smith chart showing impedance circles (bold lines) and admittance circles (thin lines).

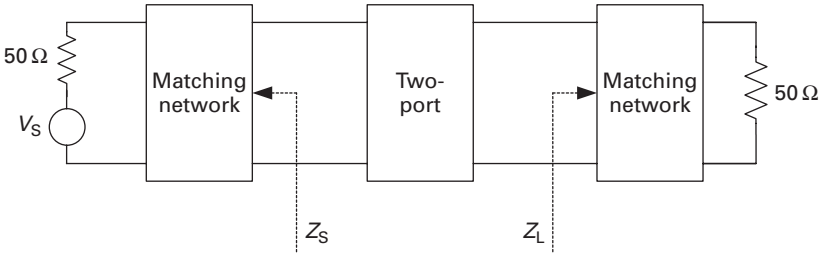

Fig. 5.8 Block diagram of a two-port embedded in impedance matching networks.

Input and output reflection coefficients for arbitrary terminations

The two-port network in Figure 5.9 is now assumed to be an amplifier circuit. We will first calculate the input and output reflection coefficients Γ_{in} and Γ_{out}, respectively, for arbitrary source and load reflection coefficients.

If Z_L is the load impedance in a transmission line system of characteristic impedance Z_0, the reflection coefficients at the source and load (Figure 5.9) are given by

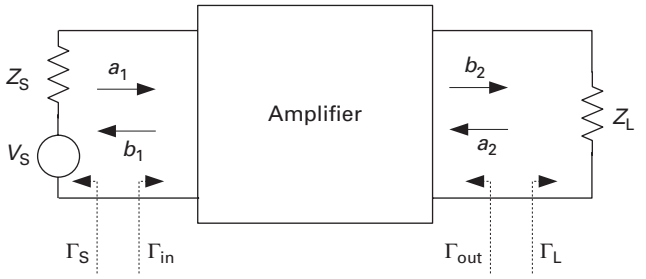

Fig. 5.9 Amplifier representation.

$$\Gamma_S = \frac{Z_S - Z_0}{Z_S + Z_0} \tag{5.32}$$

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0}. \tag{5.33}$$

Writing Equation (5.11) in expanded form, we have

$$b_1 = S_{11}a_1 + S_{12}a_2 \tag{5.34}$$

$$b_2 = S_{21}a_1 + S_{22}a_2. \tag{5.35}$$

It is evident from Figure 5.9 that reflection coefficients are related to the incident and reflected waves by the following equations:

$$\Gamma_{in} = \frac{b_1}{a_1} \tag{5.36}$$

$$a_2 = \Gamma_L b_2. \tag{5.37}$$

By substitution into Equation (5.11), it can be shown that

$$\Gamma_{in} = S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L}. \tag{5.38}$$

The output reflection coefficient is defined as

$$\Gamma_{out} = \frac{b_2}{a_2} \,|_{V_S=0}\,. \tag{5.39}$$

By substitution in Equation (5.11), it can be also shown that

$$\Gamma_{out} = S_{22} + \frac{S_{12}S_{21}\Gamma_S}{1 - S_{11}\Gamma_S}. \tag{5.40}$$

Powers at input and load

When designing amplifiers, different definitions of power gain are applied depending on the application. It is necessary to define the power at the input and the output of the amplifier in order to obtain the gain.

The power delivered to the input port of the amplifier is written in terms of the incident and reflected waves at the input. The reflected wave can be expressed in terms of the incident wave and the reflection coefficient, see Equation (5.36).

$$P_{\text{in}} = |\, a_1 \,|^2 - |\, b_1 \,|^2 \tag{5.41}$$

$$= |\, a_1 \,|^2 \left(1 - |\, \Gamma_{\text{in}} \,|^2\right). \tag{5.42}$$

Similarly, the power delivered to the load Z_{L} is

$$P_{\text{L}} = |\, b_2 \,|^2 - |\, a_2 \,|^2 \tag{5.43}$$

$$= |\, b_2 \,|^2 \left(1 - |\, \Gamma_{\text{L}} \,|^2\right). \tag{5.44}$$

The power travelling towards the load is partly due to the wave originated by the generator electromotive force b_{S}, and partly by the reflection from the source impedance Γ_{S}:

$$a_1 = b_{\text{S}} + \Gamma_{\text{S}} b_1 \tag{5.45}$$

with

$$b_{\text{S}} = \frac{V_{\text{S}}\sqrt{Z_0}}{Z_{\text{S}} + Z_0} \tag{5.46}$$

$$\Gamma_{\text{S}} = \frac{Z_{\text{S}} - Z_0}{Z_{\text{S}} + Z_0}. \tag{5.47}$$

Detailed analyses are given by Gonzalez [14].

The power available from the source is labelled P_{avs} and is defined as

$$P_{\text{avs}} = \frac{|\, b_{\text{S}} \,|^2}{1 - |\, \Gamma_{\text{S}} \,|^2} \tag{5.48}$$

$$= |a_1|^2 \frac{|1 - \Gamma_{\text{S}}\Gamma_{\text{in}}|^2}{1 - |\Gamma_{\text{S}}|^2}.$$

Note that P_{avs} is the input power under conjugate match, i.e. when $\Gamma_{\text{in}} = \Gamma_{\text{S}}^{\star}$.

The power available from the network P_{avn} is defined as the power delivered by the network when the load is conjugately matched to the output impedance, or

$$P_{\text{avn}} = \frac{|S_{21}|^2 |b_{\text{S}}|^2}{|1 - S_{11}\Gamma_{\text{S}}|^2 (1 - |\Gamma_{\text{out}}|^2)}. \tag{5.49}$$

Power gain definitions

Let us now discuss the power gain definitions which are important in amplifier design. The operating power gain G_{P} is the ratio of the power delivered to the load to the power delivered to the input of the amplifier:

$$G_{\text{P}} = \frac{P_{\text{L}}}{P_{\text{in}}}$$

$$= \left|\frac{b_2}{a_1}\right|^2 \frac{1 - |\, \Gamma_{\text{L}} \,|^2}{1 - |\, \Gamma_{\text{in}} \,|^2}.$$

Since

$$\frac{b_2}{a_1} = \frac{S_{21}}{1 - \Gamma_{\text{L}} S_{22}},$$

and using Equation (5.38), we obtain

$$G_{\mathrm{P}} = |S_{21}|^2 \frac{1 - |\Gamma_{\mathrm{L}}|^2}{|1 - \Gamma_{\mathrm{L}} S_{22}|^2 - |S_{11} - \Gamma_{\mathrm{L}} \Delta(S)|^2}, \tag{5.50}$$

where $\Delta(S)$ is the determinant of the scattering matrix.

The transducer power gain G_{T} is the ratio of the power delivered to the load to the power available from the source:

$$\begin{aligned} G_{\mathrm{T}} &= \frac{P_{\mathrm{L}}}{P_{\mathrm{avs}}} \\ &= \left|\frac{b_2}{a_1}\right|^2 \frac{(1 - |\Gamma_{\mathrm{L}}|^2)(1 - |\Gamma_{\mathrm{S}}|^2)}{|1 - \Gamma_{\mathrm{in}} \Gamma_{\mathrm{S}}|^2} \\ &= |S_{21}|^2 \frac{(1 - |\Gamma_{\mathrm{L}}|^2)(1 - |\Gamma_{\mathrm{S}}|^2)}{|1 - \Gamma_{\mathrm{L}} S_{22} - \Gamma_{\mathrm{S}} (S_{11} - \Gamma_{\mathrm{L}} \Delta(S))|^2}. \end{aligned} \tag{5.51}$$

In the absence of deliberate feedback, the reverse transmission S_{12} is frequently very small and can be neglected. $S_{12} = 0$ simplifies the denominator in Equation (5.51) and we obtain the unilateral transducer power gain

$$G_{\mathrm{TU}} = |S_{21}|^2 \frac{(1 - |\Gamma_{\mathrm{L}}|^2)(1 - |\Gamma_{\mathrm{S}}|^2)}{|(1 - S_{11}\Gamma_{\mathrm{L}})(1 - S_{22}\Gamma_{\mathrm{S}})|^2}. \tag{5.52}$$

The available power gain G_{A}, finally, is the ratio of the available gain from the network P_{avn} to the available power from the source P_{avs}:

$$\begin{aligned} G_{\mathrm{A}} &= \frac{P_{\mathrm{avn}}}{P_{\mathrm{avs}}} \\ &= |S_{21}|^2 \frac{1 - |\Gamma_{\mathrm{S}}|^2}{|1 - S_{11}\Gamma_{\mathrm{S}}|^2 (1 - |\Gamma_{\mathrm{out}}|^2)} \\ &= |S_{21}|^2 \frac{1 - |\Gamma_{\mathrm{S}}|^2}{|1 - S_{11}\Gamma_{\mathrm{S}}|^2 - |S_{22}(1 - S_{11}\Gamma_{\mathrm{S}}) + S_{12}S_{21}\Gamma_{\mathrm{S}}|^2}, \end{aligned} \tag{5.53}$$

using Equation (5.40).

5.2.5 Stability

G_{A} still contains Γ_{S} as a variable. We know already that the maximum power transfer from the source to the load occurs for $\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{in}}^{\star}$, so this appears to be an optimum choice. Before we proceed, however, let us again investigate Equation (5.53). If the denominator becomes zero, the available gain would grow beyond all bounds. This happens if

$$|1 - S_{11}\Gamma_{\mathrm{S}}| = |S_{22}(1 - S_{11}\Gamma_{\mathrm{S}}) + S_{12}S_{21}\Gamma_{\mathrm{S}}|,$$

or, written differently,

$$\left| S_{22} + \frac{S_{12}S_{21}\Gamma_{\mathrm{S}}}{1 - S_{11}\Gamma_{\mathrm{S}}} \right| = 1 = |\Gamma_{\mathrm{out}}|,$$

using Equation (5.40). Likewise, we can show from the available gain in the reverse direction that it would grow beyond all bounds for $|\Gamma_{in}| = 1$.

Both these conditions are considered as the *instability* of the amplifier, a potentially dangerous situation which may lead to malfunction or even fatal failure. In most cases (the notable exception are oscillators), it needs to be avoided.

Unconditional stability

From the above, we can deduce that a two-port will be *unconditionally stable* if

$$\left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1, \tag{5.54}$$

for all $|\Gamma_L| \leq 1$, and

$$\left| S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1 - S_{11}\Gamma_G} \right| < 1, \tag{5.55}$$

for all $|\Gamma_G| \leq 1$.

Stability circles

The following discussion follows Hoffmann [20]. For conditionally stable two-ports – where at least one of the conditions in Equations (5.54) and (5.55) is violated – we can still find generator and load admittances which allow stable operation. If we plot the locus of

$$\left| S_{22} + \frac{S_{12}S_{21}\Gamma_G}{1 - S_{11}\Gamma_G} \right| = 1$$

in the complex Γ_G plane, we obtain a circle with centre vector

$$\Gamma_{G,C} = \frac{\Delta(S)^* S_{22} - S_{11}^*}{|\Delta(S)|^2 - |S_{11}|^2}. \tag{5.56}$$

The radius is

$$r_G = \frac{|S_{12}\, S_{21}|}{\left||\Delta(S)|^2 - |S_{11}|^2\right|}. \tag{5.57}$$

Likewise, plotting the locus of

$$\left| S_{11} + \frac{S_{12}S_{21}\Gamma_L}{1 - S_{22}\Gamma_L} \right| = 1$$

in the complex Γ_L plane, we obtain a circle with centre vector

$$\Gamma_{L,C} = \frac{\Delta(S)^* S_{11} - S_{22}^*}{|\Delta(S)|^2 - |S_{22}|^2}, \tag{5.58}$$

and radius

$$r_G = \frac{|S_{12}\, S_{21}|}{\left||\Delta(S)|^2 - |S_{22}|^2\right|}. \tag{5.59}$$

Examples for stability circles in the generator and load planes are shown in Figure 5.10. The stability circles provide the boundaries between the stable and the unstable regions; however, we still need to determine whether the inside or the outside of the circle is stable. Let us do this for the Γ_G plane first.

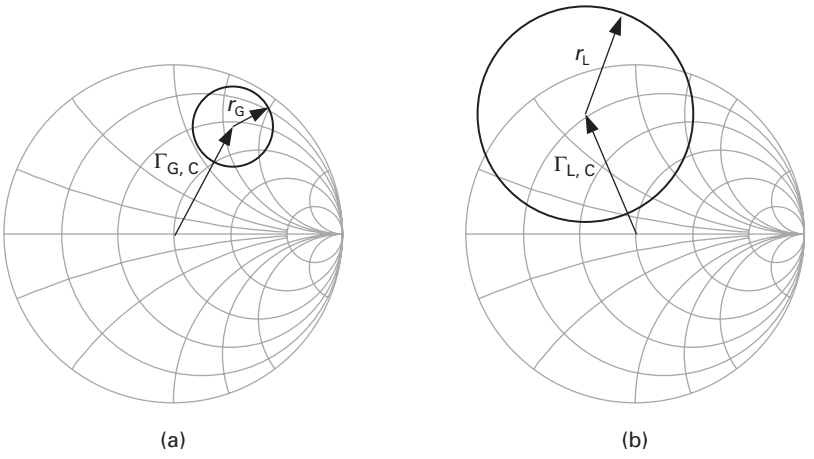

Fig. 5.10 Stability circle examples: (a) in the Γ_G plane and (b) in the Γ_L plane.

Using Equation (5.40), we conclude that $\Gamma_{out}(\Gamma_G = 0) = S_{22}$. Now we locate the area which contains $\Gamma_G = 0$ (the centre of the Smith chart). We can say:

- If $|S_{22}| < 1$, then the region which contains $\Gamma_G = 0$ is the stable region.
- Otherwise, if $|S_{22}| > 1$, the region containing $\Gamma_G = 0$ is the unstable region.

The same procedure applies for finding the stable region in the Γ_L plane.

Using the stability circles, we can look at stability in a different way. A two-port will be unconditionally stable, if all of the following conditions are fulfilled:

(i) $|S_{11}| < 1$
(ii) $|S_{22}| < 1$
(iii) $|\Gamma_{L,C}| > 1 + r_L$
(iv) $|\Gamma_{G,C}| > 1 + r_G$.

In this case, both stability circles are fully outside of the Smith chart unity circles in the reflection coefficient plane (which contain all $|\Gamma_G|, |\Gamma_L| < 1$), and the outer regions of the circles are the stable ones.

Rollet's stability factor

Conditions (iii) and (iv) above can be combined into the following:

$$k > 1 + \max\left[0, \frac{|\Delta(S)|^2 - 1}{2|S_{12}||S_{21}|}\right], \tag{5.60}$$

where

$$k = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta(S)|^2}{2|S_{12}||S_{21}|}. \tag{5.61}$$

k is the *Rollet factor*. $k > 1$ is often used as a stability criterion; however, note that it is a necessary, but not sufficient requirement for unconditional stability.

A very common, and potentially fatal, mistake is to assess stability only for the intended frequency of operation. It must be investigated over the full frequency range where instability may conceivably occur – parasitic oscillations at very low frequencies are extremely common, and instabilities may also increase with increasing frequencies such as in cascode amplifiers (see p. 333).

5.2.6 Maximum available gain and maximum stable gain

Let us now reassess the case of an amplifier with simultaneous complex conjugate match at the input and output ports, i.e. $\Gamma_S = \Gamma^{\star}_{in}, \Gamma_L = \Gamma^{\star}_{out}$. The available gain in this case is the *maximum available gain* and can be written using the Rollet factor Equation (5.61) as

$$MAG = \left|\frac{S_{21}}{S_{12}}\right| \left(k - \sqrt{k^2 - 1}\right). \tag{5.62}$$

Obviously, a real solution exists only if $k \geq 1$.

For $k < 1$, the *maximum stable gain* is frequently quoted. This is the MAG in the limit of $k = 1$, which can be obtained by the resistive loading of an otherwise not conditionally stable amplifier:

$$MSG = \left|\frac{S_{21}}{S_{12}}\right|. \tag{5.63}$$

5.2.7 Mason's unilateral gain

The concept of unilateral gain for a two-port network was first introduced by Mason in his paper [28], which is now considered a classic. A comprehensive review of the paper and its relevance today is given by Gupta [15].

The unilateral gain is defined as the maximum power gain obtained by a two-port when it is made unilateral – unilateralised. A two-port network that includes an active device is made unilateral by a lossless and reciprocal four-port network connected to input and output of the two-port under investigation. This network provides the necessary feedback to impose the unilateral condition. Mason's unilateral gain is not to be confused with the unilateral transducer power gain G_{TU}, Equation (5.52), which had been derived by *neglecting* the reverse transmission. Here, reverse transmission is eliminated by a unilateralising network.

The unilateral gain is a figure of merit which is intrinsic to the device and hence independent of the circuit in which the device is placed. The unilateral gain is, therefore, an invariant property of the device. Hence, it can be expressed in terms of the device's small-signal parameters. The scattering parameter representation of U is the most useful in microwave applications. Ku [23] has given an expression in terms of the S-matrix:

$$U = \frac{|\, S_{12} - S_{21}\,|^2}{\Delta([I] - [S^*]^T[S])}, \tag{5.64}$$

where $[I]$ is the identity matrix. U can also be expressed with the help of the stability factor k [15]:

$$U = \frac{\left|\left(\frac{S_{21}}{S_{12}}\right) - 1\right|^2}{2k\left|\frac{S_{21}}{S_{12}}\right| - 2\mathrm{Re}\left[\frac{S_{21}}{S_{12}}\right]}. \tag{5.65}$$

5.2.8 Maximum frequency of oscillation

The maximum frequency of oscillation is a criterion for a device's ability to amplify power. Because the maximum power gain at any frequency is obtained by conjugately matching the input and output ports, the MAG (see Equation (5.62)) is an obvious choice. The maximum frequency of oscillation $f_{\max}$ is then the frequency where MAG drops to unity:

$$\mathrm{MAG}(f_{\max}) = 1. \tag{5.66}$$

However, we had seen that MAG only exists when Rollet's stability factor $k \geq 1$. This raises a practical problem – for many high-performance microwave transistors, $k < 1$ in the whole measurement range, and so $f_{\max}$ cannot be determined using Equation (5.66).

Another customary definition therefore makes use of Mason's unilateral gain U (see Equation (5.65)), which does not have this limitation. $f_{\max}$ is then understood as the frequency where U drops to unity:

$$U(f_{\max}) = 1. \tag{5.67}$$

Note, however, that Equations (5.66) and (5.67) will generally not yield the same result, so it is important to check the definition used when comparing $f_{\max}$ values.

5.3 Noise in two-ports

5.3.1 Noise phenomena

Any electronic component exhibits electronic noise, provided that the absolute temperature is $T > 0$. There are several physical origins, which have been discussed in Chapter 2 in the context of active devices, so only a brief summary shall be given here.

Noise occurs in all contexts where carrier motion or carrier density is stochastic:

- As there is always a random thermal motion superimposed on any charge carrier movement, *thermal* or *Johnson noise* is omnipresent in all conductors with non-zero resistance. The rms voltage generated by the thermal noise in a resistor of value R is $V_{\mathrm{rms}} = \sqrt{4kTRB}$, where k is Boltzmann's constant and B is the measurement bandwidth.
- Emission of charge carriers over energy barriers is equally a stochastic process, and its associated noise mechanism is called *shot noise*. The rms current generated by a current of magnitude I flowing over an energy barrier is $I_{\mathrm{rms}} = \sqrt{2qIB}$, where q is the elementary charge and B again the measurement bandwidth.

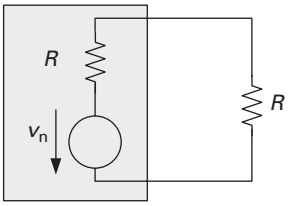

Fig. 5.11 Thévenin equivalent circuit of a noise resistor terminated by a load of equal value.

- Examples of noise due to random fluctuations in carrier density is *generation-recombination noise* or noise due to *trapping and de-trapping processes*. Unlike the first two noise phenomena, which can be considered to have a spectral density independent of frequency ('white noise'), these processes produce noise spectra with low-pass behaviour, and cutoff frequencies in the Hz–MHz range.
- Another example of noise generated by random changes in carrier densities is *avalanche noise*, due to carrier multiplication effects in high-field regions. This process also produces a low-pass limited noise spectrum with a cutoff frequency in the GHz range.

For the discussion of noise in linear two-ports, however, the physical origin of noise is irrelevant. In fact, we will occasionally assume in the following that noise is always thermal in nature. Thermal noise has an interesting property. Consider that the squared magnitude of the noise voltage phasor generated in a resistor R in a bandwidth B is

$$\left\langle |v_n|^2 \right\rangle = 8kTRB, \tag{5.68}$$

which corresponds to the rms value of $\sqrt{4kTRB}$ mentioned above. The Thévenin equivalent circuit of the noise resistor can then be drawn as in the box of Figure 5.11. The source is terminated by the same resistance R, so that the *available power* is delivered. As power is related to the peak voltage as $0.5\widehat{V}^2/R$, and the voltage drop across the resistor is $v_n/2$, the resulting available power due to the thermal noise in resistor R in a bandwidth B is

$$N = \frac{\left\langle |v_n|^2 \right\rangle}{8R} = kTB. \tag{5.69}$$

The available noise power of a resistor is hence independent of the resistor value and depends only on the absolute temperature T and the measurement bandwidth B.

5.3.2 Noise figure

We will now, in a general form, consider what happens to a signal as it traverses a noisy two-port. Figure 5.12 shows an arrangement where the two-port is connected to a generator (source) with generator resistance R_G and a load R_L. For now, we assume that we deal only with real impedances and that both generator and load constitute a power match, i.e. $R_{in} = R_G$, $R_{out} = R_L$. The condition of power match at the input will be dropped further down.

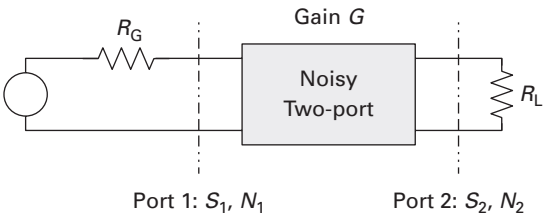

Fig. 5.12 Generic noisy two-port connected to source and load.

We further assume that the noise at the input is only due to the thermal noise of the generator resistance, and that this resistor is at a temperature of $T_0 = 290\,\mathrm{K}$.[1] The noise power at the input port is then

$$N_1 = kT_0B. \tag{5.70}$$

In a 1 Hz bandwidth, this amounts to $4 \cdot 10^{-21}$ W or -174 dBm.[2]

The two-port will also contribute noise. To simplify things, let us assume that all noise sources inside the two-port combine into a single noise source with power N_{eq}, also located at port 1. Because the two-port has a gain of G, the noise power at the output is $N_2 = G \cdot (N_1 + N_{\mathrm{eq}})$.

If there is an additional signal component S_1 present at the input, it equally is magnified by G. $S_2 = G \cdot S_1$. We can now define the ratio of the *signal-to-noise ratios* at the input and the output:

$$\frac{S_1/N_1}{S_2/N_2} = \frac{S_1G(N_1 + N_{\mathrm{eq}})}{N_1GS_1} = 1 + \frac{N_{\mathrm{eq}}}{N_1} = F \tag{5.71}$$

This defines the *noise figure* F of the two-port as the ratio of the signal-to-noise ratios at the input and the output, provided that the input carries thermal noise at a temperature $T_0 = 290\,\mathrm{K}$ only. Note that $F > 1$ under all circumstances.

We also found the relationship between the equivalent noise power at the input and the noise figure:

$$N_{\mathrm{eq}} = (F - 1)N_1 = (F - 1)kT_0B. \tag{5.72}$$

If we assume that N_{eq} is also thermally generated, $N_{\mathrm{eq}} = kT_{\mathrm{n}}B$, we can define an *equivalent noise temperature* for the two-port:

$$T_{\mathrm{n}} = (F - 1)T_0. \tag{5.73}$$

Both F and T_{n} can be used interchangeably to characterise the noise performance of two-ports. The use of T_{n} is customary in cases where F is very close to 1, for example in low-noise amplifiers (LNAs) for satellite applications, while F is more popular for general applications.

[1] 290 K is the standard temperature for noise calculations.
[2] 1 dBm is one decibel relative to a power of 1 mW.

Fig. 5.13 Cascaded noisy two-ports.

Noise figure of cascaded two-ports

The situation depicted in Figure 5.12 is too simplistic for practical applications, because amplifiers or receivers generally consist of several stages. Let us consider next what happens when several noisy two-ports are being cascaded. This is shown in Figure 5.13. Again we assume that all ports are power-matched, which is an important restriction, but is made here to simplify calculations.

First, we calculate the noise at the output (delivered to the load R_L), assuming that for each two-port, the noise sources are combined into equivalent noise sources at its input. Then,

$$N_2 = G_3 \left\{ N_{\mathrm{eq},3} + G_2 \left[N_{\mathrm{eq},2} + G_1 \left(N_{\mathrm{eq},1} + N_1 \right) \right] \right\}. \tag{5.74}$$

Now we assume that all noise sources are being transferred to the input of the cascade. The equivalent noise source there will have a noise power of

$$N_{\mathrm{eq,tot}} = N_{\mathrm{eq},1} + \frac{N_{\mathrm{eq},2}}{G_1} + \frac{N_{\mathrm{eq},3}}{G_1 G_2}. \tag{5.75}$$

Recalling Equation (5.72), we finally calculate the noise figure of the cascaded two-ports:

$$F_{\mathrm{tot}} = 1 + \frac{N_{\mathrm{eq,tot}}}{N_1} = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2}. \tag{5.76}$$

In a more generalised form, the total noise figure of a cascade of n stages is

$$F_{\mathrm{tot}} = F_1 + \sum_{i=2}^{n} \frac{F_{\mathrm{i}} - 1}{\prod_{k=1}^{i-1} G_{\mathrm{k}}} \tag{5.77}$$

This is the famed *Friis equation* [11] which in effect postulates that in a receiver chain the overall noise figure is approximately the noise figure of the first stage, provided that that stage has sufficient gain – an important observation for the design of LNAs.

A consequence of the Friis formula is that in any amplifier, a low noise figure without sufficient gain is meaningless as the noise of the following stages will take over. It is therefore useful to combine noise figure and gain in a single figure of merit. This is the *noise measure* according to Haus and Adler [19]. When using noise figures, it is written as

$$M = \frac{F - 1}{1 - 1/G}, \tag{5.78}$$

where F is the noise figure and G the gain. The noise figure of an infinite chain of identical transistors is then

$$F_\infty = M + 1. \tag{5.79}$$

The noise measure can also be written in the form of a noise temperature:

$$M_\mathrm{T} = \frac{T_\mathrm{n}}{1 - 1/G}, \tag{5.80}$$

where T_n is the equivalent noise temperature of the two-port (see Equation (5.73)).

5.3.3 Noise figure with arbitrary generator admittance

We will now abandon the condition that all ports are power-matched by allowing arbitrary terminations for the input port. The source admittance $\underline{Y}_\mathrm{S}$ is expressed by its real and imaginary parts, $\underline{Y}_\mathrm{S} = G_\mathrm{S} + \jmath B_\mathrm{S}$. As before, we assume that all noise sources internal to the two-port can be combined at the input. Due to the arbitrary generator admittance, we now have to split the noise sources into a noise voltage source v and a noise current source i. This is shown in Figure 5.14.

The equivalent noise voltage and current sources, v and i, are partially correlated. This can be accounted for by introducing a *correlation admittance* $\underline{Y}_\mathrm{C}$ and splitting the current source into an uncorrelated part i_u and a fully correlated part $\underline{Y}_\mathrm{C} v$:

$$i = i_\mathrm{u} + \underline{Y}_\mathrm{C} v. \tag{5.81}$$

The real part of the generator admittance contributes a thermal noise current with the phasor:

$$\left\langle |i_\mathrm{T}|^2 \right\rangle = 8kT_0 B G_\mathrm{S}, \tag{5.82}$$

where $T_0 = 290\,\mathrm{K}$ is the standard temperature and B the measurement bandwidth.

Note that in Figure 5.14(b), all sources share the same source admittance Y_S, so that it is sufficient to calculate v_1 for the case $i_1 = 0$:

$$v_1 = v + \frac{i + i_\mathrm{T}}{Y_\mathrm{S}}. \tag{5.83}$$

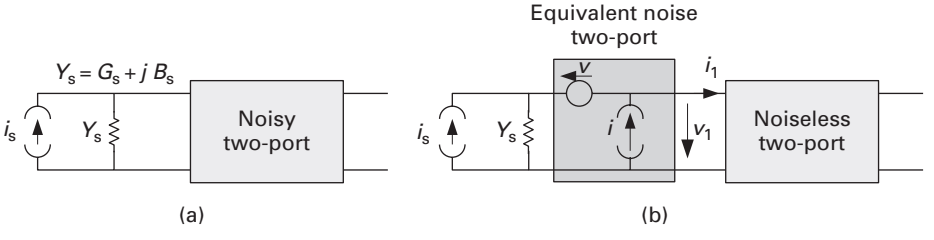

Fig. 5.14 (a) Noisy two-port with arbitrary generator admittance and (b) its equivalent circuit, introducing the equivalent noise voltage v and the equivalent noise current i.

We are really only interested in the powers, $N \propto \langle |v_1|^2 \rangle$:

$$\begin{aligned} \left\langle |v_1|^2 \right\rangle &= \langle v_1 v_1^\star \rangle \\ &= \left\langle |v|^2 \right\rangle + \frac{\langle |i|^2 \rangle + \langle |i_T|^2 \rangle}{|\underline{Y}_S|^2} + \frac{\langle i v^\star \rangle}{\underline{Y}_S} + \frac{\langle i^\star v \rangle}{\underline{Y}_S^\star} \\ &= \frac{\langle |i|^2 \rangle + \langle |i_T|^2 \rangle}{|\underline{Y}_S|^2} + \left\langle |v|^2 \right\rangle \left[1 + \frac{\underline{Y}_C}{\underline{Y}_S} + \left(\frac{\underline{Y}_C}{\underline{Y}_S} \right)^\star \right]. \end{aligned} \tag{5.84}$$

The noise figure can be expressed as

$$F = \frac{N_{1,\text{total}}}{N_{1,i_T}} = \frac{|v_1|^2}{\langle |i_T|^2 \rangle / |\underline{Y}_S|^2}, \tag{5.85}$$

where $N_{1,\text{total}}$ is the noise power due to all noise sources and N_{1,i_T} the noise power due to the generator thermal noise at $T = T_0$ alone.

Combining Equations (5.84) and (5.85) yields

$$F = 1 + \frac{\langle |i|^2 \rangle}{\langle |i_T|^2 \rangle} + \frac{\langle |v|^2 \rangle}{\langle |i_T|^2 \rangle} |\underline{Y}_S|^2 \left[1 + \frac{\underline{Y}_C}{\underline{Y}_S} + \left(\frac{\underline{Y}_C}{\underline{Y}_S} \right)^\star \right]. \tag{5.86}$$

Equation (5.86) can be written in a simpler form by introducing

- an equivalent noise admittance

$$g_n = \frac{\langle |i|^2 \rangle}{8kT_0B},$$

- an equivalent noise resistance

$$R_n = \frac{\langle |v|^2 \rangle}{8kT_0B},$$

- and remembering that

$$G_S = \frac{\langle |i_T|^2 \rangle}{8kT_0B},$$

from Equation (5.82).

This results in

$$F = 1 + \frac{g_n}{G_S} + \frac{R_n}{G_S} |\underline{Y}_S|^2 \left[1 + \frac{\underline{Y}_C}{\underline{Y}_S} + \left(\frac{\underline{Y}_C}{\underline{Y}_S} \right)^\star \right]. \tag{5.87}$$

The noise figure depends therefore on the generator admittance $Y_S = G_S + \jmath Y_S$. This was first pointed out by Rothe and Dahlke in 1955 [31, 32].

We can now search for the generator admittance where F will be minimal. This results in

$$\begin{aligned} G_{S,\text{opt}} &= \sqrt{\frac{g_n}{R_n} - B_C^2} \\ B_{S,\text{opt}} &= -B_C, \end{aligned} \tag{5.88}$$

where B_C is the imaginary part of the correlation admittance.

Introducing these terms into Equation (5.87), the noise figure can be written as follows:

$$F = 1 + 2R_{\mathrm{n}}(G_{\mathrm{S,opt}} + G_{\mathrm{C}}) + \frac{R_{\mathrm{n}}}{G_{\mathrm{S}}}\left[(G_{\mathrm{S}} - G_{\mathrm{S,opt}})^2 + (B_{\mathrm{S}} + B_{\mathrm{S,opt}})^2\right]. \tag{5.89}$$

For $Y_{\mathrm{S}} = Y_{\mathrm{S,opt}}$, the minimum noise figure is

$$F_{\min} = 1 + 2R_{\mathrm{n}}(G_{\mathrm{S,opt}} + G_{\mathrm{C}}) = 1 + 2R_{\mathrm{n}}\left(G_{\mathrm{C}} + \sqrt{\frac{g_{\mathrm{n}}}{R_{\mathrm{n}}} - B_{\mathrm{C}}^2}\right), \tag{5.90}$$

and therefore finally

$$\begin{aligned} F &= F_{\min} + \frac{R_{\mathrm{N}}}{G_{\mathrm{S}}}\left[(G_{\mathrm{S}} - G_{\mathrm{S,opt}})^2 + (B_{\mathrm{S}} - B_{\mathrm{S,opt}})^2\right] \\ &= F_{\min} + \frac{R_{\mathrm{N}}}{G_{\mathrm{S}}}\left|\underline{Y}_{\mathrm{S}} - \underline{Y}_{\mathrm{S,opt}}\right|^2. \end{aligned} \tag{5.91}$$

To describe the noise performance of a two-port, we need the following parameters:

- the minimum noise figure $F_{\min}$,
- the equivalent noise resistance R_{n},
- the noise-optimised generator admittance $\underline{Y}_{\mathrm{S,opt}}$.

For microwave applications, it is more customary to work with reflection coefficients instead of impedances or admittances. For this, we introduce a normalising admittance Y_0, typically 20 mS, and a normalised noise resistance $r_{\mathrm{n}} = R_{\mathrm{n}}Y_0$. Then Equation (5.91) turns into

$$F = F_{\min} + \frac{4r_{\mathrm{n}}\left|\Gamma_{\mathrm{S}} - \Gamma_{\mathrm{S,opt}}\right|^2}{\left(1 - |\Gamma_{\mathrm{S}}|^2\right)\left|1 + \Gamma_{\mathrm{S,opt}}\right|^2}, \tag{5.92}$$

where

$$\Gamma_{\mathrm{S}} = \frac{Y_0 - Y_{\mathrm{S}}}{Y_0 + Y_{\mathrm{S}}}.$$

Plotting $F = f(\Gamma_{\mathrm{S}})$ on a Smith chart, the contours of constant F are circles. An example is shown in Figure 5.15.

Associated gain

The condition for noise-optimised source reflection coefficient $\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{S,opt}}$ is independent of the power matching conditions derived earlier, $\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{in}}^{\star}$. The maximum gain under 'noise matching' conditions is therefore smaller than the MAG, which would be achieved when input and output of the two-port are conjugately matched. Here, $\Gamma_{\mathrm{S}} = \Gamma_{\mathrm{S,opt}}$ and the gain for the conjugately matched output becomes (see Equation (5.53))

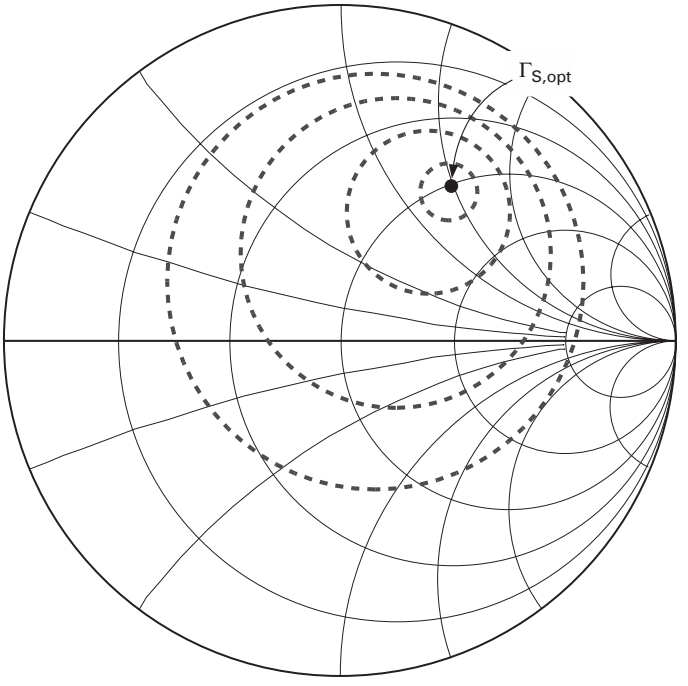

Fig. 5.15 Example for 'noise circles' – contours of constant noise figures on the Smith chart.

$$G_{ass} = \frac{1 - |\,\Gamma_{S,opt}\,|^2}{|\,1 - S_{11}\Gamma_{S,opt}\,|^2}\,|\,S_{21}\,|^2\,\frac{1}{1 - |\,\Gamma_{out}\,|^2} \tag{5.93}$$

$$= \frac{|S_{21}|^2\left(1 - \left|\Gamma_{S,opt}\right|^2\right)}{\left|1 - S_{11}\Gamma_{S,opt}\right|^2 - \left|S_{22}(1 - S_{11}\Gamma_{S,opt}) + S_{12}S_{21}\Gamma_{S,opt}\right|^2},$$

since

$$\Gamma_{out} = S_{22} + \frac{S_{21}S_{12}\Gamma_{S,opt}}{1 - S_{11}\Gamma_{S,opt}}.$$

The available gain under noise matching conditions is called *associated gain*.

5.4 Transistor amplifiers

5.4.1 A brief historical discourse

Amplifiers are such an integral part of any wireless communication system that we have to explicitly recall that in the first decades of radio, they were not used at all. It was not before the invention of the electron tube triode that amplification of alternating current signals became possible. Lee De Forest's Audion tube (conceived in 1906 and patented in 1908 [10]) is probably the first example of an amplifying device. It was 24-year-old Edwin Armstrong, the prolific and finally tragic inventor of radio's early days, who

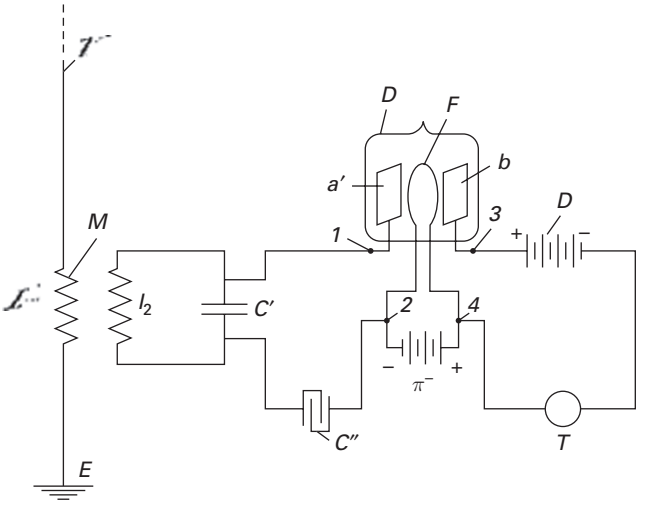

Fig. 5.16 De Forest's three-terminal audion (from US Patent No. 879,582).

actually explained its operation in 1914. As Figure 5.16 shows, the De Forest Audion already has the common three-terminal arrangement which we also find in transistor amplifiers. De Forest's claim to have invented the tube concept independently of the 'thermionic valve' patented by John Ambrose Fleming in 1905 [9] is doubtful, however.

Critical amplifier parameters vary with respect to where in a system it is being used, and what kinds of signals are being fed through the amplifiers. For example:

- When dealing with very low-level signals, for example immediately after the receiving antenna, or behind the optoelectronic converter in a fibre communication system, it is of paramount importance that the amplifier itself adds as little electronic noise as possible.
- At the output of a wireless transmitter, an amplifier should produce the required RF power with optimum efficiency and/or the required linearity, which in turn depends on the modulation format used.
- Other systems, such as in fibre-optic communications or ultra-wideband wireless systems, require extreme bandwidths, with little variation of gain and group velocity between the lower and upper cutoff frequencies.

These requirements can rarely be met simultaneously, so trade-offs have to be made which require a thorough study of the system specifications before the amplifier design is begun.

5.4.2 Fundamental amplifier configurations

Amplifiers will be discussed initially at a certain level of abstraction in order to make clear that the fundamental methods apply to FETs and bipolar transistors alike. The use of general methods will be emphasised rather than introducing a large number of circuit topologies. To this end, we will first introduce a generalised equivalent circuit for an

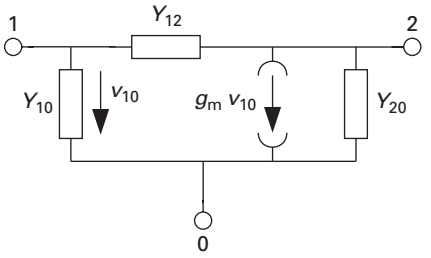

Fig. 5.17 Generic three-terminal amplifying device.

amplifying device, and then calculate its small-signal parameters – the y matrix in this case. Based on the y matrix, we develop general expressions for important amplifier parameters – the input and output admittances, and voltage and current gain. For different circuit topologies, we then calculate modified y matrices, which will immediately yield the amplifier parameters.

The move from abstract concept to actual circuit implementation will also be made occasionally. You will see that the elements of real devices can be readily matched to the general equivalent circuit.

The first important observation in amplifier design is that three-terminal devices can be used in three fundamentally different ways. Let us consider the generic three-terminal device depicted in Figure 5.17.

We restrict our discussion to linear behaviour for the time being – corresponding to the small-signal case. The amplifying action is due to the voltage-controlled current source between nodes 2 and 0. The controlling entity is the input voltage v_{10} and the parameter is the transconductance $\underline{g}_{\mathrm{m}}$, which is generally taken as a complex value. This allows us to include additional phase delays in $\underline{g}_{\mathrm{m}}$:

$$\underline{g}_{\mathrm{m}} = g_{\mathrm{m}0} e^{-\mathrm{j}\omega\tau} \tag{5.94}$$

Additionally, we included complex impedances $\underline{Y}_{10}$, $\underline{Y}_{20}$ and $\underline{Y}_{12}$. These can later be correlated with the parameters of FET or bipolar transistor small-signal equivalent circuits introduced in Chapter 2.

y matrix representation

For the 'hybrid π' equivalent circuit in Figure 5.17, the y matrix is easily calculated and will be used here. The advantage is that general amplifier properties, such as voltage and power gain or input and output admittance with arbitrary terminations, can be calculated once and can then be easily applied to the different topologies we will consider.

The y matrix expresses the following system of linear equations with respect to the two-port shown in Figure 5.18:

$$i_1 = y_{11}\, v_1 + y_{12}\, v_2 \tag{5.95}$$

$$i_2 = y_{21}\, v_1 + y_{22}\, v_2 \tag{5.96}$$

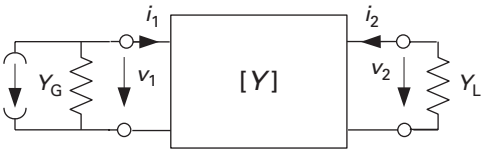

Fig. 5.18 A two-port in y matrix representation terminated with generator impedance $\underline{Y}_G$ and load impedance $\underline{Y}_L$.

The y matrix $[Y]$ can also be calculated from the scattering matrix [S] introduced earlier:

$$[Y] = Y_0 \cdot ([1] - [S])\,([1] + [S])^{-1} \tag{5.97}$$

where [1] is the identity matrix and Y_0 the normalising admittance used in the calculation of the scattering matrix – usually 20 mS.

First, consider the input admittance Y_1. The output port is terminated by an admittance $\underline{Y}_L$.

$$Y_1 = \frac{i_1}{v_1}$$
$$i_2 = -v_2\,\underline{Y}_L.$$

We find

$$Y_1 = y_{11} - \frac{y_{12}\,y_{21}}{Y_L + y_{22}}. \tag{5.98}$$

Likewise, the output admittance when the input is terminated with an arbitrary admittance $\underline{Y}_G$ is

$$Y_2 = y_{22} - \frac{y_{12}\,y_{21}}{Y_G + y_{11}}. \tag{5.99}$$

The forward voltage gain $A_V = v_2/v_1$ is

$$A_V = \frac{-y_{21}}{Y_L + y_{22}}. \tag{5.100}$$

And finally the current gain $A_I = i_2/i_1$:

$$A_I = \frac{y_{21}\,\underline{Y}_L}{y_{11}(y_{22} + \underline{Y}_L) - y_{21}\,y_{12}}. \tag{5.101}$$

The frequently used short-circuit current gain can be easily calculated from Equation (5.101) when $\underline{Y}_L \to \infty$:

$$h_{21} = \frac{i_2}{i_1}\,|(v_2 = 0) = \frac{y_{21}}{y_{11}}. \tag{5.102}$$

Common source/common emitter: node 0 as the common node

The most obvious connection is to ground node 0. In FETs, this configuration is called *common source*; in bipolar transistors, it is called *common-emitter* configuration.

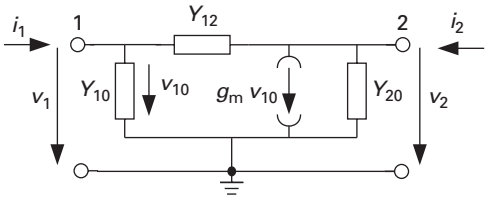

Fig. 5.19 Generic amplifier configuration with node 0 grounded.

The y matrix can be easily calculated (see Figure 5.19):

$$y_{11} = \underline{Y}_{10} + \underline{Y}_{12} \tag{5.103}$$

$$y_{12} = -\underline{Y}_{12} \tag{5.104}$$

$$y_{21} = \underline{g}_{\mathrm{m}} - \underline{Y}_{12} \tag{5.105}$$

$$y_{22} = \underline{Y}_{20} + \underline{Y}_{12}. \tag{5.106}$$

Let us first investigate the voltage gain using Equation (5.100):

$$A_{\mathrm{V}} = -\frac{y_{21}}{\underline{Y}_{\mathrm{L}} + y_{22}} = -\frac{\underline{g}_{\mathrm{m}} - \underline{Y}_{12}}{\underline{Y}_{\mathrm{L}} + \underline{Y}_{20} + \underline{Y}_{12}}. \tag{5.107}$$

To facilitate interpretation, let us assume that $\underline{g}_{\mathrm{m}}$ is real, and that the feedback admittance is weak: $g_{\mathrm{m}} - \underline{Y}_{12} \approx g_{\mathrm{m}}$, $\underline{Y}_{\mathrm{L}} + \underline{Y}_{20} + \underline{Y}_{12} \approx \underline{Y}_{\mathrm{L}} + \underline{Y}_{20}$. Then,

$$A_{\mathrm{V}} \approx -\frac{g_{\mathrm{m}}}{\underline{Y}_{\mathrm{L}} + \underline{Y}_{20}}.$$

For low frequencies, $\underline{Y}_{\mathrm{L}}$ and $\underline{Y}_{20}$ are real, and we find:

- The magnitude of voltage gain in common-source and common-emitter stages will be approximately equal to the product of transconductance and effective load resistance (the parallel connection of external load resistance and device output resistance).
- The output voltage lags the input voltage by 180°.

A useful figure of merit is the maximum voltage gain a three-terminal device can produce with node 0 grounded. We find it from Equation (5.107) by choosing $Y_{\mathrm{L}} = 0$ and assuming $\underline{Y}_{12} \ll \underline{Y}_{20}, \underline{g}_{\mathrm{m}}$:

$$A_{\mathrm{V,max}} \approx -\frac{\underline{g}_{\mathrm{m}}}{\underline{Y}_{20}}. \tag{5.108}$$

Now let us take a look at the input admittance, using Equation (5.98):

$$Y_1 = y_{11} - \frac{y_{12}\, y_{21}}{Y_{\mathrm{L}} + y_{22}} = \underline{Y}_{10} + \underline{Y}_{12}\left(1 + \frac{g_{\mathrm{m}} - \underline{Y}_{12}}{\underline{Y}_{\mathrm{L}} + \underline{Y}_{20} + \underline{Y}_{12}}\right),$$

or recognising from Equation (5.107) that the last term in parentheses is A_{V}:

$$Y_1 = \underline{Y}_{10} + \underline{Y}_{12}(1 - A_{\mathrm{V}}). \tag{5.109}$$

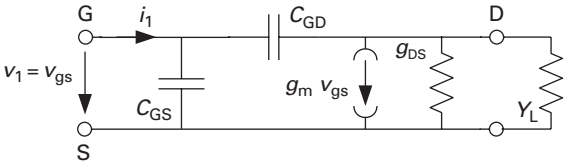

Fig. 5.20 Simplified hybrid π equivalent circuit of a FET with load admittance.

Miller effect

It is now time for a small practical example. In a FET, the small-signal equivalent circuit looks like Figure 5.20, if we neglect the series resistances. From comparison with Figure 5.17, we recognise that

$$\underline{Y}_{10} = \jmath\omega C_{GS}$$
$$\underline{Y}_{12} = \jmath\omega C_{GD}$$
$$\underline{Y}_{20} = g_{DS}.$$

The voltage gain is now

$$A_V = -\frac{g_m - \jmath\omega C_{GD}}{Y_L + g_{DS} + \jmath\omega C_{GD}} \approx -\frac{g_m}{Y_L + g_{DS}},$$

for low frequencies.

The input admittance is, using Equation (5.109),

$$Y_1 = \jmath\omega C_{GS} + \jmath\omega C_{GD}\,(1 - A_V)\,. \tag{5.110}$$

The feedback capacitance C_{GD} appears in parallel with C_{GS}, but is multiplied by the magnitude of the voltage gain, augmented by one – this is the dreaded *Miller Effect*, which we always have to be aware of in high-speed circuit design, because it may significantly increase the input capacitance. It was described as early as 1920 [29].

Going back to Figure 5.19, we calculate the output admittance to be

$$Y_2 = \underline{Y}_{20} + \underline{Y}_{12}\left(1 + \frac{\underline{g}_m - \underline{Y}_{12}}{\underline{Y}_G + \underline{Y}_{10} + \underline{Y}_{12}}\right). \tag{5.111}$$

The term

$$\frac{g_m - Y_{12}}{Y_G + Y_{10} + Y_{12}} = A_r$$

can be significantly larger than 1 and may act like a 'reverse Miller Effect'. This has to be taken into account if a tuned circuit is connected to the output node, as is the case in typical tuned amplifiers. Figure 5.21 shows an example – the detuning effect of the feedback capacitance will be much larger than expected.

The current gain of the common-source/common-emitter amplifier, finally, becomes

$$A_I = \frac{(\underline{g}_m - \underline{Y}_{12})\,\underline{Y}_L}{\underline{Y}_{12}(\underline{Y}_L + \underline{g}_m + \underline{Y}_{10} + \underline{Y}_{20}) + \underline{Y}_{10}(\underline{Y}_L + \underline{Y}_{20})}. \tag{5.112}$$

We can safely assume here that the real parts of all admittances are larger than zero – then the current gain in the quasi-static limit ($f \to 0$) is positive.

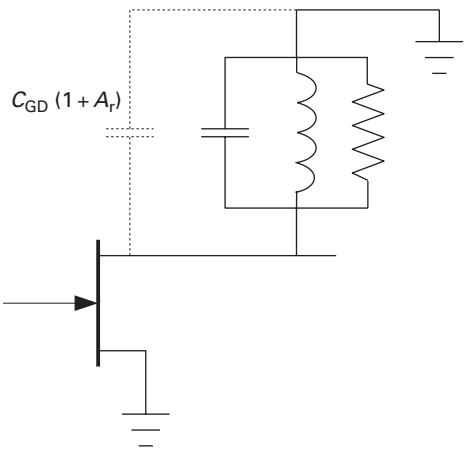

Fig. 5.21 Common-source amplifier stage with tuned load, and the output-referred Miller capacitance.

Let us check the latter equation again with our FET equivalent circuit by calculating the short-circuit current gain:

$$h_{21} = A_I|(\underline{Y}_L \to \infty) = \frac{g_m - \jmath\omega C_{GD}}{\jmath\omega(C_{GS} + C_{GD})} \approx \frac{g_m}{\jmath\omega C_{GS}}$$

assuming that $g_m \gg \omega C_{GD}$ and $C_{GS} \gg C_{GD}$.

This is the equation we had earlier used in Chapter 2 to estimate the transit frequency of a FET to be $f_T = g_m/(2\pi C_{GS})$.

Let us summarise our findings for the amplifier configuration with node 0 grounded – applicable to both the FET common-source and the bipolar common-emitter topologies:

- The configuration can provide substantial voltage and current gains.
- The voltage gain provides a phase shift of 180° in the quasi-static limit, whereas the current gain experiences no phase shift.
- Due to the presence of feedback, the input and output admittances always depend on the termination of the opposite port.
- For substantial voltage gains, the *Miller effect* has to be observed which can substantially increase the input capacitance.

Common gate/common base: node 1 as the common node

In the next set-up to be discussed, node 1 is grounded, and the input current is fed into node 0. Node 2 is still the output node. This applies to the FET common-gate and the bipolar common-base configurations. Again, we will first calculate the y matrix in a general form. For this, it is valuable to recognise in Figure 5.22 that $v_1 = -v_{10}$.

$$y_{11} = \underline{Y}_{20} + \underline{Y}_{10} + \underline{g}_m \tag{5.113}$$

$$y_{12} = -\underline{Y}_{20} \tag{5.114}$$

$$y_{21} = -\underline{Y}_{20} - \underline{g}_m \tag{5.115}$$

$$y_{22} = \underline{Y}_{20} + \underline{Y}_{12}. \tag{5.116}$$

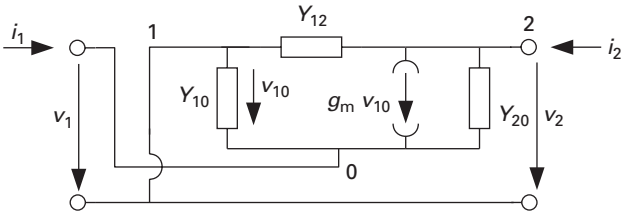

Fig. 5.22 Hybrid π equivalent circuit with node 1 grounded.

This input admittance is then

$$Y_1 = \underline{Y}_{10} + (\underline{g}_m + \underline{Y}_{20})\frac{\underline{Y}_L + \underline{Y}_{12}}{\underline{Y}_L + \underline{Y}_{20} + \underline{Y}_{12}}. \tag{5.117}$$

Let us simplify this expression somewhat. First, realise that in practical devices necessarily $\underline{g}_m \gg \underline{Y}_{20}$, otherwise it would not have a reasonable voltage gain in common-source or common-emitter configuration (see above). Further, let us assume that $\underline{Y}_{12} \ll \underline{Y}_L$. Then,

$$Y_1 \approx \underline{Y}_{10} + \underline{g}_m \frac{\underline{Y}_L}{\underline{Y}_L + \underline{Y}_{20}}.$$

An interesting observation is that now $\underline{Y}_{20}$ is the feedback admittance which determines the sensitivity of the input admittance on the load. If further, this admittance is very small, $\underline{Y}_{20} \ll \underline{Y}_L$, then

$$Y_1 \approx \underline{Y}_{10} + \underline{g}_m.$$

Using the example of Figure 5.20, we find

$$Y_1 \approx g_m\left(1 + \frac{j\omega C_{GS}}{g_m}\right) = g_m\left(1 + j\frac{\omega}{\omega_T}\right).$$

The input admittance is hence approximately the transconductance, unless we are operating close to f_T. In practical transistors, this will be a quite large value – much larger than the input admittance in the topology with node 0 grounded.

The output admittance is

$$Y_2 = \underline{Y}_{12} + \underline{Y}_{20}\frac{\underline{Y}_G + \underline{Y}_{10}}{\underline{Y}_G + \underline{Y}_{10} + \underline{Y}_{20} + \underline{g}_m}. \tag{5.118}$$

Assuming again that $\underline{g}_m \gg \underline{Y}_{20}$,

$$Y_2 \approx \underline{Y}_{12} + \frac{\underline{Y}_{20}}{1 + \dfrac{\underline{g}_m}{\underline{Y}_G + \underline{Y}_{10}}}$$

Compared to the topology with node 0 grounded (common emitter or common source), the output admittance will be substantially smaller.

The voltage gain is calculated as

$$A_V = \frac{\underline{g}_m + \underline{Y}_{20}}{\underline{Y}_L + \underline{Y}_{20} + \underline{Y}_{12}}. \tag{5.119}$$

Assuming again that $\underline{g}_m \gg \underline{Y}_{20}$, this simplifies to

$$A_V = \frac{\underline{g}_m}{\underline{Y}_L + \underline{Y}_{20} + \underline{Y}_{12}}.$$

Compare with Equation (5.107) – this is the same result, if $\underline{Y}_{12} \ll g_m$, which is a safe assumption.

Finally, the current gain is

$$A_I = -\frac{\underline{Y}_L \underline{g}_m}{(\underline{Y}_L + \underline{Y}_{12})(\underline{Y}_{20} + \underline{Y}_{10} + \underline{g}_m) + \underline{Y}_{20}\underline{Y}_{10}}. \tag{5.120}$$

The magnitude of A_I for this topology is hence less than 1. For greater simplicity, calculate the short-circuit current gain ($Y_L \to \infty$):

$$h_{21} = -\frac{\underline{Y}_{20} + \underline{g}_m}{\underline{Y}_{20} + \underline{g}_m + \underline{Y}_{10}}$$
$$\approx -\frac{1}{1 + \dfrac{\underline{Y}_{10}}{\underline{g}_m}}.$$

Using again the simple FET equivalent circuit in Figure 5.20, this reduces to

$$h_{21} = -\frac{1}{1 + j\dfrac{\omega}{\omega_T}}.$$

In other words, the short-circuit current gain of the common-gate and common-base configurations will remain independent of frequency until quite close to f_T.

Common drain/common collector: node 2 as the common node

The last fundamental configuration of the generic amplifying three-terminal device (Figure 5.17) has node 2 as the common node (see Figure 5.23). In FETs, this will be called *common drain*; in bipolar transistors, this will be called *common-collector* configuration.

Again, first calculate the y matrix of this configuration.

$$y_{11} = \underline{Y}_{10} + \underline{Y}_{12} \tag{5.121}$$
$$y_{12} = -\underline{Y}_{10} \tag{5.122}$$
$$y_{21} = -(\underline{Y}_{10} + \underline{g}_m) \tag{5.123}$$
$$y_{22} = \underline{Y}_{10} + \underline{Y}_{20} + \underline{g}_m. \tag{5.124}$$

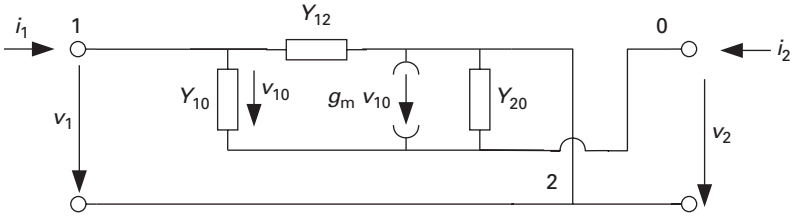

Fig. 5.23 Generic amplifier configuration with node 2 grounded.

Let us first consider the voltage gain A_V, when port 2 is terminated with a load admittance Y_L.

$$A_V = -\frac{y_{21}}{Y_L + y_{22}} = \frac{\underline{Y}_{10} + \underline{g}_m}{\underline{Y}_L + \underline{Y}_{10} + \underline{Y}_{20} + \underline{g}_m} = \left(1 + \frac{\underline{Y}_L + \underline{Y}_{20}}{\underline{Y}_{10} + \underline{g}_m}\right)^{-1}. \quad (5.125)$$

In practical devices, the second term in parentheses will be small compared to 1, at least at lower frequencies, so that $A_V \approx 1$ (but always less than 1) – v_2 follows v_1, which is why this topology is also called a *source follower* or *emitter follower* for FETs and bipolar transistors, respectively.

The input admittance is calculated to be

$$Y_1 = y_{11} - \frac{y_{21}y_{12}}{Y_L + y_{22}} = \underline{Y}_{12} + \frac{(\underline{Y}_L + \underline{Y}_{20})\underline{Y}_{10}}{\underline{Y}_L + \underline{Y}_{20} + \underline{Y}_{10} + \underline{g}_m}$$

$$= \underline{Y}_{12} + \frac{\underline{Y}_{10}}{1 + \dfrac{\underline{Y}_{10} + \underline{g}_m}{\underline{Y}_L + \underline{Y}_{20}}} \quad (5.126)$$

$$\approx \underline{Y}_{12} + \frac{\underline{Y}_{10}}{1 + \dfrac{\underline{g}_m}{\underline{Y}_L}}, \quad (5.127)$$

because typically, at least at frequencies sufficiently below f_T, $\underline{g}_m \gg \underline{Y}_{10}$, and assuming that $\underline{Y}_L \gg \underline{Y}_{20}$.

Compare this to the input admittance of the topology where node 0 was grounded, Equation (5.109), and you will notice that the influence of $\underline{Y}_{10}$ is substantially reduced, while $\underline{Y}_{12}$ does not suffer from the augmentation due to the Miller effect. We can therefore state that the topology with node 2 grounded presents a much lower input admittance.

Going to our usual FET example where the amplifying device is represented by the equivalent circuit in Figure 5.20, we can show that Equation (5.126) may have an unexpected result. In this case,

$$Y_1 = j\omega C_{GD} + \frac{\jmath\omega C_{GS}}{1 + \dfrac{g_m + \jmath\omega C_{GS}}{Y_L + g_{DS}}}.$$

Now assume that $g_m \gg \omega C_{GS}$, i.e. $\omega \ll \omega_T$, that Y_L is capacitive ($Y_L = \jmath\omega C_L$), and that $\omega C_L \gg g_{DS}$.

$$Y_1 = \jmath\omega C_{GD} - \frac{\omega^2 C_{GS} C_L}{g_m + \jmath\omega(C_L + C_{GS})},$$

or separating into real and imaginary parts,

$$Y_1 = -\frac{\omega^2 g_m C_{GS} C_L}{g_m^2 + \omega^2 (C_L + C_{GS})^2} + \jmath\omega \left(\frac{\omega^2 C_{GS} C_L (C_L + C_{GS})}{g_m^2 + \omega^2 (C_L + C_{GS})^2} + C_{GD} \right). \quad (5.128)$$

We created an admittance with a negative real part! This can be useful, for example if we want to build an oscillator (see Section 5.5), but also very dangerous for amplifier stability.

Moving back to the more general case, let us calculate the current gain:

$$A_I = -\frac{\underline{Y}_L(\underline{Y}_{10} + \underline{g}_m)}{\underline{Y}_{12}(\underline{Y}_{10} + \underline{Y}_{20} + \underline{Y}_L + \underline{g}_m) + \underline{Y}_{10}(\underline{Y}_L + \underline{Y}_{20})}. \tag{5.129}$$

Comparing this with the equation for the current gain with node 0 grounded (Equation (5.112)), we see that the denominators are equal. If further $\underline{g}_m \gg \underline{Y}_{10}, \underline{Y}_{12}$, which is typically the case, the two current gains have equal magnitude.

Finally, the output admittance of the topology with node 2 grounded will be calculated. The input port is terminated with a generator admittance Y_G.

$$Y_2 = \underline{Y}_{20} + \frac{(\underline{Y}_{10} + \underline{g}_m)(\underline{Y}_G + \underline{Y}_{12})}{\underline{Y}_G + \underline{Y}_{10} + \underline{Y}_{12}}. \tag{5.130}$$

To interpret this equation, assume that $Y_G \gg \underline{Y}_{10} + \underline{Y}_{12}, \underline{g}_m \gg \underline{Y}_{10}$. Then,

$$Y_2 \approx \underline{Y}_{20} + \underline{g}_m \approx \underline{g}_m,$$

in most cases. Compared to the output admittance of the original topology which had node 0 as the common node (Equation (5.111)), we see that now we have a substantially higher output conductance.

The combination of a very low input conductance (very high input impedance) and high output conductance (low output resistance) is the most important aspect of common-drain/common-collector topologies.

5.4.3 Feedback

Negative feedback is another important principle in amplifier design. In small-signal design, it is used for impedance matching purposes, to make an amplifier stable and to increase its bandwidth. The negative feedback amplifier was invented by Harold Black in 1927 [22].

The most important feedback implementations in high-speed amplifier design are *shunt–shunt* and *series–series* feedback, as shown in Figure 5.24.

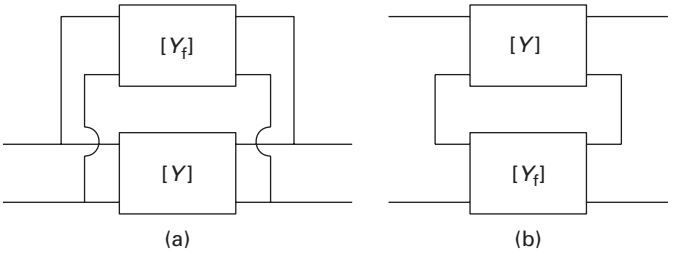

Fig. 5.24 Feedback configurations: (a) shunt–shunt feedback and (b) series–series feedback.

Shunt–shunt feedback

Case (a) is easily calculated using a y matrix representation, because the resulting y matrix is the sum of the individual matrices.

$$[Y_{\mathrm{T}}] = [Y] + [Y_{\mathrm{f}}] = \begin{bmatrix} y_{11} + y_{11,f} & y_{12} + y_{12,f} \\ y_{21} + y_{21,f} & y_{22} + y_{22,f}. \end{bmatrix} \tag{5.131}$$

Series–series feedback

Series–series feedback is better treated using a z matrix representation:

$$v_1 = z_{11} i_1 + z_{12} i_2 \tag{5.132}$$

$$v_2 = z_{21} i_1 + z_{22} i_2. \tag{5.133}$$

Conversion from y to z matrix is easy because the z matrix is simply the inverse of the y matrix:

$$[Z] = \frac{1}{\Delta(Y)} \begin{bmatrix} y_{22} & -y_{12} \\ -y_{21} & y_{11} \end{bmatrix}, \tag{5.134}$$

where $\Delta(Y)$ is the determinant of the y matrix and $\Delta(Y) = y_{11}y_{22} - y_{12}y_{21}$.

Once the z matrices have been obtained, the resulting z matrix of the circuit with series–series feedback is the sum of the individual z matrices:

$$[Z_T] = [Z] + [Z_{\mathrm{f}}] = \begin{bmatrix} z_{11} + z_{11,\mathrm{f}} & z_{12} + z_{12,\mathrm{f}} \\ z_{21} + z_{21,\mathrm{f}} & z_{22} + z_{22,\mathrm{f}} \end{bmatrix}. \tag{5.135}$$

The conversion back from z to y matrix is equally simple:

$$[Y] = \frac{1}{\Delta(Z)} \begin{bmatrix} z_{22} & -z_{12} \\ -z_{21} & z_{11} \end{bmatrix}. \tag{5.136}$$

Use of feedback in small-signal amplifiers

A very common feedback example is shown in Figure 5.25, where an admittance Y_{f} is connected between the output and the input of a common-source amplifier. The FET

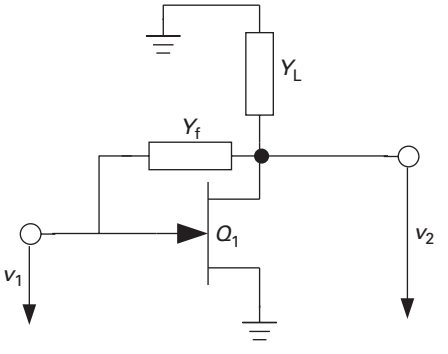

Fig. 5.25 Example of shunt–shunt feedback in a common-source amplifier.

shall be treated using the generic equivalent circuit in Figure 5.17. The feedback two-port contains only one element, Y_f. If $[Y_{Q1}]$ is the y matrix of transistor Q_1, then the y matrix of the transistor with feedback is

$$[Y_T] = \begin{bmatrix} y_{11,Q1} + Y_f & y_{12,Q1} - Y_f \\ y_{21,Q1} - Y_f & y_{22,Q1} + Y_f \end{bmatrix}. \tag{5.137}$$

Unilateralisation

An immediate application of this feedback technique is the elimination of the parameter y_{12}. Choosing

$$Y_f = -y_{12,Q1}$$

results in a *unilateralised* amplifier two-port where the input parameters no longer depend on the output load, and vice versa. This can be used to improve amplifier stability, and is referred to as *neutralisation*.

There are several ways of achieving this. In narrow-band amplifiers, the usually purely capacitive feedback may be tuned out using an inductor. The inductor is chosen to form a parallel resonance with the feedback capacitor at the frequency of operation.

A more elegant technique was invented by Harold A. Wheeler in the early 1920s for electron tubes. A current with equal magnitude – but opposite phase, as the current through the feedback admittance – is fed back from the output to the input node, where the two currents cancel out exactly. This is shown in Figure 5.26. In integrated circuits, the realisation of the autotransformer is hampered by the typically high losses of on-chip inductors. However, any kind of phase reversal will do; a particularly simple technique will be discussed further down in the context of the differential amplifier (p. 336).

Port matching

A very common task in amplifier design is matching, e.g. the input admittance Y_1 to the generator admittance: $Y_1 = Y_G^*$, where Y_G^* is the complex conjugate of the generator

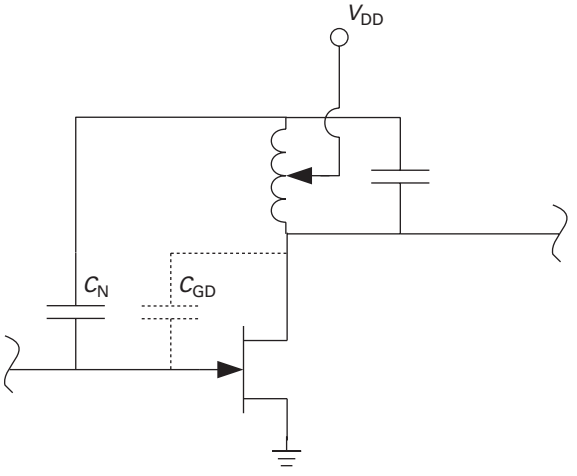

Fig. 5.26 Amplifier neutralisation using an autotransformer.

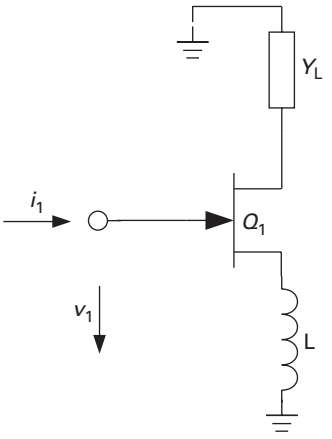

Fig. 5.27 Inductive emitter degeneration as an example for series–series feedback.

admittance. While this is commonly done by cascading a matching network and the amplifier two-port, the same result can frequently be achieved by using feedback.

With shunt–shunt feedback, the input admittance becomes

$$Y_1 = y_{11,Q1} + Y_f - \frac{(y_{12,Q1} - Y_f)(y_{21,Q1} - Y_f)}{Y_L + y_{22,Q1} + Y_f}. \tag{5.138}$$

Setting $Y_1 = Y_G^*$ and solving for Y_f, we obtain

$$Y_f = \frac{(y_{11,Q1} - Y_G^*)(Y_L + y_{22,Q1}) - y_{21,Q1}y_{12,Q1}}{Y_G^* - Y_L - (y_{11,Q1} + y_{12,Q1} + y_{21,Q1} + y_{22,Q1})}. \tag{5.139}$$

Inductive source degeneration

Series–series feedback is also commonly used in matching problems. A practical example is shown in Figure 5.27. An inductor is inserted into the source lead of a FET. This is referred to as *inductive source degeneration*, and may equally be applied to bipolar transistors. We shall now investigate its effect on the input impedance. For simplicity's sake, we assume that for transistor Q_1, $\underline{Y}_{12}$ and $\underline{Y}_{20}$ can be neglected. For a general impedance Z_f in the source lead, the input impedance becomes

$$Z_1 = \frac{v_1}{i_1} = \frac{1}{\underline{Y}_{10}} + Z_f\left(1 + \frac{\underline{g}_m}{\underline{Y}_{10}}\right). \tag{5.140}$$

In the specific case, $Z_f = \jmath\omega L$. Using the simple FET equivalent circuit in Figure 5.20, further $\underline{Y}_{10} = \jmath\omega C_{GS}$. Recall that $\omega_T = g_m/C_{GS}$, and we find

$$Z_1 = \omega_T L + \jmath\left(\omega L - \frac{1}{\omega C_{GS}}\right). \tag{5.141}$$

The inductance hence creates a real part in the input impedance.

Bandwidth improvement

Both shunt–shunt and series–series feedback can be used to increase bandwidth, however at the expense of maximum gain.

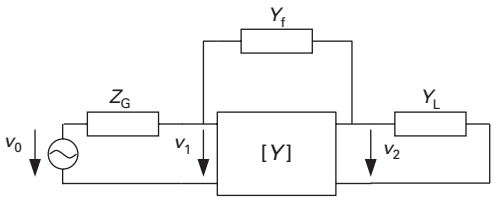

Fig. 5.28 Two-port in y matrix representation with shunt–shunt feedback.

Let us investigate the use of shunt–shunt feedback. First, realise that the influence of the generator impedance needs to be included. Figure 5.28 shows the corresponding schematic. We are interested in the voltage gain between generator and load. For $Y_f = 0$, this is

$$G_V = \frac{v_2}{v_0} = \frac{A_V}{1 + Z_G(y_{11} + A_V y_{12})} \tag{5.142}$$

where A_V is given by Equation (5.100). With $Y_f \neq 0$, we obtain

$$A'_V = -A_V \frac{1 - \dfrac{Y_f}{y_{21}}}{1 + \dfrac{Y_f}{y_{22} + Y_L}} \tag{5.143}$$

$$G'_V = \frac{A'_V}{1 + Z_G\left[y_{11} + A'_V y_{12} + Y_f(1 - A'_V)\right]}. \tag{5.144}$$

The noteworthy feature here is that Y_f appears magnified by $1 - A'_V$ in the denominator. We found this already in a different context when discussing the Miller effect.

The bandwidth enhancement effect can be seen in a simple example. The amplifying device shall be the simple FET from Figure 5.20. Then the y matrix is, with some appropriate simplifications,

$$\begin{aligned} y_{11} &= \jmath\omega(C_{GS} + C_{GD}) \approx \jmath\omega C_{GS} \\ y_{12} &= -\jmath\omega C_{GD} \\ y_{21} &= g_m - \jmath\omega C_{GD} \approx g_m \\ y_{22} &= g_{DS} + \jmath\omega C_{GD} \approx g_{DS}. \end{aligned}$$

Let us further assume that

$$Y_f \ll g_m.$$

Then,

$$A_V = -\frac{g_m}{Y_L + g_{DS}},$$

$$A'_V = A_V \frac{Y_L + g_{DS}}{Y_L + g_{DS} + Y_f},$$

and

$$G'_{\mathrm{V}} = \frac{A'_{\mathrm{V}}}{1 + Z_{\mathrm{G}}\left[(1 - A'_{\mathrm{V}})Y_{\mathrm{f}} + \jmath\omega(C_{\mathrm{GS}} - A'_{\mathrm{V}}C_{\mathrm{GD}})\right]}.$$

We calculate the 3 dB cutoff frequency as the frequency where the real and the imaginary parts in the denominator are equal. For $Y_{\mathrm{f}} = 0$, this is

$$\omega_{\mathrm{C}}(Y_{\mathrm{f}} = 0) = \frac{1}{Z_{\mathrm{G}}(C_{\mathrm{GS}} - A_{\mathrm{V}}C_{\mathrm{GD}})}.$$

Next, we apply a purely resistive feedback, $Y_{\mathrm{f}} = G_{\mathrm{f}}$:

$$\omega_{\mathrm{C}}(Y_{\mathrm{f}} = G_{\mathrm{f}}) = \frac{1 + Z_{\mathrm{G}}G_{\mathrm{f}}(1 - A'_{\mathrm{V}})}{Z_{\mathrm{G}}(C_{\mathrm{GS}} - A'_{\mathrm{V}}C_{\mathrm{GD}})}.$$

Frequently, $G_{\mathrm{f}} \ll (Y_{\mathrm{L}} + G_{\mathrm{DS}})$ and therefore $A'_{\mathrm{V}} \approx A_{\mathrm{V}}$. The resistive feedback hence results in a very substantial bandwidth enhancement by the factor $1 + Z_{\mathrm{G}}G_{\mathrm{f}}(1 - A_{\mathrm{V}})$. The low-frequency gain, however, decreases to

$$G_{\mathrm{V}}(\omega \to 0) = \frac{A_{\mathrm{V}}}{1 + Z_{\mathrm{G}}R_{\mathrm{f}}(1 - A_{\mathrm{V}})},$$

so that the product $G_{\mathrm{V}}(\omega \to 0)\omega_{\mathrm{C}} = \text{const}$.

Larger bandwidth enhancement is possible if we allow Y_{f} to have a negative imaginary part. This will be treated further down.

Bandwidth enhancement using series–series feedback will be treated for the specific example shown in Figure 5.29. This is a transadmittance stage which converts an input voltage into an output current. The series feedback using Z_{f} will first of all increase the input impedance to lower the load on the preceding stage. If it is made complex, it can be used for significant bandwidth enhancement as well.

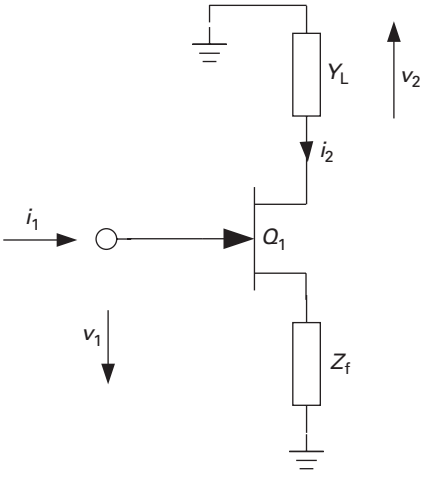

Fig. 5.29 Bandwidth enhancement using series–series feedback.

The transadmittance of this stage is

$$Y_t = \frac{i_2}{v_1} = \frac{g_m}{1 + g_m Z_f \left(1 + \frac{j\omega}{\omega_T}\right)}, \tag{5.145}$$

if the FET is described by the simple equivalent circuit in Figure 5.20, and using $\omega_T = g_m / C_{GS}$.

If now Z_f is a parallel RC network,

$$Z_f = \frac{R_f}{1 + j\omega R_f C_f},$$

and C_f is chosen,

$$C_f = \frac{1}{\omega_T R_f},$$

the frequency dependence in Y_T will disappear:

$$Y_T = \frac{g_m}{1 + g_m R_f},$$

at least for this simple equivalent circuit!

5.4.4 Amplifier configurations with two transistors

In the first part of this chapter, we have seen how the fundamental topologies we can realise with a three-terminal amplifying device have very different properties in terms of input and output admittances, as well as voltage and current gains. Further flexibility in tailoring amplifier properties is achieved when we combine two of the fundamental topologies. We will use generic FETs in order to help visualise the circuits. However, the fundamental concepts apply equally to bipolar transistors – in fact, to any three-terminal amplifying device, as was outlined in the more abstract discussion above.

Common-drain/common-source configuration

Suppose that we want to construct a buffer amplifier, which shall impose a minimal load on a generator, yet also have a significant voltage gain. This can be achieved with the combination of

- a common-drain stage (node 2 as the common node), providing the high input impedance, and
- a common-source stage (node 0 as the common node), providing the voltage gain.

Figure 5.30 shows the schematic of the common-drain/common-source (CD/CS) topology. Q_1 is the common-drain transistor, Q_2 the common-source transistor, Y_L is the load admittance and Y_G the generator admittance.

The two important aspects of this configuration are input admittance and voltage gain, as discussed.

The voltage gain of the second stage is

$$A_{V,Q2} = \frac{v_2}{v_A} \approx -\frac{\underline{g}_m}{\underline{Y}_L + \underline{Y}_{20}} = -\frac{g_{m,Q2}}{g_{DS,Q2} + Y_L},$$

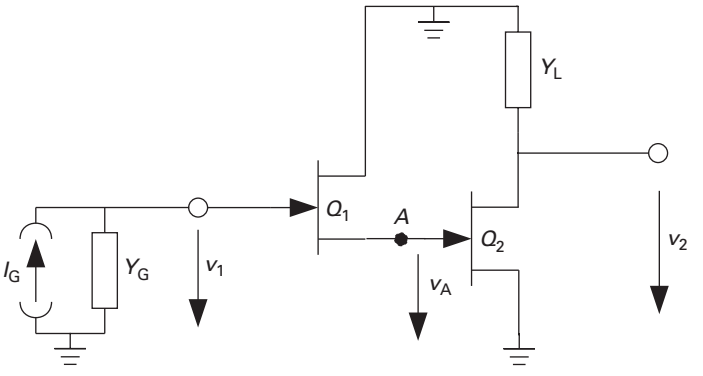

Fig. 5.30 Schematic of a CD/CS amplifier cell. Bias arrangement has been omitted for clarity's sake.

using Equation (5.107) with the simplifying assumptions indicated there, and the FET equivalent circuit in Figure 5.20.

We can now calculate the input admittance of stage two:

$$Y_{1,Q2} = \underline{Y}_{10,Q2} + \underline{Y}_{12,Q2}(1 - A_{V,Q2}) = \jmath\omega\left[C_{GS,Q2} + C_{GD,Q2}(1 - A_{V,Q2})\right].$$

Because Q_2 is in common-source configuration, we observe the Miller effect, which may significantly increase the capacitance seen from node A into Q_2. To judge the importance of this, consider the output admittance of transistor Q_1, which appears in parallel to $Y_{1,Q2}$ at node A. We can use Equation (5.130) with the appropriate simplifications:

$$Y_{2,Q1} \approx \underline{Y}_{20,Q1} + \underline{g}_{m,Q1} = g_{DS,Q1} + g_{m,Q2} \approx g_{m,Q1}.$$

The admittance from node A to ground can then be written as

$$Y_A = g_{m,Q1}\,(1 + \jmath\omega\tau_A)\,,$$

where τ_A is the characteristic time constant of node A:

$$\tau_A = \frac{C_{GS,Q2}}{g_{m,Q1}} + \frac{C_{GD,Q2}}{g_{m,Q1}}(1 - A_{V,Q2}).$$

As long as $(2\pi\tau_A)^{-1}$ is significantly outside of the intended frequency range of operation, its effect can be neglected.

The concept of the *characteristic time constant* of internal nodes is very helpful in high-speed design, especially when tracking down reasons for unexpected limitations in performance.

Here, the situation may not be so bad, because the high capacitance seen into Q_2 is compensated for by the high conductance seen into the output of Q_1 in its common-drain configuration.

The voltage gain of stage Q_1 is Equation (5.125):

$$A_{V,Q1} = \frac{v_A}{v_1} = \left(1 + \frac{Y_L + \underline{Y}_{20,Q1}}{\underline{Y}_{10,Q1} + \underline{g}_{m,Q1}}\right)^{-1} \approx \frac{g_{m,Q1}}{g_{m,Q1} + Y_{1,Q1}} = \frac{1}{1 + j\omega\tau_A}.$$

The total voltage gain is finally

$$A_V = A_{V,Q1} A_{V,Q2} \approx \frac{A_{V,Q2}}{1 + j\omega\tau_A} \approx A_{V,Q2},$$

if $\omega \ll \tau_A^{-1}$. Again we notice the importance of the characteristic impedance of node A.

The input admittance can be calculated from Equation (5.126):

$$Y_1 \approx \underline{Y}_{12,Q1} + \frac{\underline{Y}_{10,Q1}}{1 + \dfrac{\underline{g}_{m,Q1}}{Y_{1,Q2}}} = j\omega C_{GD,Q1} + \frac{j\omega C_{GS,Q1}}{1 + \dfrac{g_{m,Q1}}{Y_{1,Q2}}}.$$

Recall that in our case, $Y_{1,Q2}$ is purely capacitive. As shown earlier, a capacitive load to a common-drain stage leads to a negative real part in the input admittance (see Equation (5.128)). Whether this represents a problem for amplifier stability depends on the generator admittance value Y_G. This should be kept in mind when investigating stability problems.

Finally, we take a look at the overall current gain. Because the output current of the first stage feeds the input of the second, we expect the overall current gain to be the product of the two individual current gains. However, we have to observe that we always counted currents positive when they flow *into* the device (see Figure 5.18). Then,

$$A_I = -A_{I,Q1} \cdot A_{I,Q2}. \tag{5.146}$$

The current gain of the first (common-drain) stage is given by Equation (5.129), observing that now the load admittance is the input admittance of the second stage:

$$A_{I,Q1} = -\frac{Y_{1,Q2}(\underline{Y}_{10,Q1} + \underline{g}_{m,Q1})}{\underline{Y}_{12,Q1}(\underline{Y}_{10,Q1} + \underline{Y}_{20,Q1} + Y_{1,Q2} + \underline{g}_{m,Q1}) + \underline{Y}_{10,Q1}(Y_{1,Q2} + \underline{Y}_{20,Q1})}, \tag{5.147}$$

while the current gain of the common-source stage is given by Equation (5.112):

$$A_{I,Q2} = \frac{(\underline{g}_{m,Q2} - \underline{Y}_{12,Q2})\,\underline{Y}_L}{\underline{Y}_{12,Q2}(\underline{Y}_L + \underline{g}_{m,Q2} + \underline{Y}_{10,Q2} + \underline{Y}_{20,Q2}) + \underline{Y}_{10,Q2}(\underline{Y}_L + \underline{Y}_{20,Q2})}, \tag{5.148}$$

where Y_L is the load admittance connected to the drain of Q_2.

Since the explicit calculation of A_I presents considerable difficulty, let us make a number of simplifying assumptions. First, we make the two-ports *unilateral*, i.e. we assume $\underline{Y}_{12} = 0$. Then, we assume that the load admittances are always much larger than the elements $\underline{Y}_{20}$ for both transistors: $Y_{1,Q2} \gg \underline{Y}_{20,Q1}$, $Y_L \gg \underline{Y}_{20,Q2}$. Equation (5.146) then has a very simple solution:

$$A_I = \frac{\underline{g}_{m,Q2}}{\underline{Y}_{10,Q2}} \left(1 + \frac{\underline{g}_{m,Q1}}{\underline{Y}_{10,Q1}}\right). \tag{5.149}$$

For further interpretation, turn again to our simple FET equivalent circuit (Figure 5.20), and recall the transit (cutoff) frequency $\omega_T \approx g_m/C_{GS}$. Equation (5.149) can then be written as

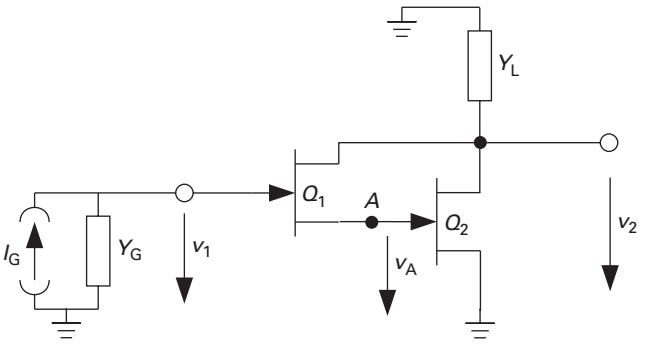

Fig. 5.31 Darlington amplifier configuration.

$$A_I = -\left(\frac{\omega_{T,Q1}\,\omega_{T,Q2}}{\omega^2} + \jmath\frac{\omega_{T,Q2}}{\omega}\right). \tag{5.150}$$

Assume now that $\omega_{T,Q1} = \omega_{T,Q2} = \omega_T$. For $\omega \ll \omega_T$, the current gain is now approximately the product of the current gains of the individual devices, but rolls off at -40 dB/decade, instead of -20 dB/decade. The frequency where $|A_I| = 1$ is $\sqrt{2/(\sqrt{5}-1)}\omega_T = 1.272\,\omega_T$.

Darlington amplifier

The CD/CS configuration is not quite the same as the popular Darlington [7] topology, shown in Figure 5.31. The difference is that in the Darlington amplifier, the drain of Q_1 is connected to the drain of Q_2. While the goal is similar, there are two noteworthy differences:

- The feedback admittance of device Q_1, $\underline{Y}_{12,Q1}$, is now in the path between the output and the input nodes, and not connected directly to ground as in the CD/CS configuration. Therefore, the Miller effect will be present at the input.
- The output current of Q_1 now also flows through the load. This changes the current gain equation. Using the same strong simplifications as in deriving Equation (5.150), we now find

$$A_I = -\left[\frac{\omega_{T,Q1}\,\omega_{T,Q2}}{\omega^2} + \jmath\left(\frac{\omega_{T,Q1}}{\omega} + \frac{\omega_{T,Q2}}{\omega}\right)\right]. \tag{5.151}$$

  Compared to the CD/CS amplifier, the Darlington has slightly more short-circuit current gain close to ω_T. If again both transistors are equal and equally biased, the frequency where $|A_I| = 1$ is $2\omega_T$. This is why this configuration is sometimes also called f_T *doubler*. The expression should be taken with a grain of salt. Remember that f_T is derived here from current gain, and that we neglected the feedback admittances in calculating Equation (5.151). The Miller effect disadvantage of the Darlington stage therefore does not show up in the simplification, but can significantly affect circuit performance for small values of Y_L. Further, the current gain rolls off with -40 dB/decade, which may lead to stability problems when feedback is applied around the stage. So your mileage may vary.

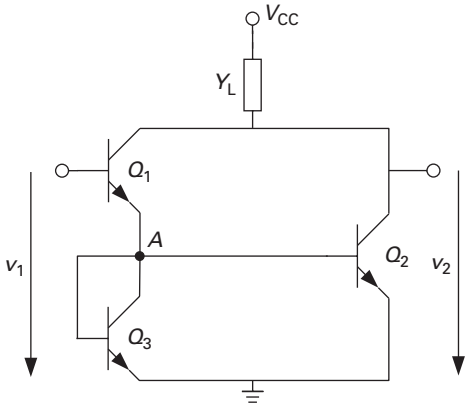

Fig. 5.32 Battjes f_T doubler circuit.

Battjes f_T Doubler

The well-known circuit shown in Figure 5.32, patented by C. R. Battjes [2], is essentially a Darlington amplifier (Q_1, Q_2) combined with a current mirror (Q_2, Q_3), which makes sure that both transistors in the signal path are operated with the same current. If they are also of equal size, they will have the same transit frequency. The circuit shown uses bipolar transistors (as in the patent), but the concept equally works with FETs. Note that the input capacitance of Q_3 needs to be accounted for – if Q_1/Q_3 and Q_2 are supposed to have the same current, then Q_2 and Q_3 need to have the same size, and the effective capacitance attached to node A approximately doubles (neglecting the Miller capacitance seen into Q_2).

Cascode amplifier

The cascode amplifier is a combination of a common-source (or common-emitter) with a common-gate (or common-base) topology. It was conceived as a way to overcome the Miller effect and first described in 1939 using two triode tubes [21], where the cathode of tube 2 was series-connected ('cascaded') to the anode of tube 1. The term *cascode* hence refers to *casc*aded an*ode*. Figure 5.33 shows a cascode realised using FETs.

Let us assess the input admittance first. Because Q_1 is in common-source configuration (node 0 grounded), we use Equation (5.109):

$$Y_1 = \underline{Y}_{10,Q1} + \underline{Y}_{12,Q1}\,(1 - A_{V,Q1}).$$

When calculating $A_{V,Q1}$, we recognise that the load admittance is the input admittance of Q_2 at node A:

$$A_{V,Q1} = -\frac{\underline{g}_{m,Q1} - \underline{Y}_{12,Q1}}{Y_{1,Q2} + \underline{Y}_{20,Q1} + \underline{Y}_{12,Q1}}.$$

The input admittance $Y_{1,Q2}$ is calculated using Equation (5.117), because Q_2 is in common-gate configuration:

$$Y_{1,Q2} = \underline{Y}_{10,Q2} + (\underline{g}_{m,Q2} + \underline{Y}_{20,Q2})\frac{Y_L + \underline{Y}_{12,Q2}}{\underline{Y}_L + \underline{Y}_{20,Q2} + \underline{Y}_{12,Q2}}.$$

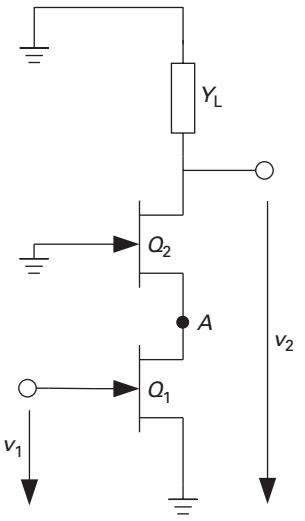

Fig. 5.33 Schematic of a cascode stage built with FETs (bias arrangement omitted).

Assume that $\underline{g}_{m,Q2} \gg \underline{Y}_{20,Q2}$, further $Y_L \gg \underline{Y}_{12,Q2}, \underline{Y}_{20,Q2}$. Then, the input admittance of Q_2 simplifies to

$$Y_{1,Q2} \approx \underline{Y}_{10,Q2} + \underline{g}_{m,Q2}.$$

With this simplification, the voltage gain of Q_1 is then

$$\begin{aligned} A_{V,Q1} &= -\frac{\underline{g}_{m,Q1} - \underline{Y}_{12,Q1}}{\underline{Y}_{10,Q2} + \underline{g}_{m,Q2} + \underline{Y}_{20,Q1} + \underline{Y}_{12,Q1}} \\ &\approx -\frac{\underline{g}_{m,Q1}}{\underline{Y}_{10,Q2} + \underline{g}_{m,Q2}}, \end{aligned}$$

further assuming that $\underline{g}_{m,Q2} \gg (\underline{Y}_{12,Q1} + \underline{Y}_{20,Q1})$.

The input admittance of the cascode finally becomes

$$Y_1 \approx \underline{Y}_{10,Q1} + \underline{Y}_{12,Q1} \left(1 + \frac{\underline{g}_{m,Q1}}{\underline{g}_{m,Q2}} \frac{1}{1 + \dfrac{\underline{Y}_{10,Q2}}{\underline{g}_{m,Q2}}} \right). \tag{5.152}$$

Using our simple FET equivalent circuit (Figure 5.20) this finally becomes

$$Y_1 = \jmath\omega \left[C_{GS,Q1} + C_{GD,Q1} \left(1 + \frac{g_{m,Q1}}{g_{m,Q2}} \frac{1}{1 + j\frac{\omega}{\omega_T}} \right) \right].$$

The suppression of the Miller effect is simply explained by the low voltage gain of the common-source stage $-g_{m,Q1}/g_{m,Q2}$ for low frequencies. Frequently, transistors Q_1 and Q_2 are chosen the same size, and since they share the same drain (or collector) current, it follows that $g_{m,Q1} = g_{m,Q2}$ and $A_{V,Q1} = -1$.

The calculation of the output admittance starts with the output admittance of transistor Q_2, which is in common-gate configuration (node 1 grounded), using Equation (5.118):

$$Y_2 = Y_{2,\mathrm{Q2}} = \underline{Y}_{12,\mathrm{Q2}} + \underline{Y}_{20,\mathrm{Q2}} \frac{Y_{2,\mathrm{Q1}} + \underline{Y}_{10,\mathrm{Q2}}}{Y_{2,\mathrm{Q1}} + \underline{Y}_{10,\mathrm{Q2}} + \underline{Y}_{20,\mathrm{Q2}} + \underline{g}_{\mathrm{m,Q2}}},$$

where $Y_{2,\mathrm{Q1}}$ is the output admittance of transistor Q_1 in common-source configuration (see Equation (5.111)):

$$Y_{2,\mathrm{Q1}} = \underline{Y}_{20,\mathrm{Q1}} + \underline{Y}_{12,\mathrm{Q1}} \left(1 + \frac{\underline{g}_{\mathrm{m,Q1}} - \underline{Y}_{12,\mathrm{Q1}}}{Y_{\mathrm{G}} + \underline{Y}_{10,\mathrm{Q1}} + \underline{Y}_{12,\mathrm{Q1}}} \right),$$

where Y_{G} is the admittance terminating the input port. We simplify the expressions by assuming that the feedback admittances are small and the corresponding terms can be neglected. Then $Y_{2,\mathrm{Q1}} \approx \underline{Y}_{20,\mathrm{Q1}}$ and the overall output conductance becomes

$$Y_2 \approx \underline{Y}_{20,\mathrm{Q2}} \left(1 + \frac{\underline{g}_{\mathrm{m,Q2}} + \underline{Y}_{20,\mathrm{Q2}}}{\underline{Y}_{10,\mathrm{Q2}} + \underline{Y}_{20,\mathrm{Q1}}} \right)^{-1}. \tag{5.153}$$

Because $\underline{g}_{\mathrm{m}}/\underline{Y}_{20}$ is the magnitude of the maximum voltage gain in common-source configuration (see Equation (5.108))), an additional sensible assumption is that $\underline{g}_{\mathrm{m,Q2}} \gg \underline{Y}_{20,\mathrm{Q2}}$. Then,

$$Y_2 \approx \underline{Y}_{20,\mathrm{Q2}} \left(1 + \frac{\underline{g}_{\mathrm{m,Q2}}}{\underline{Y}_{10,\mathrm{Q2}} + \underline{Y}_{20,\mathrm{Q1}}} \right)^{-1}. \tag{5.154}$$

In our simple FET example, we finally find

$$Y_2 \approx g_{\mathrm{DS,Q2}} \left(1 + \frac{g_{\mathrm{m,Q2}}}{g_{\mathrm{DS,Q1}} + \jmath\omega C_{\mathrm{GS,Q2}}} \right)^{-1},$$

and for the quasi-static case, $\omega \to 0$:

$$Y_2 \approx \frac{g_{\mathrm{DS,Q2}}}{1 + \dfrac{g_{\mathrm{m,Q2}}}{g_{\mathrm{DS,Q1}}}}.$$

The output admittance is therefore significantly reduced compared to the common-gate or common-source configurations.

The voltage gain of the cascode stage, $A_{\mathrm{V}} = A_{\mathrm{V,Q1}} \cdot A_{\mathrm{V,Q2}}$, is

$$A_{\mathrm{V}} \approx -\frac{\underline{g}_{\mathrm{m,Q1}}}{Y_{\mathrm{L}} + \underline{Y}_{20,\mathrm{Q2}}} \cdot \frac{\underline{g}_{\mathrm{m,Q2}} + \underline{Y}_{20,\mathrm{Q2}}}{\underline{g}_{\mathrm{m,Q2}} + \underline{Y}_{10,\mathrm{Q2}}}, \tag{5.155}$$

neglecting the feedback admittances. In the FET example,

$$A \approx \frac{g_{\mathrm{m,Q1}}}{Y_{\mathrm{L}} + g_{\mathrm{DS,Q2}}} \cdot \frac{1 + g_{\mathrm{DS,Q2}}/g_{\mathrm{m,Q2}}}{1 + \jmath\omega/\omega_{\mathrm{T}}},$$

or provided that $g_{m,Q2} \gg g_{DS,Q2}$ and $\omega \ll \omega_T$:

$$A_V = -\frac{g_{m,Q1}}{Y_L + g_{DS,Q2}}.$$

In summary, the cascode configuration provides a comparable voltage gain to the common-source topology, but its input admittance is significantly lower due to the reduction of the Miller capacitance and its output admittance is significantly higher.

Finally, an important side effect of the cascode shall be pointed out here: the real part of the output admittance may become negative. In practical devices, the assumption that parameter $\underline{Y}_{20}$ is purely real is not correct; a better approximation is $\underline{Y}_{20} = g_{DS} + \jmath\omega C_{DS}$. If we insert this into Equation (5.153) and separate real and imaginary parts, we find that the real part becomes negative if

$$g_{DS,Q1} \cdot g_{DS,Q2} < \omega^2 C_{DS,Q2} \cdot (C_{DS,Q1} + C_{GS,Q2}),$$

assuming that $g_{m,Q2} \gg g_{DS,Q1}$ along the way. Frequently, this can lead to amplifier instabilities, but it may also be used to compensate losses in travelling-wave amplifiers, as will be discussed later.

5.4.5 Differential amplifiers

An important component in many high-speed electronic circuits is the differential amplifier. One of the most influential pioneers of biomedical engineering, Otto Schmitt, is frequently held to be the father of the differential amplifier topology [34] – the ability of the differential amplifier to reject common-mode signals at its input is crucial for the measurement of weak bio-electric signals. Incidentally, he also invented the Schmitt trigger circuit.

A generic differential amplifier topology realised with FETs is shown in Figure 5.34. A first noteworthy difference between the amplifiers discussed so far is that the input and output voltages are not referenced to ground, but to the other input and output electrodes, respectively.

Figure 5.35 shows the small-signal representation of the differential amplifier, where the transistors are represented using the generic small-signal equivalent circuit from Figure 5.17. The transistors are identical.

Differential mode

Any combination of nodal input voltages $v_A = v'_A + v''_A$, $v_B = v'_B + v''_B$ can be split into a differential mode ($v'_A = -v'_B,\ v''_A = v''_B = 0$) and a common mode ($v''_A = v''_B,\ v'_A = v'_B = 0$). First, we concentrate on the differential mode. To calculate the voltage of the common mode v_0, we first loop through $v_A,\ v_{1,1},\ v_{1,2}$ and v_B:

$$-v'_A + v_{1,1} - v_{1,2} - v'_A = 0 \rightarrow v'_A = \frac{v_{1,1} - v_{1,2}}{2}.$$

On the other hand, $v_0 = v'_A - v_{1,1}$ and therefore

$$v_0 = -\frac{v_{1,1} + v_{1,2}}{2}.$$

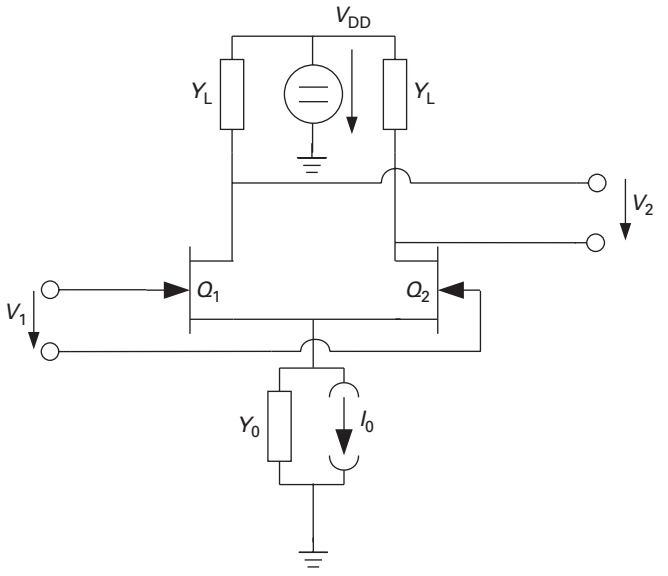

Fig. 5.34 Generic differential amplifier topology.

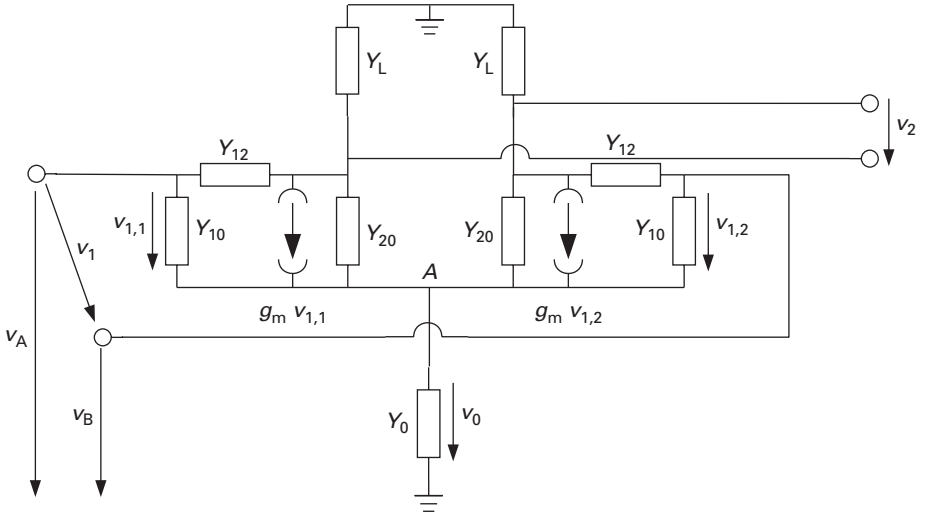

Fig. 5.35 Small-signal representation of the differential amplifier.

For symmetry reasons, $v'_A = -v'_B$ also implies $v_{1,1} = -v_{1,2}$ and hence

$$v_0 = 0$$

in differential mode! The common node *A* constitutes a *virtual ground*. This is a very important concept in high-speed circuit design, as it dramatically reduces problems with common-node impedances, such as in bond wires to ground, which otherwise may lead to a variety of feedback problems.

Now that A is grounded, the two halves of the differential amplifier reduce to standard common-source (or common-emitter) circuits which we have already analysed. The input voltage to the left half is $v_{1,1} = v_1/2$, while the right half receives $v_{1,2} = -v_1/2$.

Using Equation (5.109) to calculate the common-source input admittance $Y_{1,\mathrm{CS}}$ for the individual transistors, the differential mode input admittance is

$$Y_{1,\mathrm{d}} = \frac{i_1}{v_1} = \frac{Y_{1,\mathrm{CS}}}{2}. \tag{5.156}$$

Likewise, the output admittance is half the output admittance $Y_{2,\mathrm{CS}}$ for the common-source stage given by Equation (5.111):

$$Y_{2,\mathrm{d}} = \frac{i_2}{v_2} = \frac{Y_{2,\mathrm{CS}}}{2}. \tag{5.157}$$

Equation (5.107) is used to calculate the common-source voltage gain $A_{\mathrm{V,CS}}$. The differential voltage gain is then

$$A_{\mathrm{V,d}} = \frac{v_2}{v_1} = A_{\mathrm{V,CS}}. \tag{5.158}$$

Common mode

In common mode, both input terminals have the same potential to ground: $v''_{\mathrm{A}} = v''_{\mathrm{B}}$. The individual transistors are connected in parallel then at input and output, resulting in the equivalent circuit shown in Figure 5.36, and their y matrices can simply be added. We arrive at an equivalent transistor Q_{e} with the following y matrix:

$$[y_{\mathrm{Qe}}] = 2\begin{bmatrix} \underline{Y}_{10,1} + \underline{Y}_{12,1} & -\underline{Y}_{12,1} \\ \underline{g}_{\mathrm{m},1} - \underline{Y}_{12,1} & \underline{Y}_{20,1} + \underline{Y}_{12,1} \end{bmatrix}.$$

The parameters are those of the individual transistor. This problem can be treated using the results of the feedback discussion (see p. 323), converting the y matrix first to a z matrix, and adding the z matrix corresponding to Y_0, which is

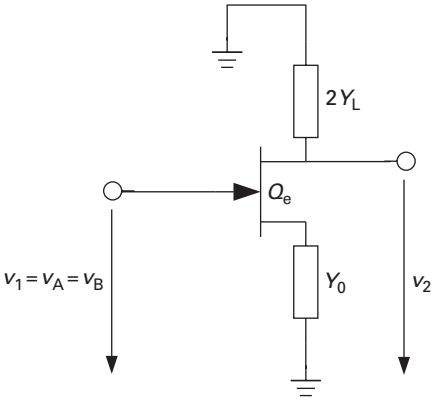

Fig. 5.36 Equivalent circuit of the differential amplifier under common mode excitation.

$$[Z_f] = \begin{bmatrix} \frac{1}{Y_0} & \frac{1}{Y_0} \\ \frac{1}{Y_0} & \frac{1}{Y_0} \end{bmatrix},$$

and converting back to a y matrix.

Here, we will consider a quick solution using a simplified equivalent circuit where we neglect both $\underline{Y}_{12,e}$ and $\underline{Y}_{20,e}$ for the transistor Q_e. Let $\underline{g}_{m,e} = 2g_{m,1}$ and $\underline{Y}_{10,e} = 2\underline{Y}_{10,1}$, as discussed. The voltage gain $A_{V,cm}$ for common-mode excitation is then

$$\begin{aligned} A_{V,cm} = \frac{v_2}{v_1} &= \frac{\frac{-2g_{m,1}}{(2Y_L)}}{1 + \frac{2}{Y_0 g_m}} \\ &= -\frac{g_{m,1}}{Y_L\left(1 + \frac{2g_m}{Y_0}\right)} \\ &\approx -\frac{Y_0}{2Y_L}, \end{aligned} \tag{5.159}$$

assuming $g_{m,1}/Y_0 \gg 1$. The output voltage here is taken between one of the output terminals and ground – the differential output voltage for common-mode excitation is 0, provided that the circuit is perfectly symmetrical. The voltage gain under differential excitation, but with the output voltage taken between one of the output terminals and ground, is

$$A_{V,d,gnd} = -\frac{g_{m,1}}{2Y_L}.$$

The ratio

$$\left|\frac{A_{V,d,gnd}}{A_{V,cm}}\right| = \left|\frac{g_{m,1}}{Y_0}\right| \tag{5.160}$$

is the *common-mode rejection ratio*, a measure for the suppression of common-mode input signals. We see that Y_0 should be as small as possible.

Neutralisation of differential amplifiers

The fact that the output voltages of a differential amplifier are exactly 180° out of phase can be used to elegantly eliminate the effect of the feedback capacitances in the transistors. This is shown in Figure 5.37. The capacitances C_n, which have to be exactly equal to C_{GD}, feed a current into the gate nodes of the two transistors, which is equal in magnitude, but of opposite sign, compared to the currents flowing through C_{GD}, cancelling these capacitances.

A more complex differential amplifier example

Differential amplifiers for high-speed applications are frequently more complex and exploit the special properties discussed in the section on basic amplifier topologies using two transistors. Figure 5.38 shows a cell common to many high-speed differential amplifiers. Transistors Q_1 and Q_2 are in common-collector configuration, connected to the transistor pairs Q_3, Q_5 and Q_4, Q_6, respectively, which form a differential cascode.

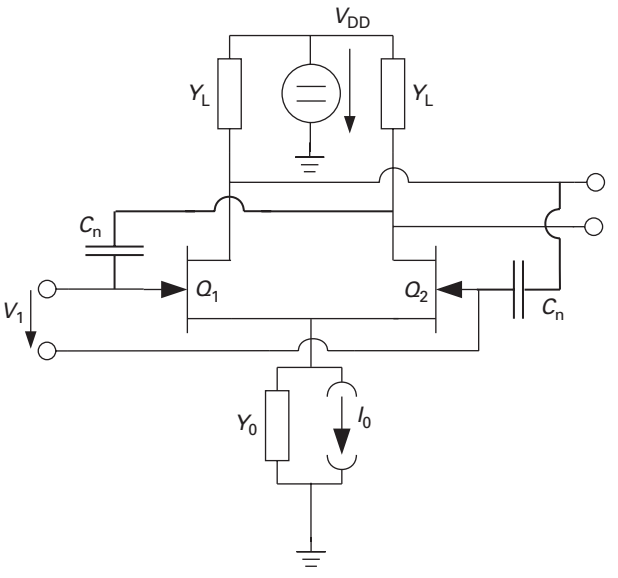

Fig. 5.37 Differential amplifier with neutralisation.

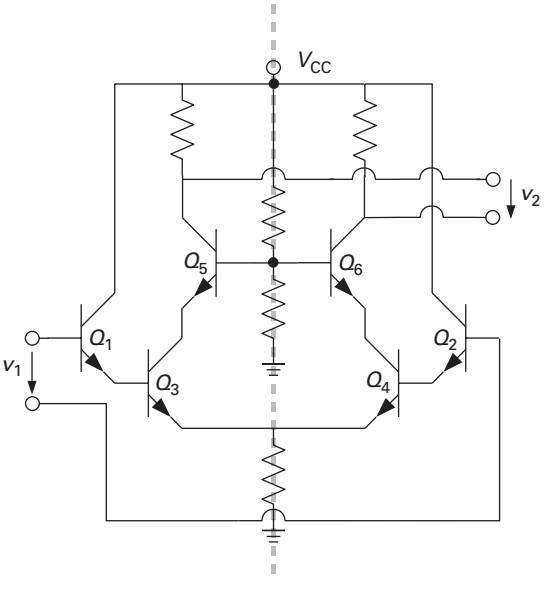

Fig. 5.38 A more complex differential amplifier example. The dashed line indicates the symmetry plane; all nodes along this plane are virtual grounds.

All nodes along the median, which is indicated as a dashed line, are virtual grounds, provided that the circuit is driven fully differentially. This is particularly interesting for the bases of Q_5 and Q_6, because proper grounding of the base terminal can be a problem in cascode stages – here, it is easy due to the virtual ground property. Equally, the emitters of Q_3 and Q_4 are properly grounded. The DC bias voltage terminal, V_{CC}, is also an RF ground, facilitating RF/DC decoupling.

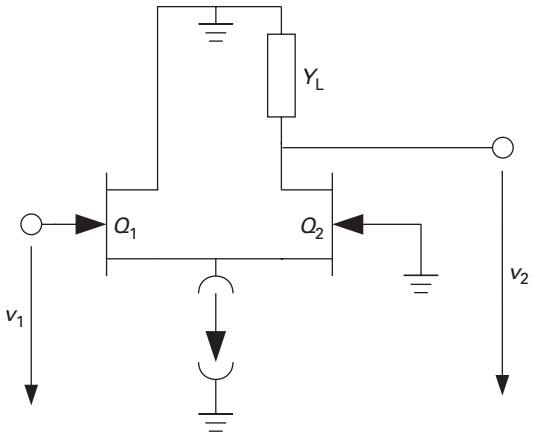

Fig. 5.39 Source-coupled amplifier schematic (bias elements omitted).

These advantages lead to an increasing use of differential topologies in micro- and millimetre-wave circuits. Drawbacks are the increased power consumption due to the doubled component count and the increased area consumption. Another problem may be on-wafer testing, due to the necessity for differential probes.

5.4.6 Source-coupled amplifier

The amplifier topology shown in Figure 5.39 has, at first glance, a configuration very similar to the differential amplifier. We immediately recognise the source-coupled pair and the common-current source. However, the amplifier is driven single-endedly and also has only a single output. Upon closer investigation, Q_1 is in common-drain and Q_2 in common-gate topology.

The idea is therefore very similar to the CD/CS amplifier discussed earlier. The common-drain input transistor creates a low input admittance, while the common-gate stage delivers the voltage gain. The Miller effect is eliminated, and the input is well isolated from the output.

Compared to the CD/CS amplifier, the input admittance is higher, because the input admittance of the common-gate transistor Q_2 is much higher than that of a comparably biased common-source transistor: $Y_{1,\mathrm{Q2}} \approx \underline{Y}_{10,\mathrm{Q2}} + \underline{g}_{\mathrm{m}}$; see Equation (5.117) with $Y_{\mathrm{L}} \gg \underline{Y}_{20,\mathrm{Q2}}$. The input admittance of the source-coupled amplifier is then

$$Y_1 = \underline{Y}_{12,\mathrm{Q1}} + \frac{\underline{Y}_{10,\mathrm{Q1}}}{1 + \dfrac{\underline{g}_{\mathrm{m,Q2}} + \underline{Y}_{10,\mathrm{Q2}}}{\underline{g}_{\mathrm{m,Q1}} + \underline{Y}_{10,\mathrm{Q1}}}}. \tag{5.161}$$

Using our simple FET equivalent circuit, $\underline{Y}_{10} = \jmath\omega C_{\mathrm{GS}}$ and

$$\underline{g}_{\mathrm{m}} + \underline{Y}_{10} = g_{\mathrm{m}}\left(1 + \jmath\frac{\omega}{\omega_{\mathrm{T}}}\right).$$

If the transistors therefore have the same ω_T, the input admittance is

$$Y_1 = \jmath\omega \left(C_{GD,Q1} + \frac{C_{GS,Q1}}{1 + \frac{g_{m,Q2}}{g_{m,Q1}}} \right).$$

It is purely capacitive and does not show the risk of a negative real part, which the CD/CS amplifier had posed.

The circuit can also be compared to the cascode – the source-coupled amplifier has a lower input admittance, is non-inverting and requires a lower supply voltage than the cascode, but the cascode requires less current, because the current through the common-gate stage is recycled in the common-source transistor.

5.4.7 Tuned amplifiers

Tuned amplifiers are commonly used at micro- and millimetre-wave frequencies when the fractional bandwidth is small. The fractional bandwidth is the required operational bandwidth divided by the centre frequency. For example, the 24 GHz license-free ISM band has a total allowed spectral width of 250 MHz, so any amplifier will need sensibly only a fractional bandwidth of 10^{-2}. Other applications, such as emerging ultra-wideband sensor and communications standards, will have fractional bandwidths which are orders of magnitude larger – the design of amplifiers for such systems will be treated in the next section (p. 350).

A typical tuned amplifier will use three fundamental circuit techniques:

(i) A resonant load – the load admittance goes through a minimum at the frequency of operation, maximising the voltage gain for a given transconductance.
(ii) Complex conjugate match at the input, ensuring that the available power from the source is delivered to the amplifier.
(iii) Complex conjugate match at the output, ensuring that the available power from the amplifier is delivered to the load.

For LNAs and power amplifiers, other matching strategies may apply for the input and output ports, respectively. These will be treated in the sections on LNA design (p. 365) and power amplifier design (p. 376). For now, we assume that achieving the maximum gain is our objective.

Resonant loads

Let us first investigate the resonant load, using the simple example of Figure 5.40. The schematic also indicates the generator and the equivalent input admittance of the following stage – it is essential to include at least the next-stage input admittance in the calculations, and due to feedback, the generator admittance will also have an effect, albeit more weakly.

The admittances $Y_{2,Q1}$ and $Y_{1,Q2}$ can typically be represented by a conductance in parallel with a capacitive reactance (exception – if the following stage is a common-gate

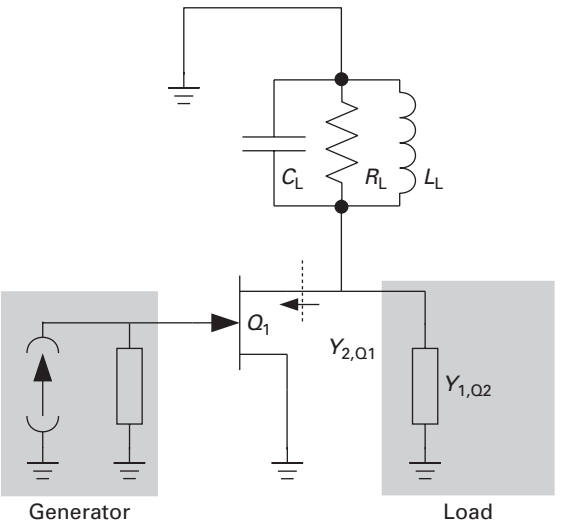

Fig. 5.40 Tuned amplifier stage with resonant load.

or common-base stage, the reactance may be inductive). These elements are absorbed into the load. The resulting reactances of the tank circuit are then

$$\begin{aligned} G_T &= R_L^{-1} + \mathrm{Re}(Y_{2,Q1} + Y_{1,Q2}) \\ C_T &= C_L + \frac{1}{\omega}\mathrm{Im}(Y_{2,Q1} + Y_{1,Q2}) \\ L_T &= L_L. \end{aligned}$$

The transfer function of the voltage gain is

$$A_V(\omega) = -\frac{g_m}{G_T}\frac{1}{1 + \jmath\left(\omega\frac{C_T}{G_T} - \frac{1}{\omega L_T G_T}\right)}, \tag{5.162}$$

which has its maximum at

$$\omega_0 = \frac{1}{\sqrt{L_T C_T}},$$

and its -3 dB corner points at

$$\omega_{\frac{1}{2}} = \sqrt{\frac{G_T^2}{4C_T^2} + \frac{1}{L_T C_T}} \pm \frac{G_T}{2C_T}.$$

The bandwidth between the -3 dB points is therefore

$$\Delta\omega = \omega_1 - \omega_2 = \frac{G_T}{C_T}. \tag{5.163}$$

This equation can be used to choose the proper G_T for the required bandwidth of the amplifier.

Using Equations (5.162) and (5.163), we find the product of the voltage gain at $\omega = \omega_0$ and the -3 dB bandwidth:

$$-A_V(\omega_0) \cdot \Delta\omega = \frac{g_m}{C_T}, \tag{5.164}$$

which is interestingly independent of frequency. This is due to simplifying assumptions, of course. In the *ansatz* for Equation (5.162) we used Equation (5.107) with the assumption that $\underline{g}_m \gg \underline{Y}_{12}$, hence that feedback is negligible, which is no longer true at very high frequencies.

A tuned tank circuit always bears the risk of amplifier instability. For the common-source amplifier in the example, we use Equation (5.109) to calculate the input admittance of the circuit, using the expression in Equation (5.162) for $A_V(\omega)$:

$$\begin{aligned} Y_1 &= \underline{Y}_{10} + \underline{Y}_{12}\left\{1 + \frac{\underline{g}_m}{G_T}\frac{1}{1 + \jmath\left[\omega\frac{C_T}{G_T} - \frac{1}{L_T G_T}\right]}\right\} \\ &= \underline{Y}_{10} + \underline{Y}_{12} + \underline{Y}_{12}\frac{\underline{g}_m}{G_T}\frac{1 - \jmath\left[\omega\frac{C_T}{G_T} - \frac{1}{\omega L_T G_T}\right]}{1 + \left[\omega\frac{C_T}{G_T} - \frac{1}{\omega L_T G_T}\right]^2}. \end{aligned}$$

If, as is usually the case, $\underline{Y}_{12} \approx \jmath\omega C_{GD}$, the third term in the sum has a negative real part for $\omega < \omega_0$. The risk of parasitic oscillations increases with increasing peak gain. Neutralisation measures as discussed already (p. 325) may become necessary in such cases.

Input and output matching networks

A common requirement in microwave amplifiers is that input and output admittances need to have a predefined value. There are two major reasons for this:

(i) If the input and output admittances are the complex conjugates of the source and load admittances, the source's available power is transferred to the amplifier, and the amplifier's available power is transferred to the load, resulting in the maximum power gain – this value is called the *maximum available gain* and will be discussed shortly.

(ii) To avoid standing waves on interconnecting transmission lines, the lines need to be terminated by their characteristic impedances at least at one end.

The characteristic impedance of a lossless transmission line is real; hence, input and output admittances are normally tuned to a purely real value where the circuit will interface with a transmission line. For internal nodes, however, this is not necessary – in fact, as we will see in our discussion of broadband amplifier techniques, at internal nodes impedance matching is often abandoned altogether, in favour of increased bandwidth, but with penalties in power gain.

Because the Smith chart (see p. 295) is the most important tool in solving matching problems, we will conduct the matching discussions using scattering parameters.

Most importantly, we need to translate source and load admittances as well as two-port input and output admittances into reflection coefficients. This is easily done:

$$\Gamma = \frac{Y_0 - Y}{Y_0 + Y} = \frac{Z - Z_0}{Z + Z_0}, \tag{5.165}$$

where $Y_0 = 1/Z_0$ is the normalising admittance, which is frequently 20 mS (correspondingly $Z_0 = 50\,\Omega$), but can be chosen arbitrarily.

We have seen that in two-ports which are not unilateral ($y_{12} \neq 0$, correspondingly $S_{12} \neq 0$), the input admittance depends on the load admittance, and the output admittance depends on the source admittance. In general terms and with the two-port expressed as a scattering matrix, the input (Γ_1) and output (Γ_2) reflection coefficients are (p. 300)

$$\Gamma_1 = S_{11} + \frac{S_{12}S_{21}\Gamma_\mathrm{L}}{1 - S_{22}\Gamma_\mathrm{L}} \tag{5.166}$$

$$\Gamma_2 = S_{22} + \frac{S_{12}s_{21}\Gamma_\mathrm{G}}{1 - S_{11}\Gamma_\mathrm{G}}, \tag{5.167}$$

where Γ_G and Γ_L are the generator and load reflection coefficients, respectively.

For simultaneous power match at input and output ports, we need these coupled equations to be satisfied:

$$\Gamma_\mathrm{G}^* = S_{11} + \frac{S_{12}S_{21}\Gamma_\mathrm{L}}{1 - S_{22}\Gamma_\mathrm{L}}$$

$$\Gamma_\mathrm{L}^* = S_{22} + \frac{S_{12}S_{21}\Gamma_\mathrm{G}}{1 - S_{11}\Gamma_\mathrm{G}},$$

where Γ^* is the complex conjugate of Γ. Solving these equations for the necessary generator and load reflection coefficients $\Gamma_{\mathrm{G,m}}$ and $\Gamma_{\mathrm{L,m}}$, we find

$$\Gamma_{\mathrm{G,m}} = \frac{C_1^*}{|C_1|}\left(\frac{B_1}{2|C_1|} - \sqrt{\frac{B_1^2}{|C_1|^2} - 1}\right), \tag{5.168}$$

with

$$B_1 = 1 - |S_{22}|^2 + |S_{11}|^2 - |\Delta(S)|^2$$

$$C_1 = S_{11} - \Delta(S)\,S_{22}^*,$$

where $\Delta(S)$ is the determinant of the scattering matrix. For the load reflection coefficients, we find likewise:

$$\Gamma_{\mathrm{L,m}} = \frac{C_2^*}{|C_2|}\left(\frac{B_2}{2|C_2|} - \sqrt{\frac{B_2^2}{|C_2|^2} - 1}\right), \tag{5.169}$$

with

$$B_2 = 1 - |S_{11}|^2 + |S_{22}|^2 - |\Delta(S)|^2$$

$$C_2 = S_{22} - \Delta(S)\,S_{11}^*.$$

Simultaneous input and output power match is not always possible, but requires a two-port to be *unconditionally stable* (see p. 303).

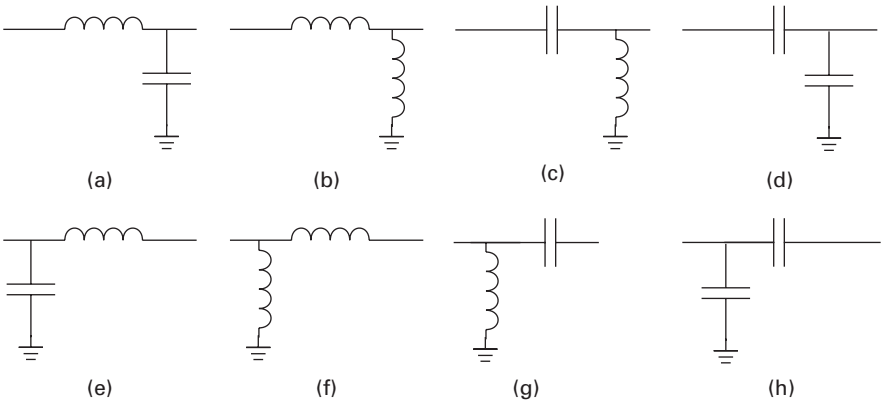

Fig. 5.41 Generic L network topologies.

Generally speaking, impedance transformation can be achieved using

- reactances L, C,
- transformers,
- transmission line impedance transformation.

At micro- and millimetre-wave frequencies, 'true' transformers based on coils are rarely used, because when realised on-chip using planar inductors, they tend to be very lossy, and additionally have high parasitic capacitances. So only impedance transformations using reactive networks and transmission line impedance transformation will be discussed here.

The most fundamental impedance transforming network is the *L network*, which can have any of the shapes shown in Figure 5.41.

There is always more than one topology which achieves the desired impedance transformation. This is an important observation, because other considerations need to be taken into account also. For example, the input port may have to be DC-blocked, in which case a topology with a series C may be suitable(cases c, d, g, or h in Figure 5.41), or DC bias may have to be supplied through the port, in which case a series L and no shunt L are needed (cases a or e). Likewise, it may be advantageous to ground the input port at low frequencies, favouring a topology with a shunt L (cases b, c, or f).

Figure 5.42 shows an example of an impedance matching problem, solved using several topologies. In all cases, the impedance in the lower left quadrant is the starting point and the centre of the Smith chart is the target.

- Path 1 uses an L in series with the start impedance and then a shunt L.
- Path 2 also starts with a series L, but a larger one, and then uses a shunt C.
- Path 3 starts with a shunt L, and then continues with a series L.
- Path 4 starts equally with a shunt L, but a smaller one, and then uses a series C to reach the required impedance.

The several options are best visualised using the Smith chart.

Table 5.1 Matching a complex load (100 Ω parallel to 2.5 pF) to 50 Ω using different L network topologies, $f = 1$ GHz

Path	Components	Bandwidth
1	$L_{S,1}$=3.1 nH, $L_{P,2}$=8.8 nH	508 MHz
2	$L_{S,1}$=11 nH, $C_{P,2}$=2.75 pF	547 MHz
3	$L_{P,1}$=26 nH, $L_{S,2}$=7.9 nH	561 MHz
4	$L_{P,1}$=6.1 nH, $C_{S,2}$=3.15 nH	324 MHz

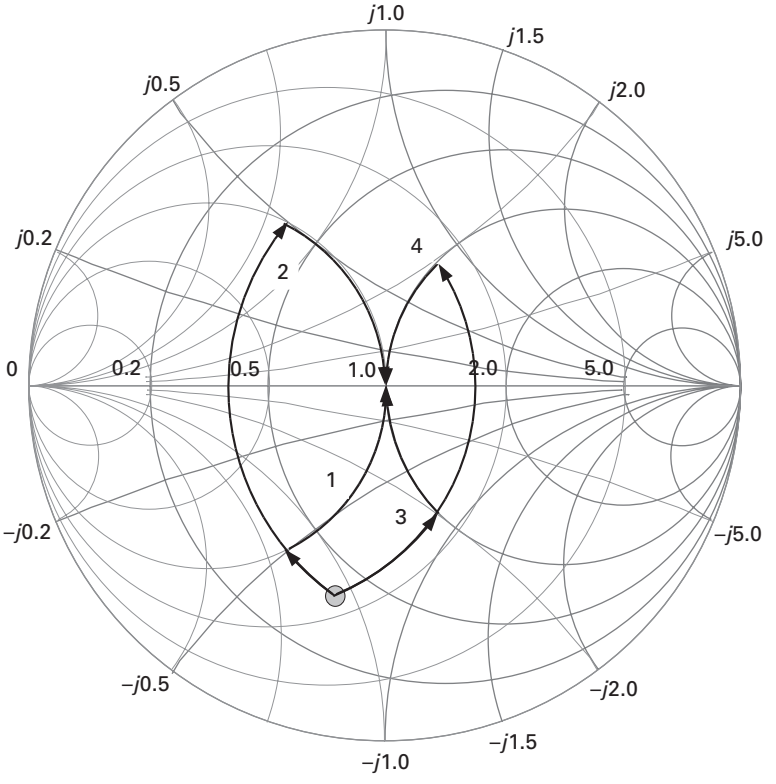

Fig. 5.42 Example for multiple impedance transformation paths using L networks.

Other aspects of matching networks need to be considered as well. This shall be done in an example, where a parallel RC load ($R = 100\,\Omega$, $C = 2.5\,\text{pF}$) is matched to $Z_0 = 50\,\Omega$ using the different topologies in Figure 5.42. The results are shown in Table 5.1.

First of all, we note that the matching bandwidth, defined as the difference between the frequencies where the reflection coefficient becomes $|\Gamma| > 0.32$ (return loss less than 10 dB), is vastly different – path 4 has less than 60% of the bandwidth of the others. Also, component values may become impractically larger for on-chip implementation – for example, $L_{P,1}$ for path 3.

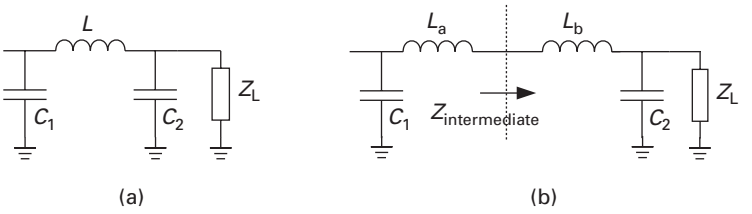

Fig. 5.43 Example of (a) a π-type matching network and (b) its decomposition into two cascaded L networks.

These calculations have been performed using ideal components. In practice, large-value spiral inductors also come with considerable series resistances, which is another aspect to consider.

π networks are an extension of L networks – they are best thought of as being separated into two L networks, as shown in Figure 5.43. The first L network transforms to an intermediate impedance $Z_{\text{intermediate}}$, which is then transformed by the second L network to the desired value. π networks offer an additional degree of freedom, so we can additionally design for different matching bandwidths. They are additionally attractive, because they allow the absorption of interconnect parasitics into the matching network – e.g. bond pad parasitics on chip and in the package (or on the PCB board) can form part of C_1 and C_2, while the bond wire inductance can be absorbed into L.

Other combinations of L-type networks exist and can be useful for specific matching problems, but this is beyond the scope of this book.

Figure 5.44 shows three examples of compact tuned amplifiers in an 80 GHz f_T Si/SiGe HBT technology [6]. The amplifiers share the same basic topology – three cascaded cascode stages with resonant loads and LC interstage matching using spiral inductors. Additionally, inductive emitter degeneration (Equation (5.141)) is used to assist the match by increasing the real part of the input impedance. The use of concentrated reactances, even at millimetre-wave frequencies, leads to an extremely compact layout.

Transmission line segments can also be used to transform impedances. Assuming lossless transmission lines, the input impedance looking into a transmission line of length l and characteristic impedance Z_0, terminated by an impedance Z_L, is

$$Z_1 = Z_0 \frac{Z_L + \jmath Z_0 \tan\left(2\pi \frac{l}{\lambda'}\right)}{Z_0 + \jmath Z_L \tan\left(2\pi \frac{l}{\lambda'}\right)}, \tag{5.170}$$

where λ' is the wavelength on the transmission line,

$$\lambda' = \frac{c_0}{f\sqrt{\epsilon_{r,\text{eff}}}}.$$

A very popular example is the *quarter-wavelength transformer*. In case $l = \lambda'/4$, the input impedance becomes

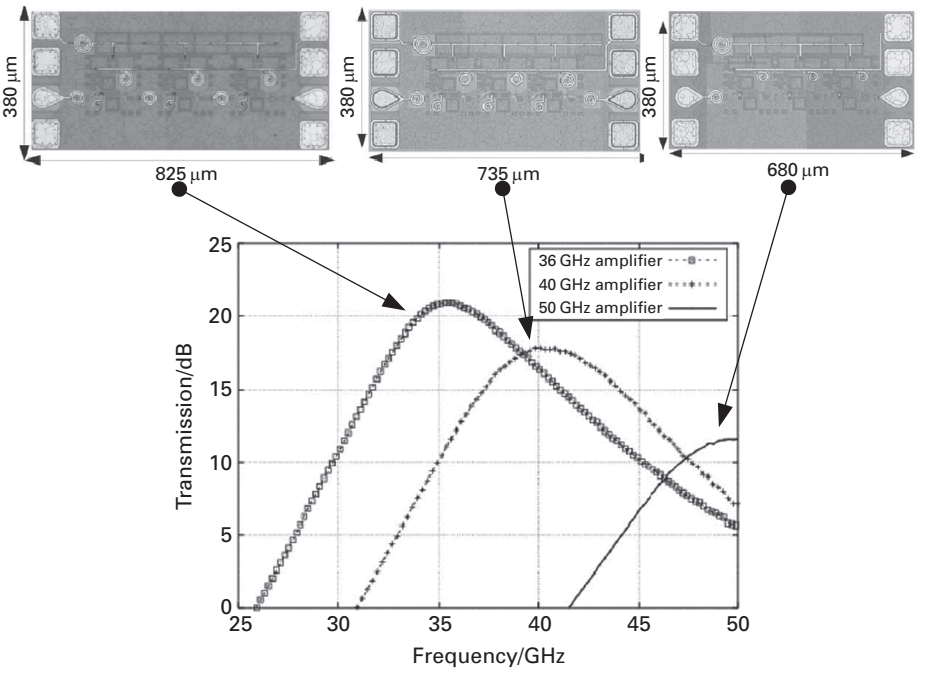

Fig. 5.44 Tuned millimetre-wave amplifiers in a Si/SiGe HBT technology, using LC loads and matching networks. (After [6])

$$Z_1 = \frac{Z_0^2}{Z_L}. \tag{5.171}$$

In other words, to match two impedances Z_A, Z_B, they need to be connected with a transmission line which is $\lambda'/4$ long and has a characteristic impedance of $Z_0 = \sqrt{Z_A Z_B}$. Quarter-wave transmission line sections are also called *impedance inverters* – the reason is obvious from Equation (5.171).

Transmission lines open up additional possibilities in matching. This is shown in Figure 5.45, again using the same start impedance as above:

- In path 1, a transmission line section of impedance $Z_0 = 50\,\Omega$ is used first to make the impedance real. The intermediate impedance is $14.5\,\Omega$; hence the quarter-wave section must have an impedance of $\sqrt{50 \cdot 14.5} = 26.9\,\Omega$.
- Path 2 first uses a series inductance to make the impedance real, the intermediate impedance is $28.7\,\Omega$. The quarter-wave section then needs to have a characteristic impedance of $37.8\,\Omega$.
- Path 3, finally, uses a shunt inductance to make the impedance real ($Z_{\text{intermediate}} = 100\,\Omega$) and a quarter-wave section with $Z_0 = 70.7\,\Omega$.

Option 2 has the widest matching bandwidth, but the transmission line in option 3 is likely the easiest to realise.

With increasing frequency, tuned amplifiers using transmission line segments become increasingly interesting, because spiral inductors are especially difficult to realise and

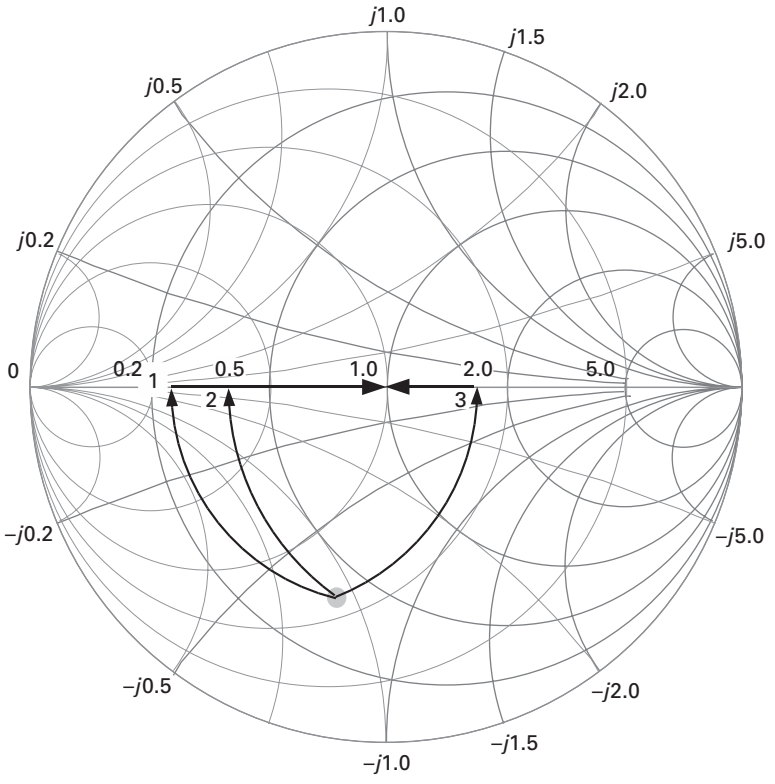

Fig. 5.45 Matching examples using quarter-wave transmission line transformers.

model, and the main objection against the use of transmission lines – their physical size in layout – becomes irrelevant as the wavelength shrinks. Figure 5.46 shows an example. The IC represents a three-stage fully differential amplifier for 77 GHz automotive RADAR systems, realised in a 190 GHz f_T Si/SiGe BiCMOS technology [5]. The amplifier provides 16 dB gain while consuming 90 mW from a 3 V supply. Thin-film microstrip[3] lines (TFMSLs) are used here for impedance matching purposes. Due to the high frequency, the resulting IC is still very compact ($740 \times 540\,\mu m^2$ chip size).

5.4.8 Broadband amplifier techniques

Tuned loads and reactive impedance matching networks are not suitable for amplifiers with large fractional bandwidths, such as those used in high-speed fibre-optic systems, micro/millimetre-wave instruments, many military systems with high frequency ability, or impulse-radio ultra-wideband systems. All of these applications need amplifiers where the gain must be flat over a wide frequency range (often the ratio of upper to

[3] In TFMSLs, the ground plane is realised on top of the substrate. This shields the signal line from the lossy Si substrate, but leads to very narrow signal lines.

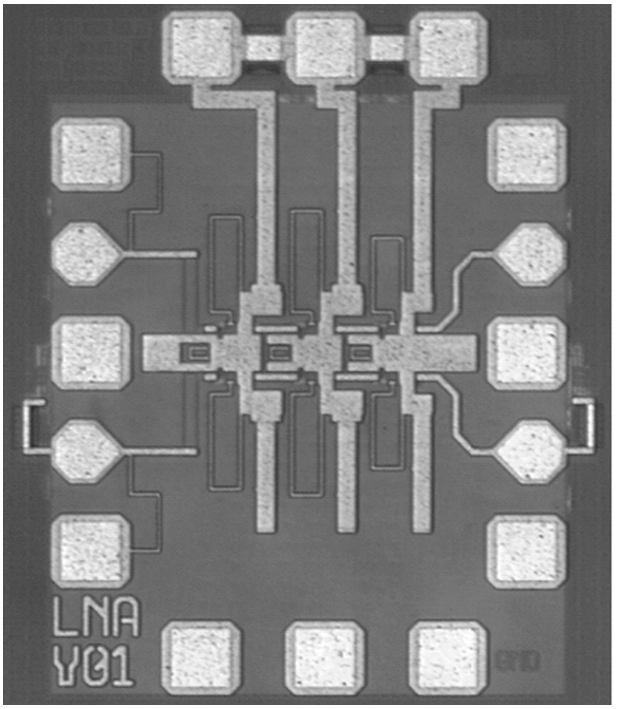

Fig. 5.46 Fully differential Si/SiGe HBT amplifier for 77 GHz, using tuned transmission lines. (After [5])

lower cutoff frequency exceeds the factor of two – *multi-octave* bandwidths), and almost always the input and output return loss also needs to stay below a specified value over the full frequency range.

In the following section, we will discuss some common techniques which prove useful in the realisation of amplifiers with very large bandwidths using concentrated circuit components. Discussion of distributed amplification, which is also a very important concept for wideband amplifiers, will start on p. 354.

Shunt peaking

We have already emphasised the importance of the characteristic time constant in the discussion of multi-stage amplifier topologies (p. 330). We will see that broadband amplifier design always comes down to modifying these internal characteristic time constants.

Consider the simple cascading of common-source amplifiers, shown in Figure 5.47 together with a strongly simplified equivalent circuit. The load resistance and the input capacitance of the following stage are combined into an equivalent impedance to ground Z_{eq}. Using Z_{eq}, the transadmittance of the cascaded stage can be expressed as

$$Y_T = \frac{i_2}{v_1} = -g_{m,1} g_{m,2} Z_{eq} = -\frac{R_L g_{m,1} g_{m,2}}{1 + \jmath\omega R_L C_{GS,2}}. \tag{5.172}$$

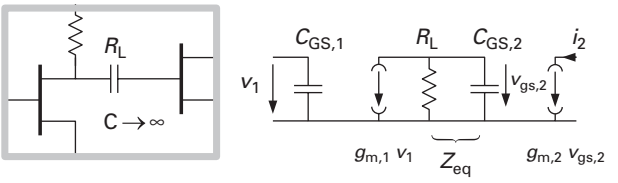

Fig. 5.47 Intermediate node of two cascaded common-source amplifiers, with small-signal equivalent circuit.

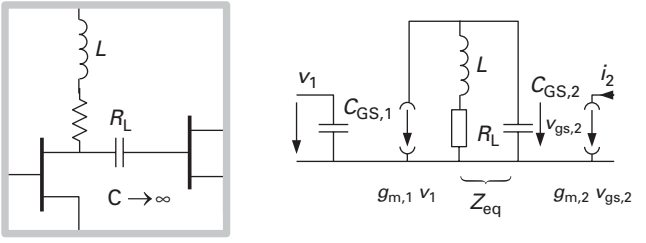

Fig. 5.48 Cascade connection of two common-source amplifiers with shunt peaking inductor.

Obviously, $R_L C_{GS,2}$ is the characteristic time constant of the intermediate node, which limits the bandwidth to

$$\omega_1 = \frac{1}{R_L C_{GS,2}} = \frac{\omega_{T,2}}{g_{m,2} R_L}, \tag{5.173}$$

using $\omega_T = g_m / C_{GS}$.

We will now partially compensate the capacitive reactance by connecting an inductor in series with the load resistor (see Figure 5.48). The transadmittance now becomes

$$\begin{aligned} Y_T &= -g_{m,1} g_{m,2} Z_{eq} \\ &= -g_{m,1} g_{m,2} R_L \frac{1 + \frac{\jmath \omega L}{R_L}}{1 - \omega^2 L C_{gs,2} + \jmath \omega R_L C_{gs,2}}. \end{aligned} \tag{5.174}$$

Introducing

$$\tau = \frac{L}{R_L};\ m = \frac{R_L^2 C_{gs,2}}{L} = \frac{1}{\omega_1 \tau},$$

we rewrite Equation (5.174) [25]:

$$Y_T = -g_{m,1} g_{m,2} R_L \frac{1 + \jmath \left(\frac{\omega}{\omega_1} \right) m^{-1}}{1 - \left(\frac{\omega}{\omega_1} \right)^2 m^{-1} + \frac{\jmath \omega}{\omega_1}}. \tag{5.175}$$

The new -3 dB cutoff frequency is

$$\omega_2 = \omega_1 \sqrt{\left(\frac{-m^2}{2} + m + 1 \right) + \sqrt{\left(\frac{-m^2}{2} + m + 1 \right)^2 + m^2}}. \tag{5.176}$$

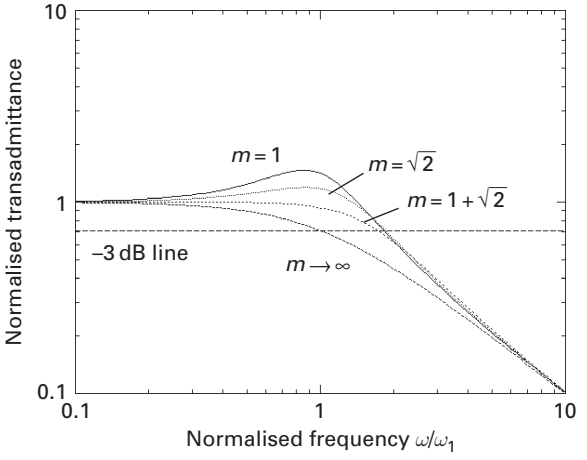

Fig. 5.49 Normalised transadmittance of an amplifier cascade with shunt peaking versus frequency, for different values of parameter m.

Equation (5.176) is maximum for

$$m = \sqrt{2},$$

or finally

$$\tau = \frac{1}{\sqrt{2}\omega_1}. \tag{5.177}$$

Figure 5.49 plots the normalised transadmittance $Y_T/\left(g_{m,1}g_{m,2}R_L\right)$ versus the normalised frequency, and for several values of m. We note that

- we can achieve 1.8-fold increase in bandwidth;
- the increase in bandwidth comes at the expense of gain flatness;
- however, for $m = 1 + \sqrt{2}$, the response becomes *maximally flat* with only a marginal decrease in bandwidth.

Feedback techniques

We had already seen (Figure 5.29) that a parallel RC combination in series–series feedback can be used to completely eliminate the dominant pole in the frequency response of the transadmittance. Let us consider a somewhat more complicated example now where the amplifier is loaded by a complex load formed by a resistor and a capacitor in parallel – the typical equivalent circuit of a following amplification stage. The small-signal equivalent circuit is shown in Figure 5.50. The voltage gain is

$$A_V = -g_m R_L \frac{1 + \jmath\omega\tau_S}{(1 + \jmath\omega\tau_L)\left[1 + \frac{g_m}{G_S} + \jmath\omega\left(\tau_S + \frac{g_m}{G_S}\frac{\omega}{\omega_T}\right)\right]}, \tag{5.178}$$

where

$$\tau_S = R_S C_S;\ \tau_L = R_L C_L;\ \omega_T = \frac{g_m}{C_{GS}}.$$

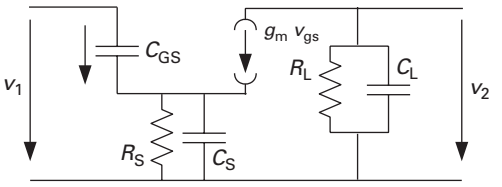

Fig. 5.50 Bandwidth enhancement using series–series feedback.

The enumerator term can now be used to cancel one of the denominator poles:

- If $1 + \jmath\omega\tau_L$ dominates, then $\tau_S = \tau_L$ is the proper choice.
- If the second term dominates, then choose $\tau_S = \omega_T^{-1}$. This corresponds to the solution already discussed in Equation (5.145).

5.4.9 Distributed amplification

The amplifier topologies discussed so far employed concentrated circuit elements and are as such not very different from topologies employed at lower frequencies. The distributed nature of components, especially interconnect lines, only comes in at the layout stage. In the wideband amplifier technique we will discuss now, the transmission line nature is consciously used to establish *distributed amplification*.

A common problem in achieving high gain at microwave frequencies is that the necessary large transconductance of the amplifying device requires a large device size (source width or emitter area), which in turn invariably increases the input capacitance. In FETs, in a first-order approximation, the ratio of transconductance to input capacitance is the transit frequency: $g_m/C_{GS} = \omega_T$. In a common-source amplifier, the dominant time constant at the input is therefore

$$\tau_1 = Z_G C_{GS} = Z_G \frac{g_m}{\omega_T} \approx -Z_G\, Y_L\, \frac{A_V}{\omega_T},$$

where Z_G is the generator admittance Y_L the load admittance and A_V the quasi-static voltage gain in common-source configuration. The input time constant is therefore directly linked to the voltage gain of the cell, for a given load admittance.

In narrowband amplifiers, we may be able to compensate for the input capacitance using a matching network, as we have seen. Very wideband amplifiers, however, preclude the use of tuned networks.

To find a way around the input capacitance limitation, we follow two fundamental steps:

(i) Instead of using one large device, we will use several smaller ones to deliver the needed overall transconductance.
(ii) The input (and output) capacitances will then be absorbed into an artificial transmission line.

The second step is the most crucial one. To understand this concept, remember that any transmission line can be modelled using a ladder-type network of concentrated

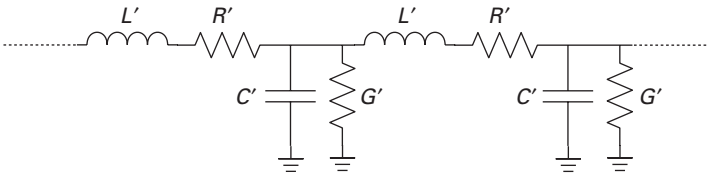

Fig. 5.51 Lumped-element equivalent circuit of a transmission line.

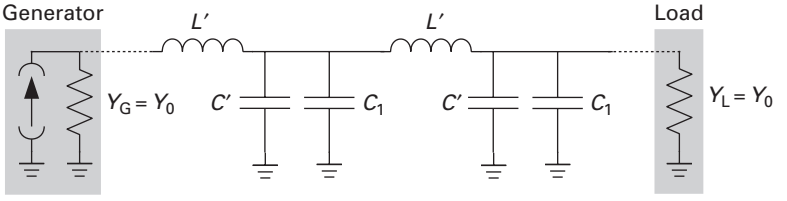

Fig. 5.52 Lossless transmission line loaded with additional shunt capacitances.

elements, such as shown in Figure 5.51. The line is characterised by its distributed inductance L', capacitance C', and the distributed series (R') and shunt (G') losses. The characteristic impedance Z_0 and the propagation constant γ of the line are then:

$$Z_0 = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \tag{5.179}$$

$$\gamma = \sqrt{(R' + j\omega L')(G' + j\omega C')}. \tag{5.180}$$

Note that the propagation constant $\gamma = \alpha + j\beta$, where α is the attenuation constant and β is the phase constant. In many cases, the losses can be neglected ($R' \ll \omega L'$, $G' \ll \omega C'$) and we obtain the simple relationships:

$$Z_0 \approx \sqrt{\frac{L'}{C'}} \tag{5.181}$$

$$\beta = \omega\sqrt{L'C'}. \tag{5.182}$$

This opens up a fundamental idea: any capacitance to ground can be made to disappear if it is absorbed into a transmission line – it will simply lower the characteristic impedance, and increase the phase constant.

Consider Figure 5.52. The lossless transmission line is loaded by additional shunt capacitances C_1. The transmission line parameters are now

$$Z_0 = \sqrt{\frac{L'}{C' + \frac{C_1}{l}}} \tag{5.183}$$

$$\beta = \omega\sqrt{L'\left(C' + \frac{C_1}{l}\right)}. \tag{5.184}$$

The parameter l is the length of the transmission line segment between each shunt capacitance.

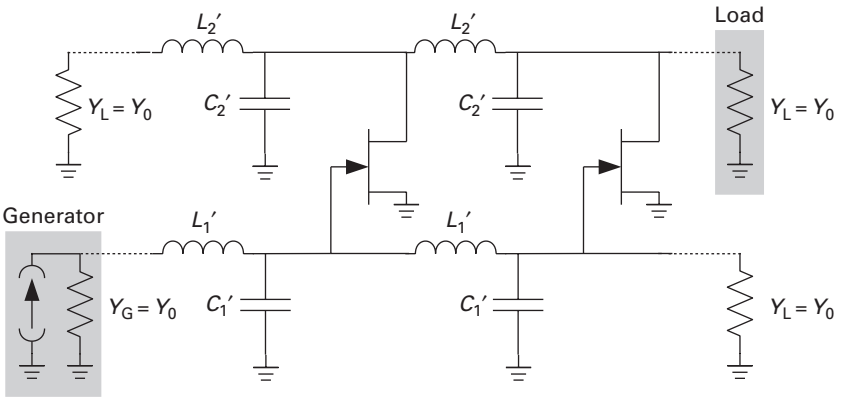

Fig. 5.53 Distributed amplifier concept using FETs in common-source configuration.

Provided that L' and C' are chosen in such a way that $Y_G = Y_L = Z_0^{-1}$, the transmission between generator and load is unaltered by the presence of the additional shunt capacitances!

This observation is not new at all. Its earliest implementation is in the *Pupin coils*, periodically inserted series loading coils (increasing L' in our example) which compensate for the capacitance to ground of telegraph and telephony lines. They were invented in 1894 by Serbian physicist Mihajlo Idvorski Pupin, following earlier suggestions by Oliver Heaviside in 1893.

Of course, the LC combination also acts as a low-pass filter. The frequency

$$\omega_{\text{Bragg}} = \frac{1}{l\sqrt{L'(C' + C_1/l)}} \tag{5.185}$$

is called the *Bragg frequency* of the transmission line structure. The length l must be chosen such that the Bragg frequency is significantly above the intended frequency of operation.

Distributed amplifier structures using electron tubes were first described by W. S. Percival in his 1937 patent [30].

General design procedure

We will now apply the concept to an arrangement of FETs in common-source configuration along two transmission lines, connecting the inputs and outputs, as shown in Figure 5.53. Note that the transmission lines at input and output have different inductance and capacitance per unit area. The loading capacitances are now the imaginary parts of the input and output admittances of the common-source gain cells. Using Equations (5.109) and (5.111) and a simplified FET equivalent circuit, we write for the shunt capacitance loading the input line:

$$C_1 = C_{GS} + C_{GD}\left(1 + \frac{g_m}{2Y_0}\right), \tag{5.186}$$

provided that the output transmission line is terminated in its characteristic admittance Y_0.

The shunt capacitance loading the output transmission line is

$$C_2 = C_{DS} + C_{GD}\left(1 + \frac{g_m}{2Y_0}\right), \tag{5.187}$$

where C_{DS} is the parasitic drain–source capacitance.

The unloaded input and output transmission lines must be chosen such that

- the loaded characteristic impedances correspond to generator and load impedances and
- the phase delays between corresponding nodes on the (loaded) input and output lines are equal.

Assuming identical generator and load impedances, $Z_G = Z_L = Z_0$, we find

$$Z_1 = \sqrt{\frac{L_1'}{C_1' + \frac{C_1}{l_1}}} \stackrel{!}{=} Z_0 \tag{5.188}$$

$$Z_2 = \sqrt{\frac{L_2'}{C_2' + \frac{C_2}{l_2}}} \stackrel{!}{=} Z_0. \tag{5.189}$$

The phase synchronism requirement translates into

$$\beta_1 l_1 = \beta_2 l_2$$

$$l_1 \cdot \sqrt{L_1'\left(C_1' + \frac{C_1}{l_1}\right)} = l_2 \cdot \sqrt{L_2'\left(C_2' + \frac{C_2}{l_2}\right)}. \tag{5.190}$$

The difference in the unit amplifier cell input and output capacitances may result in very different design parameters for the input and output transmission lines. Figure 5.54 shows this in a practical example. The distributed amplifier shown was fabricated in an experimental Si/SiGe HFET technology [1]. The transmission lines are realised in coplanar waveguide form. The difference in geometry for the input (gate) and output (drain) lines is clearly visible.

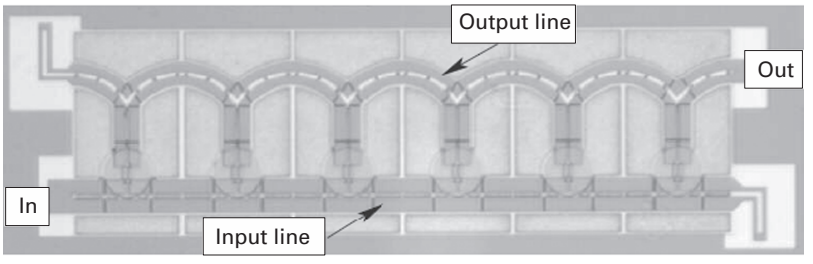

Fig. 5.54 Chip micrograph of a distributed amplifier with 32 GHz bandwidth, realised in a Si/SiGe HFET technology (P. Abele, I. Kallfass, M. Zeuner, J. Müller, Th. Hackbarth, D. Chrastina, H.v.Känel, U. König, and H. Schumacher, *Electronics Letters*, Vol. 39, pp. 1448–1449, 2003. © 2003 IEEE).

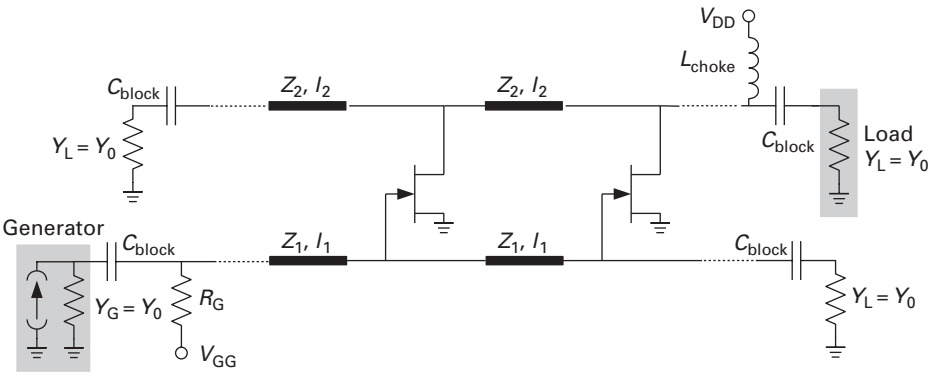

Fig. 5.55 Distributed amplifier with bias arrangement.

The terminating impedances for the input and output lines are placed off-chip in this example – which brings us to a general problem we did not address so far. The distributed amplifier concept in Figure 5.53 did not include the bias arrangement. If we apply a gate voltage to the input and a drain voltage to the output line, a constant current would flow through the terminating impedances attached to the ends of the transmission lines opposite to the input and output ports – resulting in generally unacceptable power dissipation there. The terminating impedances therefore need to be galvanically isolated from the transmission lines. A more practical schematic for a distributed amplifier would therefore look like Figure 5.55. The bias-related elements C_{block} and L_{choke} set the lower cutoff frequency. If a very low lower cutoff frequency is desired, then the on-chip realisation especially of the blocking capacitors may be a significant challenge. L_{choke} is generally placed off-chip.

Gain and loss in distributed amplifiers

Without any losses, the theoretical voltage gain of a distributed amplifier with n stages should be

$$A_{\text{V}} = n\, g_{\text{m}} \frac{Z_0}{2}, \tag{5.191}$$

where g_{m} is the transconductance of the individual cell and Z_0 the characteristic impedance of the output line.

So far, we assumed that the transmission lines were lossless, and that the input and output admittances of the unit amplifier cells were purely capacitive. The latter assumptions particularly are too bold, of course, and we need to assess how the resistive parts of the input and output admittances impact distributed amplifier performance.

In most calculations so far, the gate (or base) series resistance was neglected. This we will abandon here. For the case of a FET, the input line is then loaded with an complex admittance:

$$Y_1 = \frac{\jmath\omega C_1}{1 + \jmath\omega C_1 R_G} \tag{5.192}$$

$$= \omega^2 \frac{R_G C_1^2}{1 + \omega^2 R_G^2 C_1^2} + \jmath\omega \frac{C_1}{1 + \omega^2 R_G^2 C_1^2}, \tag{5.193}$$

where R_G is the gate resistance and C_1 the input capacitance as before. As long as $\omega \ll (R_G C_1)^{-1}$, losses due to R_G need not be accounted for, but they will increase strongly for higher frequencies.

For the output line, some attenuation is always present due to the real part of $\underline{Y}_{20}$ in Equation (5.111), which is g_{DS} in FETs:

$$Y_2 = g_{DS} + \jmath\omega C_2, \tag{5.194}$$

where C_2 is the output capacitance as before. The loss introduced to the drain line is hence frequency-independent.[4]

When the number of stages, n, is increased, the power consumption scales linearly. However, with increasing n, the losses introduced by the amplifier cells become more important and lead to a situation where the gain scales sub-linearly. This introduces a practical limitation to the number of stages. For a detailed analysis, refer to Beyer *et al.* (1984) [4].

Distributed amplifier variations

Matching input and output capacitances

A common problem in distributed amplifiers is that the amplifier cell input capacitance C_1 is much larger than the output capacitance C_2. In turn, the unloaded characteristic impedance of the output line will be significantly smaller than that of the input line. This is significant because the dispersion characteristics of the lines depend on their geometries – different geometries lead to different dispersions, and phase synchronism between input and output lines is increasingly lost with rising frequency.

A simple technique is to increase the output capacitance. This can be done easily using a transmission line stub between the amplifier cell output and the output transmission line. As the amplifier output shows a reasonably high impedance, the transmission line stub acts capacitively when seen from the output line. Figure 5.56 shows this simple concept, which is used in many practical amplifier examples.

The input capacitance can also be lowered by introducing a series capacitance in the unit cell input port. This leads to a capacitive voltage division between the series capacitor and the input impedance of the amplifier, and hence a reduction in gain, but depending on the application, this may be tolerated for the benefit of an increased bandwidth. To allow proper biasing, the capacitor must be bridged with a high-value resistor which has no influence on the RF performance. The measure is shown in Figure 5.57.

[4] The loss due to additional drain (or collector) resistances can be neglected unless they are excessive.

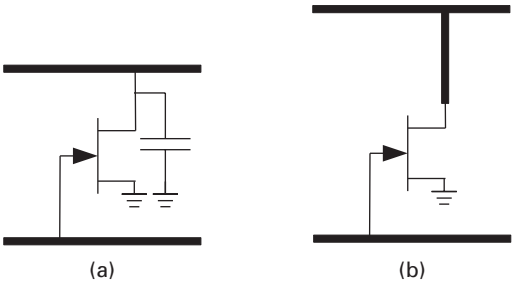

Fig. 5.56 Distributed amplifier unit cell with increased output capacitance: (a) concept and (b) implementation using a transmission line stub.

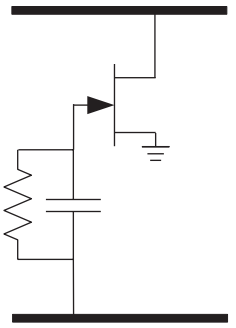

Fig. 5.57 Input capacitance reduction using a series capacitor.

By changing the series capacitance value along the input transmission line (lower towards the generator and higher towards the termination), the input voltage across the amplifying device can be made equal despite the decreasing signal on the transmission line.

Distributed amplifiers with a cascode cell

Despite the potential of the distributed amplifier concept to eliminate input and output capacitances by embedding them into an artificial transmission line, there are good reasons to keep input and output capacitances low. One reason is that high input and output capacitances force the unloaded characteristic impedances of the lines to be very high – the signal-carrying lines then have to be very narrow and will exhibit high ohmic loss. Further, a high input capacitance means that the loss due to the gate resistance will start to matter at much lower frequencies (see Equation (5.192)).

Choosing a cascode as the amplifier unit cell is therefore a logical choice. A simplified configuration is shown in Figure 5.58.

We had seen that the cascode gain cell is prone to producing a negative real part of the output admittance (see p. 336). Here, this effect may be used with benefit to compensate for losses on the output line, but amplifier stability has to be carefully checked, especially at higher frequencies.

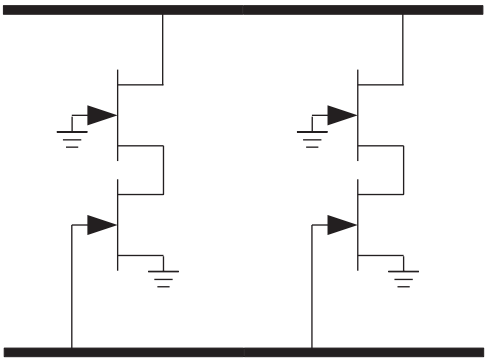

Fig. 5.58 Cascode gain cells in a distributed amplifier structure (bias elements not shown).

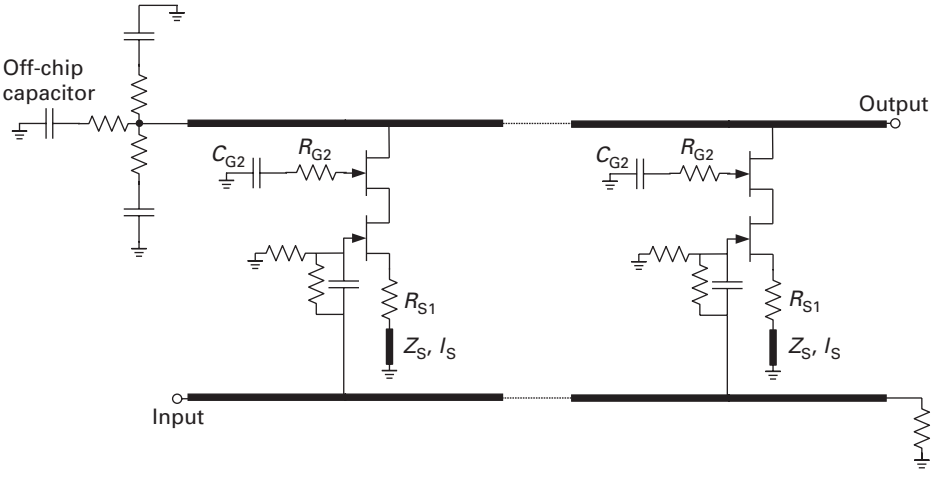

Fig. 5.59 Practical distributed amplifier design using (Al,Ga)As/InGaAs pHEMTs (bias circuitry omitted).

Practical distributed amplifier examples

40 GHz bandwidth distributed amplifier using GaAs pHEMTs

Figure 5.59 shows the schematic diagram of a practical distributed amplifier using a pseudomorphic HEMT process [16]. Several of the measures discussed above have been taken here. The unit cell has a cascode topology, but additionally the input capacitance was reduced using a series capacitor in the input line. The series capacitor is bridged using a high-value resistor; the additional resistor to ground at the gate node improves gain flatness at low frequencies.

The gate termination does not have a DC blocking capacitor here, because the gate line is held at 0 V – the source resistor R_{S1} provides the slightly negative gate–source voltage. Note the elaborate drain termination. This is rather typical of distributed amplifiers for fibre-optic systems where a lower cutoff frequency in the kHz range is required: a broadband termination is created using several RC networks with staggered time constants. The largest capacitor (100 nF in this case) is necessarily placed off-chip.

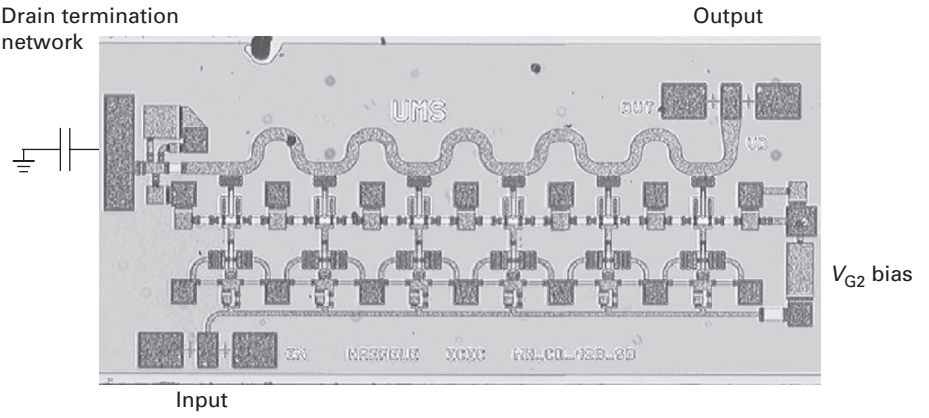

Fig. 5.60 Chip photo of the amplifier shown in Figure 5.59.

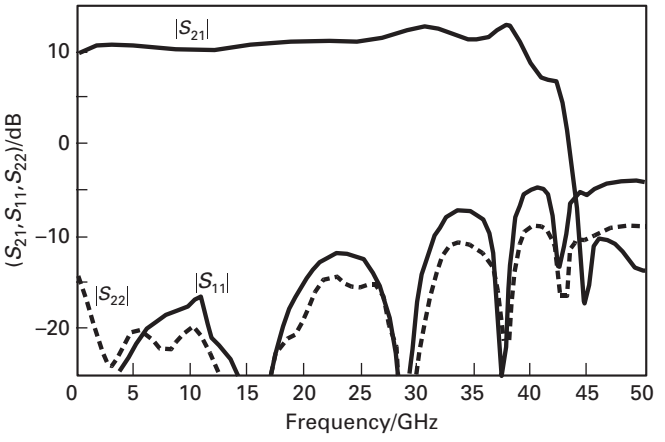

Fig. 5.61 Frequency response of gain ($|S_{21}|$), and input and output reflection coefficients ($|S_{11}|$, $|S_{22}|$) of the distributed amplifier in Figure 5.59.

The design deliberately uses the negative real part of the cascode cell output admittance to compensate for drain–line losses. R_{G2} and R_{S1} improve stability together with the transmission line in the source lead of the cascode, which acts as a small inductor and reduces the cell's gain with increasing frequency, avoiding instability at higher frequencies.

Figure 5.60 shows the chip micrograph of the distributed amplifier. It has six gain stages and is implemented using standard microstrip line technology (the back of the chip is metallised). Two adjacent stages share via the connections to ground – this requires careful assessment of interstage cross-talk issues, but is very efficient in reducing the necessary chip area.

The experimental frequency response (Figure 5.61), shows a very flat gain up to about 40 GHz, where the gain drops sharply. This is a very typical feature of distributed

amplification. Another noteworthy feature is the low reflection coefficient for both input and output over a very wide frequency range, which is due to the distributed nature of the input and output impedances.

The midband gain is 11 dB, the output power at 1 dB gain compression (for a definition, see Figure 5.72 on p. 372) is 22.6 dBm measured at 20 GHz.

A distributed amplifier on Si using Si/SiGe HBTs

The distributed amplifier concept is not restricted to FETs. They can also be realised using bipolar transitors or HBTs. In the example used here, the goal is to realise a distributed amplifier in a production Si/SiGe HBT process on lossy substrates.

The latter issue, the lossy substrate (20 Ωcm specific resistivity), introduces an additional complication because neither standard microstrip transmission lines (which use the substrate as the dielectric) nor coplanar waveguides (which would equally introduce large substrate losses) can be used. Instead, a thin-film microstrip transmission line technique (Figure 5.62) was chosen, which creates the microstrip line entirely above the substrate. Here, the signal line was placed in metal 3, while metal 1 acts as the ground plane, shielding the signal completely from the lossy substrate. The thin dielectric, however, leads to very narrow signal lines for the characteristic impedances in question (50–100 Ω) and strongly increases series resistance losses.

Furthermore, the input admittance of a bipolar transistor is not purely capacitive, as we could safely assume for FETs. Using the hybrid π equivalent circuit of Figure 5.17, we can estimate the admittance $\underline{Y}_{10}$ for a bipolar transistor:

$$\underline{Y}_{10,\text{bipolar}} \approx \frac{I_\text{C}}{\beta_\text{f}\, V_\text{T}} + \jmath\omega \left(C_\text{JBE} + \tau_\text{B} \frac{I_\text{C}}{V_\text{T}} \right), \tag{5.195}$$

where β_f is the small-signal current gain in common-emitter configuration, τ_B is the base transit time, I_C is the collector current in this bias point and $V_\text{T} = kT/q$ is the thermal voltage. The real part of $\underline{Y}_{10}$ would strongly attenuate the signal travelling on the input line and has to be eliminated.

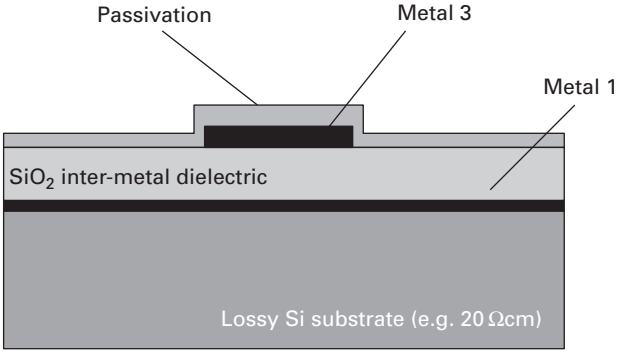

Fig. 5.62 Example of a TFMSL on a silicon substrate.

The latter problem can be solved using a common-collector (emitter follower) input stage (Equation (5.126)):

$$Y_1 \approx \underline{Y}_{12} + \frac{\underline{Y}_{10}}{1 + \frac{g_m}{Y_L}}.$$

It is evident that the input admittance is much smaller. Furthermore, we had seen in Equation (5.128) that given a capacitive component of Y_L, the real part of the input admittance becomes negative. This can be used to compensate for ohmic losses on the input line, but always bears the risk of instability.

If a cascode gain cell is chosen, the negative real part of its output admittance can equally be used to compensate for ohmic losses on the output line, with the same stability caveat.

Figure 5.63 shows an example of a differential amplifier where all of these measures have been taken [33]. It was realised in Si/SiGe HBT technology, with transistors of f_T, $f_{max} = 80\,\text{GHz}$, on a $20\,\Omega\text{cm}$ substrate.

Three cascaded emitter followers are used in the input to achieve the appropriate low input capacitance and negative input conductance. The differential cascode gain cell has open collector outputs which connect directly to the output transmission lines. The capacitively shunted emitter degeneration resistors in the common-source pair improve the bandwidth through a positive gain slope of this stage.

Note the extensive use of level shifting diodes (transistors with their base–collector contacts tied together). This is necessary due to the low collector–emitter breakdown voltages typical of high-f_T Si/SiGe HBTs.

The unusual differential topology solves an additional problem of silicon-based MMICs: The absence of through-the-substrate via holes makes low-inductance grounding highly critical. The differential topology eases packaging by creating an on-chip

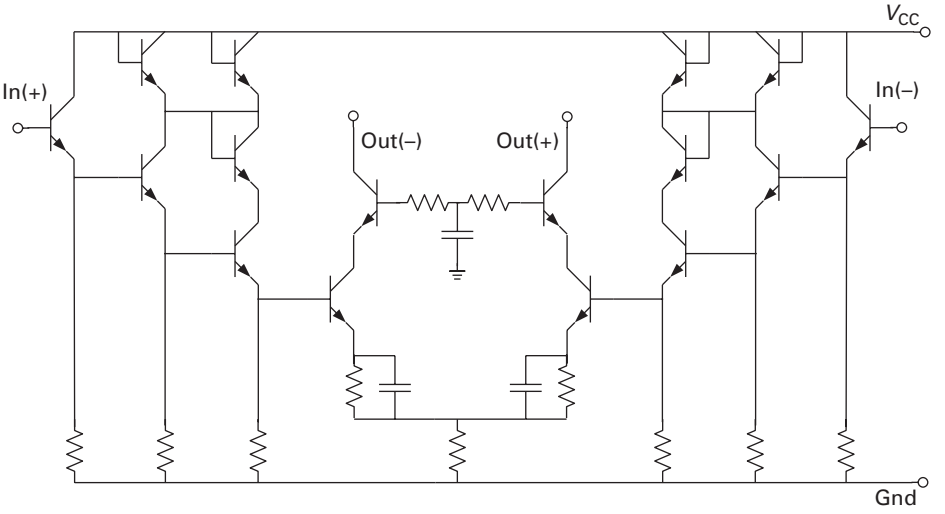

Fig. 5.63 Schematic of a differential distributed amplifier gain cell using Si/SiGe HBTs.

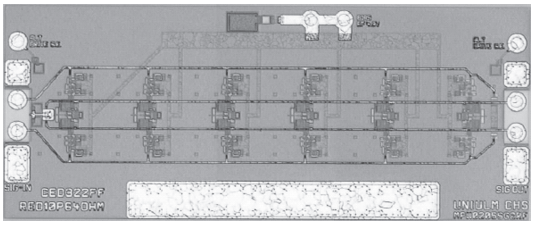

Fig. 5.64 Chip photo of the differential distributed amplifier.

ground, as already discussed. In wideband amplifiers, it is not suitable for all system architectures, however, due to the need for ultra-wideband baluns.

Figure 5.64 shows the chip micrograph of the structure. The chip size is $1.7 \times 0.7\,\text{mm}^2$. The narrow width of the thin film Microstrip line is very apparent. The differential gain is 13.6 dB and the −3 dB bandwidth is 32.2 GHz.

5.4.10 Low-noise amplifier

A very frequent requirement is the design of an amplifier with minimum noise figure – an LNA. This is especially important in weak signal reception environments such as in satellite receivers.

We have seen earlier that the noise figure of any two-port depends on the source reflection coefficient presented to it (see p. 310). The parameters needed for noise-optimum design are

(i) the noise-optimised source reflection coefficient for which the two-port noise figure is minimal: $\Gamma_{S,opt}$;
(ii) the minimum noise figure F_{min} which provides the two-port noise figure under the condition that the source reflection coefficient is the noise-optimised one: $\Gamma_S = \Gamma_{S,opt}$;
(iii) the normalised equivalent noise resistance r_n, which describes the sensitivity of the noise figure F on deviations from the noise-optimised source reflection coefficient $\Gamma_{S,opt}$.

Using these parameters, the noise figure is given by

$$F = F_{min} + \frac{4r_n \left|\Gamma_S - \Gamma_{S,opt}\right|^2}{\left(1 - |\Gamma_S|^2\right)\left|1 + \Gamma_{S,opt}\right|^2}. \tag{5.196}$$

In practical two-ports using active devices, the noise parameters are also bias-dependent. Of particular interest is the dependence of F_{min} on the drain or collector current. Qualitatively, it is shown in Figure 5.65.

An additional aspect needs to be considered – while in principle any reflection coefficient $|\Gamma| \leq 1$ can be transformed into any other using reactive matching networks, practical limitations need to be considered. If the end points of the transformation are located too far apart, the resulting matching network will either be very narrow band (if the reacting matching elements are sufficiently low loss) or introduce significant

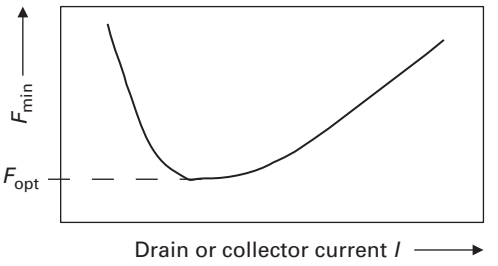

Fig. 5.65 Qualitative dependence of the minimum noise figure on the source or collector current.

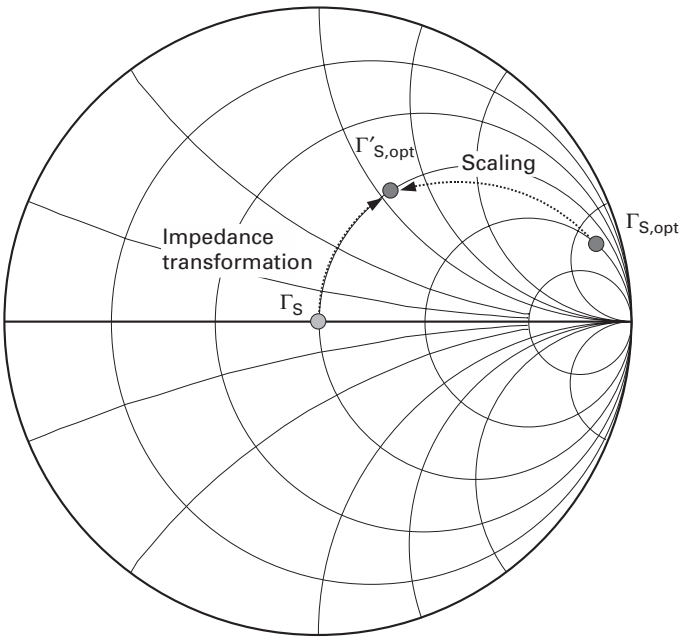

Fig. 5.66 Noise matching example using device scaling and impedance transformation.

additional losses, which deteriorate the noise figure according to Friis' formula. For LNA design, this means that $\Gamma_{S,opt}$ should be suitably located. $\Gamma_{S,opt}$ can be changed by changing the device width ('scaling') – a larger device width results in larger values of $Y_{S,opt}$.

For a better understanding, refer to Figure 5.66. We assume that the original source reflection coefficient Γ_S, e.g. the feed point impedance of an antenna at resonance, is real, and the corresponding impedance is equal to the normalising impedance of the Smith chart, hence $\Gamma_S = 0$. The original noise-optimised reflection coefficient $\Gamma_{S,opt}$ is located too far towards the outside of the Smith chart. By choosing a larger device, $\Gamma_{S,opt}$ is achieved in a location which is much closer to Γ_S. In fact, this location is ideal because the transformation from Γ_S to $\Gamma'_{S,opt}$ can be achieved conveniently using only a series inductance.

The fundamental design steps of the LNA's input stage are hence the following:

(i) Pick a suitable device size which puts $\Gamma_{S,opt}$ into a convenient location with respect to the original source reflection coefficient Γ_S.
(ii) Adjust the bias point such that the optimum F_{min} is achieved.
(iii) Design the input matching network.

Because the bias point affects $\Gamma_{S,opt}$, a few iterations may be necessary.

In principle, matching for optimum noise performance ($\Gamma_S = \Gamma_{S,opt}$) and matching for optimum power transfer at the input ($\Gamma_S = \Gamma_{in}^{\star}$) are unrelated. A frequent requirement, however, is the combination of optimum noise performance and a minimum return loss, hence $\Gamma_{in}^{\star} \approx \Gamma_{S,opt}$. This cannot be achieved using impedance transformation networks between the source and the LNA input, because that would modify $\Gamma_{S,opt}$ for the resulting two-port and Γ_{in} in the same way. Instead, Γ_{in} can be modified in two ways which leave $\Gamma_{S,opt}$ invariant:

(i) through lossless feedback;
(ii) by mismatching the output for non-unilateral two-ports, utilising the fact that the input reflection coefficient depends also on the load reflection coefficient:

$$\Gamma_{in} = S_{11} \frac{S_{21} S_{12} \Gamma_L}{1 - S_{22} \Gamma_L}.$$

Figure 5.67 summarises the individual reactive networks surrounding the LNA core, which can be used in the design to fulfil noise and return loss specifications.

The feedback elements Z_A (series or current–voltage feedback) and Z_B (parallel or voltage–current feedback) are used to set Γ_{in} while leaving $\Gamma_{S,opt}$ invariant, as discussed. $M1$ provides noise match or, after suitable modification of Γ_{in} using feedback techniques, simultaneous noise and power match (minimum noise figure and minimum input return loss). $M2$ can be chosen either to present the needed Γ_L to the LNA core for adjustment of Γ_{in} (see above), or to achieve power match at the output (minimum output return loss).

Very commonly, Z_A is an inductor. As already shown in Equation (5.141) this provides an increased real part of the input impedance. Consider the case depicted in Figure 5.68.

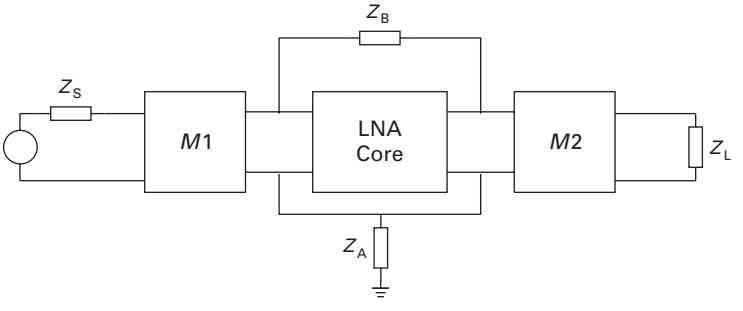

Fig. 5.67 Matching and feedback networks in LNA design.

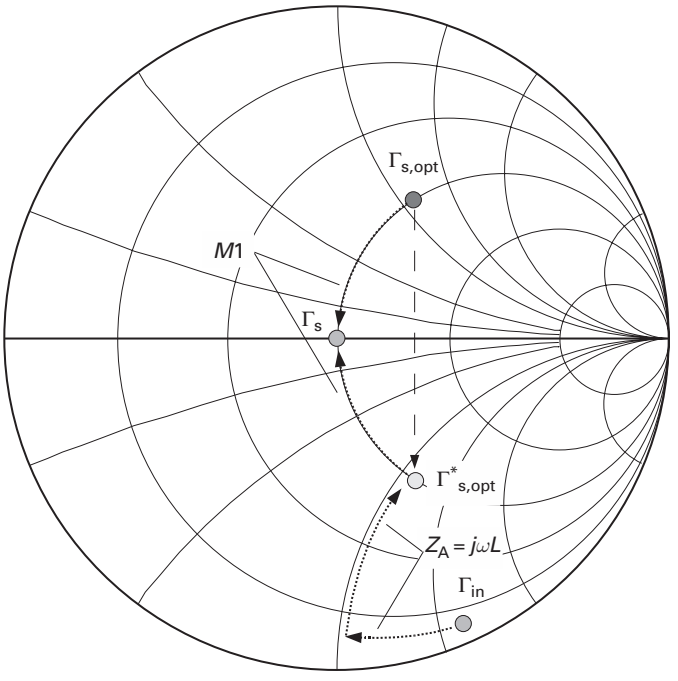

Fig. 5.68 Simultaneous noise and power match example using inductive series feedback and an input matching network $M1$.

Without any feedback or matching network, the input reflection coefficient is Γ_{in}, corresponding to an input impedance of $R_{in} - \jmath X_{in}$. The location indicated in the example would be typical for a FET. The goal is now to transform Γ_{in} to a new location $\Gamma'_{in} \approx \Gamma^{\star}_{S,opt}$. We connect an inductor L in series to the LNA core. Applying Equation (5.141), we find for the input impedance of the LNA core with feedback:

$$Z'_{in} = R_{in} + \omega_T L + \jmath\,(\omega_0 L - X_{in})\,. \tag{5.197}$$

On the Smith chart, the transformation path corresponding to the effect of L can be interpreted as first increasing the imaginary part, starting from Γ_{in} and then increasing the real part, as shown in the lower part of Figure 5.68.[5]

In a second step, matching network $M1$ (which in the example is simply a series inductor) transforms both Γ'_{in} and $\Gamma_{S,opt}$ towards Γ_S, achieving the required simultaneous optimisation of noise and input return loss.

For the LNA core, cascode stages (see p. 333) are very frequently being used at microwave frequencies. This is because the aforementioned scaling, placing $\Gamma_{S,opt}$ in an 'easily matchable' location results frequently in relatively large transistors, where the Miller effect (discussed on p. 318) can be significant – adoption of a cascode topology is a proven way to reduce the increased input capacitance associated with the Miller effect.

[5] As an aside, you may notice that with increasing L, F_{min} decreases – the associated gain G_{ass}, however, also decreases. The entity invariant to reactive feedback is the *noise measure* $M = F_{min}/\left(1 - G_{ass}^{-1}\right)$.

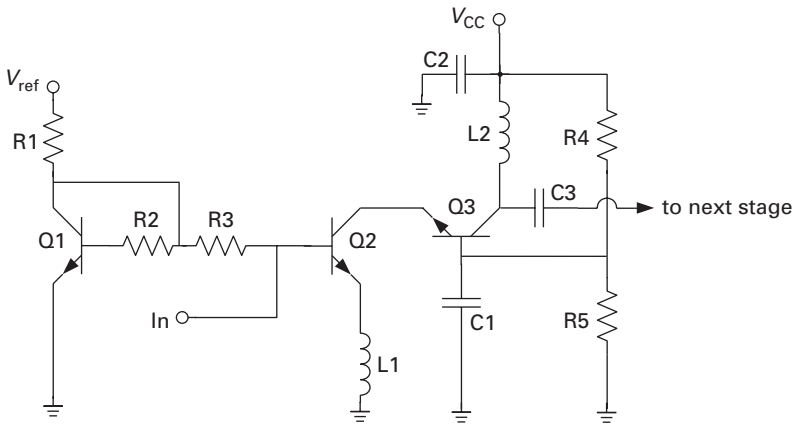

Fig. 5.69 First stage of a three-stage LNA for 24 GHz using Si/SiGe HBTs.

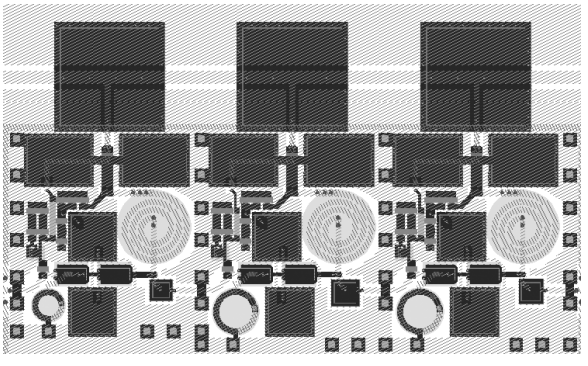

Fig. 5.70 Layout of the three-stage 24 GHz LNA.

As a practical example, we will discuss a three-stage LNA for 24 GHz using Si/SiGe HBTs [35]. The schematic of the first stage is shown in Figure 5.69.

Transistors Q2 and Q3 form the cascode LNA core; Q1 forms a current mirror with Q2 to set the latter's collector current. Q3's base voltage is then set using the voltage divider R4/R5. All capacitors are large-value bypass capacitances.

Inductor L1 is used to allow simultaneous noise and power match along with the proper sizing of Q2. There is no on-chip inductance in series with the *In* port, because the bond wire is used instead, efficiently including this parasitic into the design. L2 forms, together with the capacitance between the collector Q3 and ground and the input capacitance of the following stage, a parallel resonance which provides the LNA with a bandpass characteristic.

The other two stages are identical in topology, but due to the different source impedances of the preceding stages, the inductive source degeneration of the common-emitter transistor (Q2) is adjusted.

Figure 5.70 shows the layout of the three-stage amplifier. Note that the source degeneration inductors are constructed as two-layer stacked inductors, while the tank circuit inductors (L2) are conventional spiral inductors.

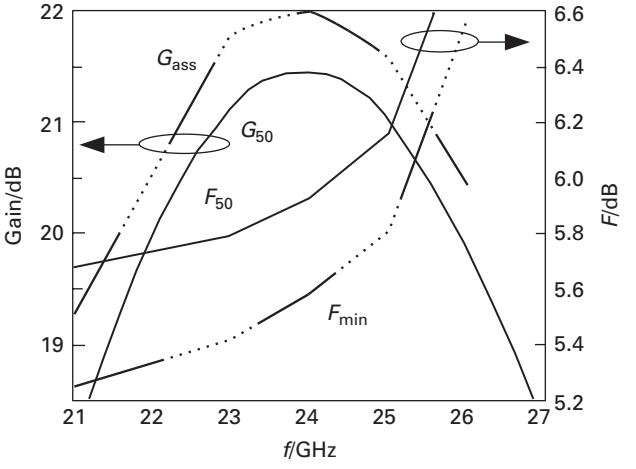

Fig. 5.71 Gain and noise performance of the 24 GHz LNA using Si/SiGe HBT technology.

The circuit was realised in a Si/SiGe HBT technology with f_T, $f_{max} = 80$ GHz and characterised on wafer. Results of a small-signal characterisation are shown in Figure 5.71. The circuit shows the targeted bandpass performance with the gain peak at 24 GHz (the intended application is in the 24 GHz ISM band). The minimum noise figure at 24 GHz is 5.6 dB, while the 50 Ω noise figure is slightly below 6 dB. This deviation is not surprising, as the circuit was designed to provide optimum noise figure with the bond wire parasitic included. The gain with a 50 Ω source impedance is $G_{50} = 21.4$ dB, while the associated gain under noise match conditions is $G_{ass} = 22$ dB – in this circuit, the noise-optimised source impedance is actually slightly closer to 50 Ω than the input impedance.

5.4.11 Amplifier linearity

So far, we treated amplifiers as perfectly linear systems – the output signal can always be described as a linear combination of the input signals. In reality, however, any circuit including active devices will show a non-linear behaviour and the assumption of linearity holds only for small deviations around a given operating point.

In practice, the non-linear behaviour of amplifiers will generate *nonlinear distortions*, which create non-linear deviations in time-domain signal shape, and additional spectral components in the frequency domain which have to be reckoned with.

Single-tone excitation

A common way to treat general non-linear functions is the Taylor series expansion. A non-linear function $f(x)$ is expanded around $x = x_0$ as

$$f(x) = \sum_{\nu=0}^{\infty} \frac{f^{(\nu)}(x_0)}{\nu!}(x - x_0)^{\nu}, \tag{5.198}$$

where $f^{(\nu)}(x_0)$ is the ν^{th} derivative of f with respect to x in $x = x_0$.

Now assume that we apply a single sinusoidal signal to our non-linear system: $a(\omega t) = A_0 + A\sin(\omega t)$.

The output signal $f(\omega t)$ can now be described by the following Taylor series expansion ($x_0 = 0$):

$$\begin{aligned} f(\omega t) = {} & k_0 A_0 \\ & + k_1 \sin(\omega t) \\ & + k_2 \sin^2(\omega t) \\ & + k_3 \sin^3(\omega t) \\ & + \ldots . \end{aligned} \tag{5.199}$$

The first two lines in Equation (5.199) provide the linear response, while the following terms are non-linear distortions. Consider that

$$\sin^2(\omega t) = \frac{1}{2}\left[1 - \cos 2\,\omega t\right]$$

$$\sin^3(\omega t) = \frac{1}{4}\left[3\sin(\omega t) - \sin(3\omega t)\right],$$

and we find that Equation (5.199) turns into

$$\begin{aligned} f(\omega t) = {} & k_0 A_0 + \frac{k_2}{2} \\ & + \left(k_1 + \frac{3k_3}{4}\right)\sin(\omega t) \\ & - \frac{k_2}{2}\cos(2\omega t) \\ & - \frac{k_3}{4}\sin(3\omega t) \\ & + \ldots . \end{aligned} \tag{5.200}$$

We easily see that the non-linear distortion results in new spectral components (harmonics) being generated, which are related to the fundamental components as integer multiples.

A simple procedure to assess an amplifier's linearity is the single-tone excitation test. A test generator with high spectral purity and adjustable power is connected to the input of the amplifier, and a spectrum analyser to the output. Increasing the input power (P_{in}), the power of individual spectral components at the output is recorded. Plotting the output power levels as a function of the input power on a double-logarithmic scale, we obtain a graph similar to the one shown in Figure 5.72.

For low power levels, the output power of the fundamental spectral line will increase linearly with the input power. Gradually, it will, however, rise more slowly – *gain saturation* sets in. When the power ratio between the extrapolated linear increase and the actual curve is 1 dB, the 1 dB compression point ($P_{-1\text{dB}}$) has been reached. It is a measure of the maximum power the amplifier can deliver in linear operation. Depending on the application, it is referred to the input (e.g. LNAs) or the output (e.g. power amplifiers).

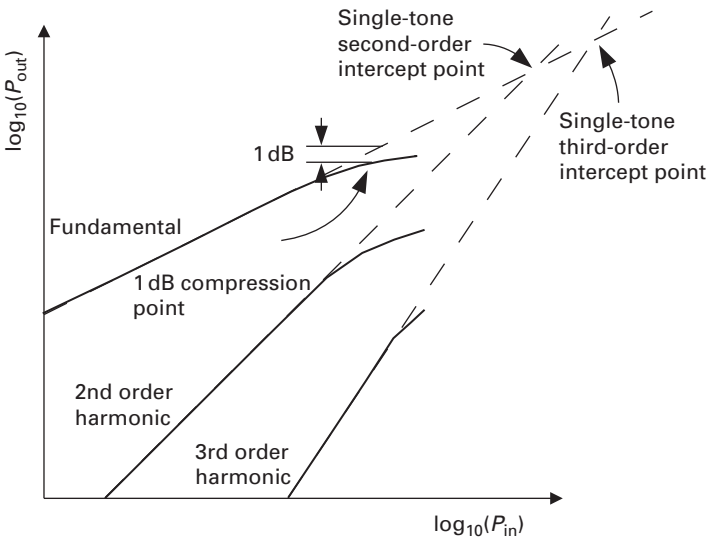

Fig. 5.72 Single-tone excitation test of an amplifier, showing definitions for the 1 dB compression, second-order intercept and third-order intercept points.

The spectral power at the second harmonic increases twice as fast as the fundamental power, before it also shows saturation. Extrapolating the curve at low input powers, we find the *single-tone second-order intercept point* at the point where the extrapolation intersects the extrapolated fundamental power.

The spectral power at the third harmonic increases three times as fast as the fundamental power. Its extrapolation intersects the extrapolated fundamental power at the *single-tone third-order intercept point*.

The intercept points can also be referred to the input or the output, depending on the application.

In many applications where the operational bandwidth is only a small fraction of the carrier frequency, the generation of harmonics is not necessarily a problem, because they can easily be removed by filtering. For example, frequency modulated (FM) transmitters are operated under strongly non-linear conditions (class C, see p. 377), and the resulting harmonics in the output signal are simply removed by low-pass filtering.

Two-tone excitation

An FM signal is a particularly simple example of modulation, because the resulting signal has only a single spectral component (which varies in frequency, but that is irrelevant here). Most modulated signals, however, consist of many spectral components which are present at the amplifier input simultaneously.

To understand what amplifier non-linearity will do to these signals, let us construct a simple experiment, where the input signal is formed by two spectral components (at ω_1 and ω_2) of equal amplitude, applied to the input of the amplifier. The output is connected to a spectrum analyser again. The corresponding block diagram is shown in Figure 5.73.

Again, the test generators need to have very high spectral purity.

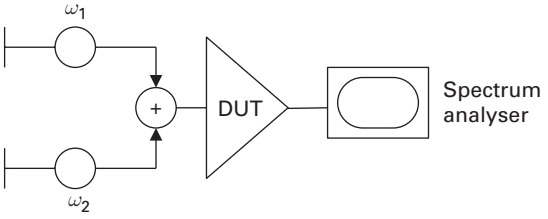

Fig. 5.73 Schematic representation of a two-tone excitation test of an amplifier (DUT = device under test).

Mathematically, the description of the distorted output signal becomes much more complex. We obtain:

- Fundamental components at ω_1 and ω_2.
- Harmonics of the input signals (at $2\omega_{1,2}$, $3\omega_{1,2}$, ...).
- Components due to the product of the two input signals – consider that

$$\sin(\omega_1 t)\sin(\omega_2 t) = \frac{1}{2}\{\cos[(\omega_1 - \omega_2)t] - \cos[(\omega_1 + \omega_2)t]\}. \tag{5.201}$$

  The multiplication term therefore produces spectral components at the sum and difference of the two input spectral lines. These components are called *two-tone second-order intermodulation products*.
- Components due to the product of a fundamental component and a second-order harmonic

$$\sin^2(\omega_1 t)\sin(\omega_2 t) = \frac{1}{2}[\sin(\omega_2 t) - \sin(2\omega_1 t + \omega_2 t) + \sin(2\omega_1 t - \omega_2 t)] \tag{5.202}$$

$$\sin^2(\omega_2 t)\sin(\omega_1 t) = \frac{1}{2}[\sin(\omega_1 t) - \sin(2\omega_2 t + \omega_1 t) + \sin(2\omega_2 t - \omega_1 t)]. \tag{5.203}$$

  These terms hence generate spectral components at $2\omega_{1,2} \pm \omega_{2,1}$, which are called *two-tone third-order intermodulation products.*
- Higher-order components which are neglected here.

Figure 5.74 schematically shows the spectral components generated by non-linear distortion of a two-tone signal, up to the third order. Note that second- and third-order harmonics as well as the second-order intermodulation products are significantly far away from the original signal and can most often be removed by filtering. Of particular concern are two third-order intermodulation products at $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, because they are close to the original spectral components and cannot be removed by filtering.

Just as in case of the single-tone excitation, we can plot the output powers at the fundamental tones and the close-in third-order intermodulation components as a function of the input power. Figure 5.75 shows an example of such a measurement. The two-tone third-order intercept point is found by extrapolating the low-power portions of the curves, where the relationship between input and output powers has a linear shape on a double-log plot. It can be referred to the input or the output.

Often the system requirement will be formulated in terms of the *intermodulation distance (IMD)* or the *dynamic range*, not in terms of the intercept points.

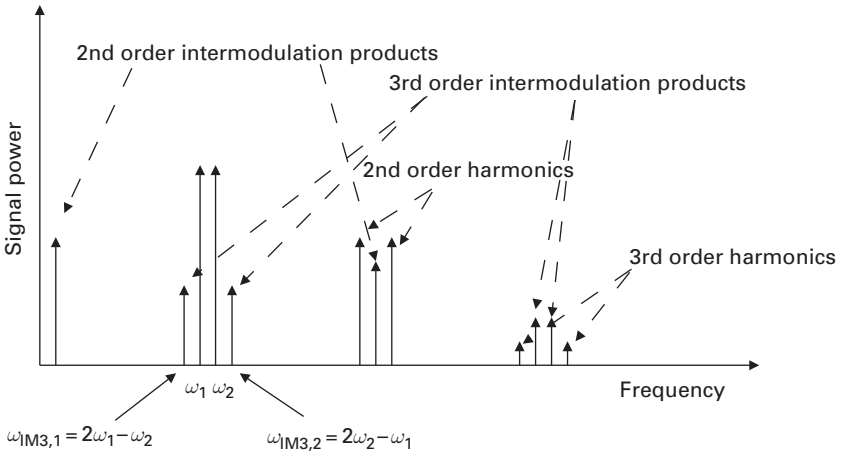

Fig. 5.74 Schematic representation of spectral components generated from a two-tone excitation through second- and third-order non-linearities.

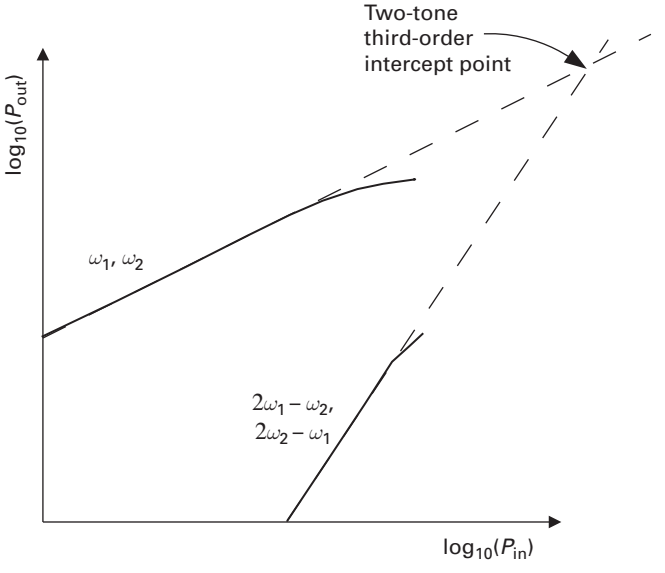

Fig. 5.75 Determination of the two-tone third-order intercept point.

The IMD, measured in a two-tone excitation test, is simply the power ratio between the power level of the two carriers at the amplifier output and the highest intermodulation spectral lines. The most prominent ones will typically be third-order intermodulation products. Then, the IMD can be calculated from the third-order intercept point. Consider again Figure 5.75 and remember that on the double-log scale, the $P_{out} = f(P_{in})$ transfer curve for the fundamental component has a slope of 1, while it is 3 for the third-order intermodulation products. The distance between the curves for the fundamental and the intermodulation products is the IMD on the log–log scale. The

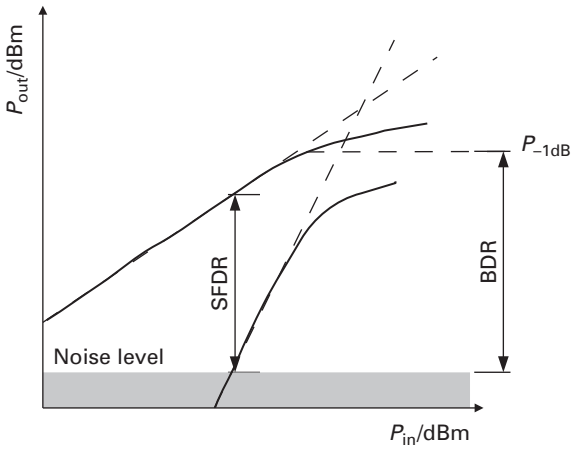

Fig. 5.76 Determination of the SFDR and the BDR in a two-tone excitation measurement.

powers are expressed most often in dBm[6] on a logarithmic scale. Therefore,

$$\frac{\text{IMD}}{\text{dB}} = 2\left(\frac{\text{IIP3}}{\text{dBm}} - \frac{P_{\text{in}}}{\text{dBm}}\right), \tag{5.204}$$

where IIP3 is the third-order intercept point referred to the input.

On a linear scale (powers in W), the IMD can be expressed as

$$\text{IMD} = \left(\frac{\text{IIP3}}{P_{\text{in}}}\right)^2. \tag{5.205}$$

Specification of an amplifier in terms of dynamic range combines linearity and noise. There are two definitions, which are compared in Figure 5.76.

The *spurious-free dynamic range (SFDR)* is the IMD at the point where the power of the third-order intermodulation products is equal to the noise floor. The *blocking dynamic range (BDR)* is the distance between the 1 dB compression point $P_{-1\text{dB}}$ and the noise floor.

Adjacent channel power ratio

Modern communication systems have frequently very complex modulation schemes, with many spectral components present. They are, therefore, very sensitive to intermodulation effects in non-linear amplifiers. A two-tone measurement can only give an indication of linearity, but is no solid proof of the amplifier's suitability.

A very realistic test is the ACPR (adjacent channel power ratio) test, which is always specific to a certain modulation technique. Figure 5.77 shows an example for a UMTS signal. In a first step, the integral powers within the channel bandwidth need to be calculated from the spectral analysis. Then, the ACPR is calculated as the ratio of the power in the band of operation to either the lower or higher adjacent channel. It is a

[6] dBm means decibels relative to 1 mW, i.e. 0 dBm = 1 mW, 20 dBm = 100 mW, etc.

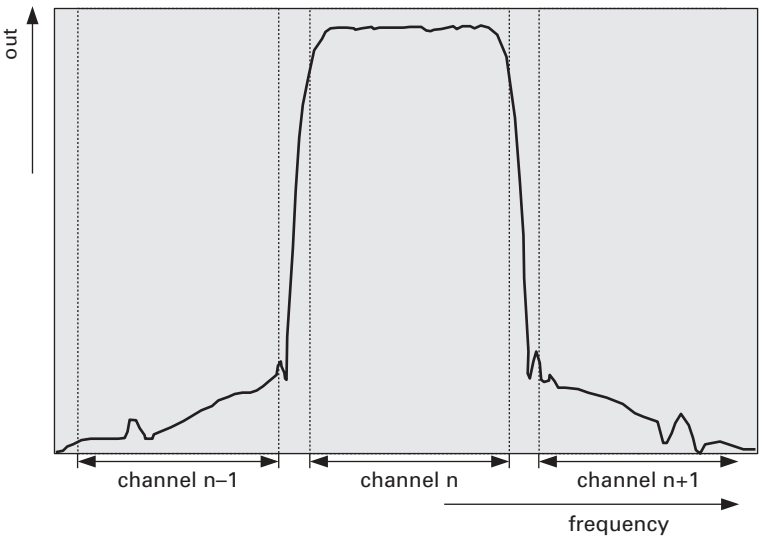

Fig. 5.77 Example of an Adjacent Channel Power ratio (ACPR) measurement: Power spectral density (PSD) versus frequency, with channel limits indicated.

direct measure of the interference generated by transmitter non-linearities in adjacent channels.

5.4.12 Power amplifiers

Power amplifiers have the task of amplifying signals before they are delivered to loads, such as antennas or cables. Critical criteria are

- maximum output power, for example measured in terms of output power at the 1 dB compression point $P_{-1\,\mathrm{dB}}$ (see Figure 5.72);
- gain (either small-signal gain or large-signal gain at a given output power);
- gain and potentially phase deviation across the operational bandwidth;
- linearity, defined by parameters such as the output-referred two-tone third-order intercept point, the IMD at a given output power, or the ACPR at a given output power;
- efficiency – at microwave frequencies, it is customary to use the *power added efficiency (PAE)*, the ratio of the power difference between output and input to the DC power:

$$PAE = \frac{P_{\mathrm{out}} - P_{\mathrm{in}}}{P_{\mathrm{DC}}} = \eta \left(1 - \frac{1}{G}\right), \tag{5.206}$$

where η is the collector or drain efficiency ($\eta = P_{\mathrm{out}}/P_{\mathrm{DC}}$) and G the amplifier gain.

Classes of operation

Since the days of vacuum tubes, amplifier operation has been described by *classes*, which describe where the amplifying devices are biased in a quiescent state.

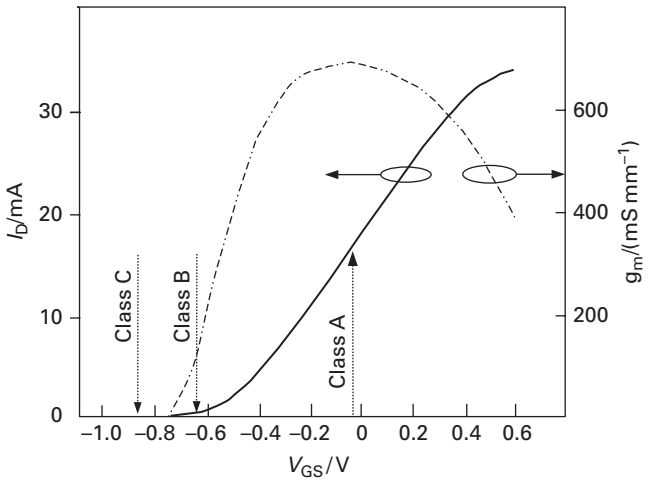

Fig. 5.78 HEMT drain current I_D and transconductance g_m as a function of the gate-source voltage V_{GS} with bias points for power amplifier classes A, B and C indicated.

For an understanding of the 'classical' classes A, B and C, refer to Figure 5.78. The example shows the drain current and transconductance of a HEMT. For the classification, we observe the drain current curve.

In a class A amplifier, the gate-source voltage V_{GS} is set in the region where the output current I_D is a linear function of the input voltage V_{GS} – the transconductance is approximately constant. For both positive and negative half-waves of the input signal, current will flow – the *conduction angle* is 360°. In this bias point, the amplifier will exhibit a very high linearity, but low efficiency. The theoretical maximum is 50%, but at microwave frequencies, values of 30% would already be very satisfactory.

For class B, the device is biased at pinch-off. Only the positive half-wave of the input signal will then generate an output current flow – the conduction angle is 180°. The efficiency will increase theoretically to 78.5% ($\pi/4$), and at microwave frequencies it can still reach 60% or higher, but the deviation from a sine wave in the output current creates non-linear distortions.

A class C amplifier has a quiescent bias point where V_{GS} is significantly below the threshold voltage V_{th}. Output current will flow only if the momentary $V_{GS}(t) > V_{th}$, therefore the conduction angle is $<180°$. The efficiency can still be higher; however, due to the lower conduction angle, the non-linear distortions are also increased.

Switched amplifiers

There is another interpretation of class C operation, which is helpful in understanding the way that amplifiers in class D, E and F operate. For this, look at Figure 5.79. It shows the output I–V characteristics of a HEMT (but this could be any FET). Provided that the driving voltage is large enough, the transistor simply changes between two saturated states with very different differential resistances. In the quiescent point, the transistor is in cut-off and the differential resistance between drain and source is very

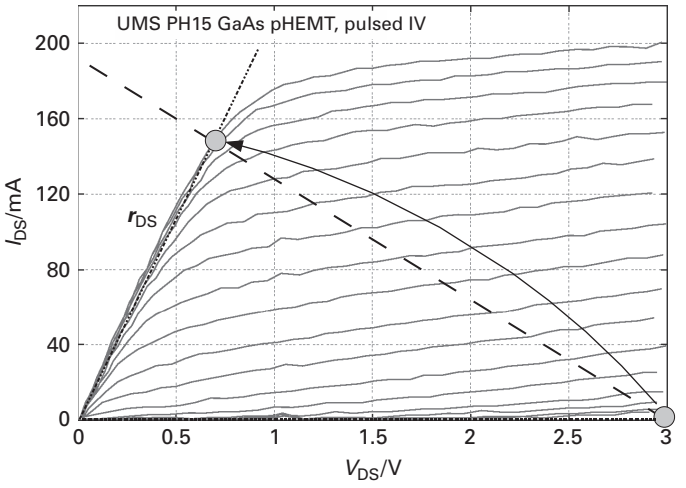

Fig. 5.79 Saturated class C operation in the output I–V characteristics of a HEMT.

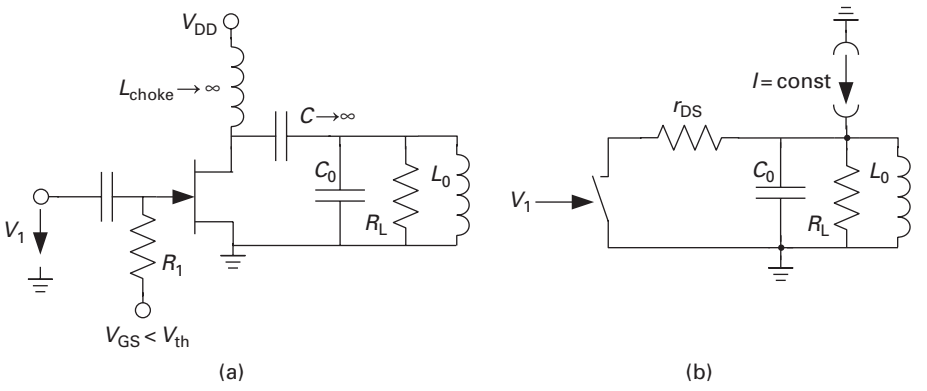

Fig. 5.80 (a) Simple class C amplifier topology and (b) its equivalent circuit.

high. For a sufficiently high input voltage, the transistor reaches another saturated state with small r_{DS}.

We can, therefore, model the transistor in saturated class C operation simply by a switch in series with its residual differential resistance r_{DS}. Figure 5.80(a) shows a simple class C amplifier stage. The load is embedded in an LC resonant circuit which acts as a bandpass filter to suppress the harmonic frequency components other than the fundamental. The RF choke (L_{choke}) provides a constant current, at least on the time scale of interest. This circuit can also be realised with an LC parallel resonant circuit, by the way.

Replacing the choke with a constant current source, and the FET with a periodically actuated switch and its series resistance r_{DS}, we arrive at the equivalent circuit in Figure 5.80(b). The class C amplifier operates in this configuration by periodically shunting current away from the load.

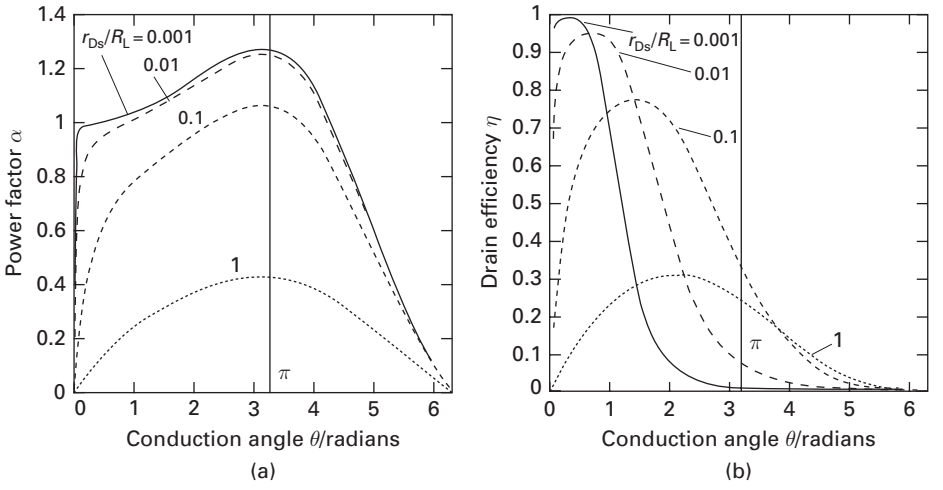

Fig. 5.81 Class C amplifier: (a) power factor α and (b) drain efficiency β as a function of frequency.

A detailed analysis of class C operation is found in [17]. First, note that due to the RF choke, the average voltage across the load is the supply voltage V_{DD}. The peak voltage is $(1+\alpha)V_{DD}$ and the minimum voltage $(1-\alpha)V_{DD}$, where

$$\alpha(\theta) = \frac{4\sin\left(\frac{\theta}{2}\right)}{\theta + \sin(\theta) + \frac{2\pi r_{DS}}{R_L}}. \tag{5.207}$$

Here, θ is the conduction angle. Note that for $r_{DS} \to 0$, $\alpha_{max} = \alpha(\theta = \pi) = 1.27$ – the maximum voltage across the transistor can, therefore, exceed the supply voltage by a factor of 2.27.

The output power in saturated class C operation is

$$P_{out} = \frac{(\alpha V_{DD})^2}{2R_L}. \tag{5.208}$$

The drain efficiency is

$$\eta(\theta) = \pi \frac{r_{DS}}{R_L} \frac{\alpha^2}{\theta - 2\alpha \sin\left(\frac{\theta}{2}\right)}. \tag{5.209}$$

Both the power factor α and the drain efficiency η are shown in Figure 5.81 as a function of the conduction angle θ. Note that the output power always peaks at $\theta = \pi$, but that the efficiency has its maximum at much lower conduction angle. The normalised on-resistance of the FET, r_{DS}/R_L, has a significant influence on both the output power and the drain efficiency.

Class D amplifier

Above, we interpreted the class C amplifier as a resonant circuit driven by current pulses, where for maximum efficiency the current flow angle was $\theta < \pi$. We can, of course, not only turn the current on and off, but actually reverse it, as shown schematically in Figure 5.82.

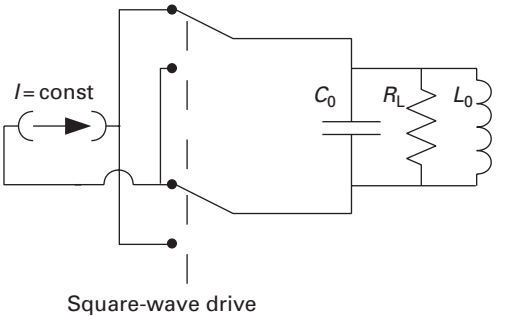

Fig. 5.82 Class D amplifier principle.

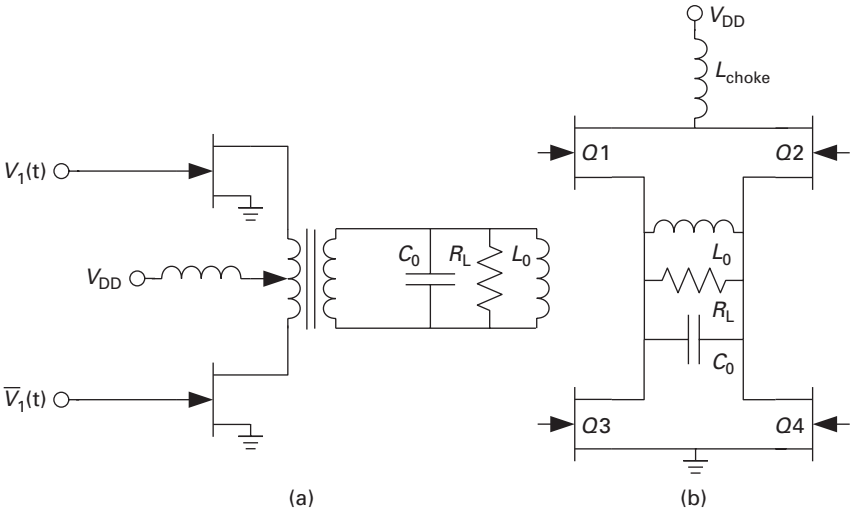

Fig. 5.83 Example implementations of class D amplifiers. After [17].

Instead of a single-pole, single-throw switch, the equivalent circuit shows a double-pole, double-throw switch which periodically reverses the current through the load. The parallel resonant circuit again eliminates all harmonic frequency components except the fundamental one.

In practice, the switches are realised with transistors, of course. Figure 5.83 shows two examples. In Figure 5.83(a), the current reversal is achieved using a transformer where the current is fed into the centre tap, and the ends of the primary coil are connected alternatingly to ground. This is a very common solution at lower frequencies.

At microwave frequencies, the transformers are difficult to realise, and in any case they do not lend themselves easily to monolithic integration. The circuit in Figure 5.83(b) is then more practical – it avoids transformers altogether; however, now we have four transistors to control instead of two: in this bridge configuration, transistors Q1 and Q4, and Q2 and Q3 conduct alternately to achieve the current phase reversal across the load.

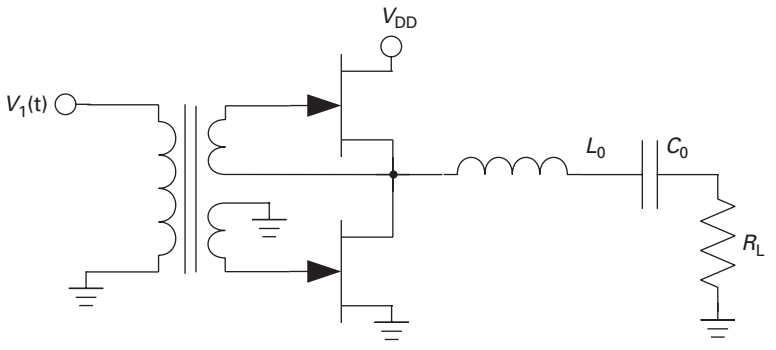

Fig. 5.84 Class D amplifier example using a series-fed load. After [17].

Note that in both implementations, the load floats – it has no direct ground reference, which is often a problem for microwave systems where ground-referenced (single-ended) transmission is more common. This can be avoided in a class D amplifier if a series-fed load is applied. This is shown in Figure 5.84. The bandpass function is now realised with a series resonant circuit (L_0, C_0), and the voltage is alternated, not the current. Still a balun is needed at the input, unless $V_1(t)$ is already available in differential form.

The class D amplifier output power is in saturation [17]:

$$P_{\text{out}} = \frac{2V_{\text{DD}}^2}{\pi^2 R_{\text{L}}}, \tag{5.210}$$

while the drain efficiency is

$$\eta = \frac{R_{\text{L}}}{R_{\text{L}} + r_{\text{DS}}}, \tag{5.211}$$

at least for the circuits according to Figures 5.83(a) and 5.84. r_{DS} is again the channel resistance of the FET for low V_{DS} ('on-resistance'). For the circuit in Figure 5.83(b), the efficiency is lower because the switch resistance doubles.

Class D amplifiers place high demands on the ideality of the switches and on the timing. This is particularly true for circuits such as in Figure 5.83(b) or 5.84, where switches are stacked – they must never conduct at the same time, not even for brief periods. Therefore, class D amplifiers are mostly restricted to lower RF frequencies.

Class E and F amplifiers

Class E and F amplifiers are derived from class C. The idea in a class E amplifier [38] is to make sure that the drain-source voltage of the switching transistor (see Figure 5.80) is zero when the transistor switches, leading to a reduction of losses due to capacitive charging. This can be achieved by a modified output network. Consider Figure 5.85. At first glance, it looks like a class C amplifier with a series resonant feed, with the addition of shunt capacitor C_2. Additionally, the series resonant filter is tuned to a frequency below the intended frequency of operation ω_0 by increasing L_1. Adjusting L_1 and C_2, the voltage-free switching condition is found and the efficiency is maximised.

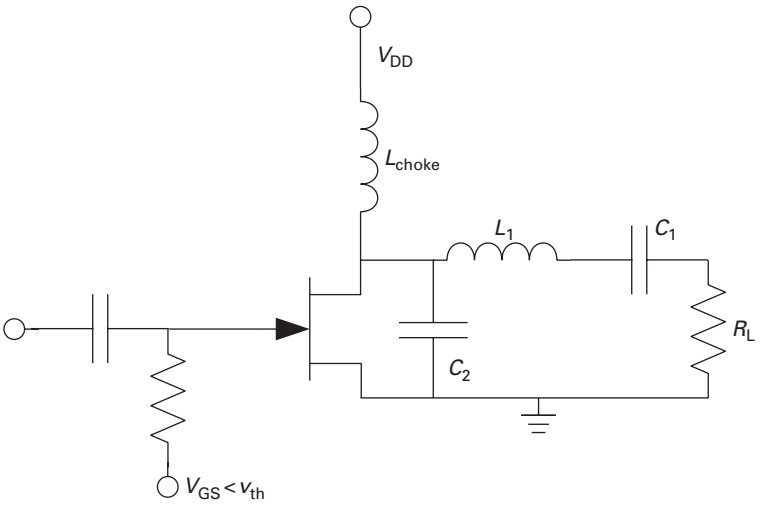

Fig. 5.85 Simplified class E amplifier schematic.

An in-depth analysis of class E operation can be found, e.g. in [8]. The inductance L_1 is chosen as

$$L_1 = \frac{1 + 1.153\omega_0 C_1 R_L}{\omega_0^2 C_1}. \tag{5.212}$$

The capacitance C_2 is

$$C_2 = \frac{2}{3.467\omega_0 \pi R_L}. \tag{5.213}$$

A problem may be that the maximum drain-source voltage is even higher than in the class C amplifier, $V_{DS,max} \approx 3.56 V_{DD}$.

The class F amplifier increases the efficiency by appropriately terminating the harmonics. The idea is to achieve square-wave voltage excitation with respect to the drain-source voltage, as in case of the series-fed class D amplifier (Figure 5.84).

A well-known fact from Fourier analysis is that a square wave (rectangular signal with 50% duty cycle) in the time domain produces only odd harmonics in the frequency domain. We must, therefore, make sure that all odd-numbered harmonics ($n = 1, 3, 5, \ldots$) are still present in the drain-source voltage. The load should, therefore, present an open to the transistors at these frequencies.

A way to achieve this is to use the transforming properties of quarter-wave transmission lines, which we discussed much earlier; see Equation (5.171) on p. 349. Assume that a transmission line, which is a quarter wavelength long at the fundamental frequency, is terminated by a short at all harmonic frequencies except the fundamental. Then an open will appear at the input for all odd-numbered harmonics, while a short results for all even-numbered harmonics, where the electric length of the transmission line is a multiple of $\lambda/2$. The modification of the original class C topology (Figure 5.80) is quite straightforward, as Figure 5.86(a) shows. λ is the wavelength at the frequency of operation ω_0. The resonant circuit formed by L_0, C_0 is resonant at ω_0; C_0 then acts

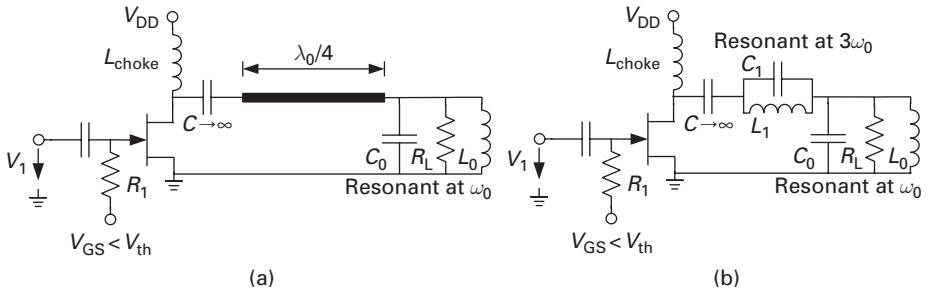

Fig. 5.86 Topology of a class F amplifier: (a) using a quarter-wave transformer and (b) using a parallel resonant trap tuned to the third harmonic.

as a short at the higher-order harmonics. This short is transformed into an open by the quarter-wave transformer at all odd-numbered harmonics.

In monolithic integration, the transmission line transformer is frequently much too long, and it may introduce significant losses. In many cases, it is perfectly acceptable to just use the third harmonic. This is shown in Figure 5.86(b). Here, a simple parallel LC circuit blocks the third harmonic ($3\omega_0$), while it acts as a short for all other harmonics and the fundamental frequency.

5.5 Oscillators

Oscillators are crucial components in almost any microwave system. Their fundamental task is to generate AC energy at a well-defined frequency from DC sources. A typical use of an oscillator is shown in the generic receiver block diagram of Figure 5.87, where it converts the input signal to a lower intermediate frequency. The mixer circuit, which is also needed for the frequency translation, will be discussed in the next section.

The class of oscillators discussed here has three aspects in common:

(i) a *resonator* to set the frequency of oscillation,
(ii) the generation of *instability* to allow the onset of oscillation, and
(iii) *amplitude control* to establish a stable amplitude in steady state.

Simple relaxation-type oscillators, such as found in simple digital timing circuits, will not be covered here.

5.5.1 Resonators – a brief overview

The resonator has the task of setting the oscillator's natural frequency.

The most common resonator types are

- lumped-element *LC resonators*, which come in either series or parallel resonance forms;
- *transmission-line resonators*;

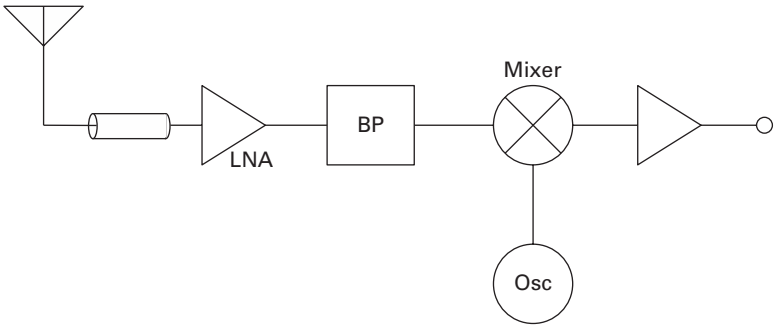

Fig. 5.87 Generic receiver block diagram.

- *cavity resonators* using waveguide elements;
- *dielectric resonators*, which use high ϵ_r ceramics and are typically combined with transmission line coupling structures;
- *quartz crystals* and similar devices which use the piezoelectric effect to derive electrical from a mechanical resonance – surface acoustic wave (SAW) and bulk acoustic wave (BAW) resonators also fall into this category.

Other resonator types, such as the magnetically tuned YIG (yttrium iron garnet) resonators, have only very limited use in speciality applications.

Critical aspects for resonators are

- the *quality factor*, which will be discussed in more detail below;
- the *reproducibility* of the resonant frequency – this can be a significant problem in BAW and SAW resonators;
- the *stability* of the resonant frequency against changes in temperature, mechanical shock and aging;
- the *tunability* of the resonant frequency – mostly established using variable capacitance elements.

For fixed-frequency oscillators, quartz crystals can still be considered to be the gold standard. Replacement of quartz resonators by elements which can be monolithically integrated is highly desirable and a hot research topic.

Quality factor

A very generic definition of the quality factor compares the stored and the dissipated energy in the resonator [20]:

$$Q = 2\pi \frac{\text{stored energy in the resonator}}{\text{dissipated energy during one cycle}}, \tag{5.214}$$

for $\omega = \omega_0$.

Let us consider RLC resonators (Figure 5.88) as an important example – via equivalent circuits, other resonator types can be transferred into RLC type resonators, at least in the immediate vicinity of the resonant frequency.

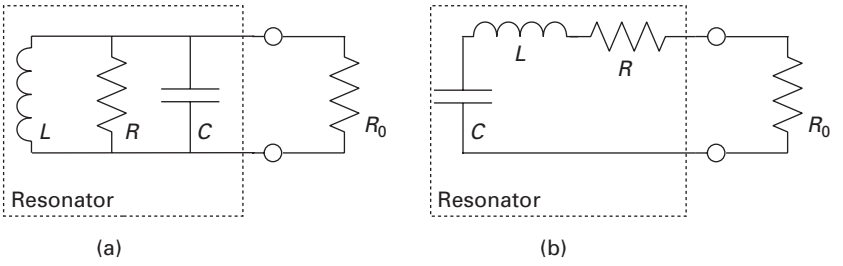

Fig. 5.88 RLC resonator circuits: (a) parallel topology and (b) series topology.

For the parallel resonant circuit (Figure 5.88(a)) at resonance $\omega = \omega_0 = 1/\sqrt{LC}$, the impedance is purely resistive and the dissipated energy during one cycle is

$$E_{\text{diss}} = \frac{1}{2}\frac{\hat{I}^2 R}{\frac{\omega_0}{2\pi}} = \pi\frac{\hat{I}^2 R}{\omega_0}, \tag{5.215}$$

where $\hat{I}$ is the amplitude of the sinusoidal current flowing through the resonator.

The stored energy moves back and forth between the inductor and the capacitor; therefore, it suffices to calculate it for the capacitor:

$$E_{\text{stored}} = \frac{1}{2}C\hat{V}^2 = \frac{1}{2}C\left(\hat{I}R\right)^2. \tag{5.216}$$

Inserting Equations (5.215) and (5.216) into (5.214) yields the Q factor for the parallel RLC resonator. This quality factor is called the *unloaded* Q because the loading resistor (R_0) has not been taken into account:

$$Q_{\text{u}} = \omega_0 RC = \frac{R}{\omega_0 L} = \frac{R}{\sqrt{\frac{L}{C}}}. \tag{5.217}$$

The *loaded* Q is calculated by connecting R_0 in parallel to R:

$$Q_{\text{l}} = \omega_0 C\frac{RR_0}{R + R_0} = \frac{Q_{\text{u}}}{1 + \frac{R}{R_0}}. \tag{5.218}$$

Similarly, we can calculate the unloaded Q for the series resonator (Figure 5.88):

$$Q_{\text{u}} = \frac{\omega_0 L}{R} = \frac{\sqrt{\frac{L}{C}}}{R}, \tag{5.219}$$

and the loaded Q as

$$Q_{\text{l}} = \frac{Q_{\text{u}}}{1 + \frac{R_0}{R}}. \tag{5.220}$$

5.5.2 Self-excitation criteria

Early in this chapter, we considered two-port stability from the viewpoint of avoiding parasitic oscillations in amplifiers. Now the task is to deliberately create instability in a certain frequency range.

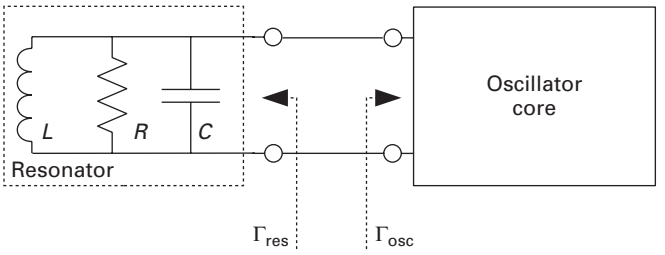

Fig. 5.89 Connection of a resonator to an oscillator core. A parallel RLC resonator is chosen as an example.

In the above resonator examples, the resonators always had a dissipative element associated with it. In practice, this is indeed always the case as resistive and radiative losses are never fully avoidable. In terms of the reflection coefficient seen looking into the resonator, this means

$$|\Gamma_{res}| < 1.$$

The stability boundary can be written as

$$\Gamma_{res}\Gamma_{osc} = 1, \tag{5.221}$$

where Γ_{osc} is the reflection coefficient looking into the oscillator core. Because the reflection coefficients are generally complex entities, Equation (5.221) has to be decomposed into a magnitude and a phase condition:

$$|\Gamma_{res}| \cdot |\Gamma_{osc}| = 1 \tag{5.222}$$

$$\angle(\Gamma_{res}) + \angle(\Gamma_{osc}) = 0. \tag{5.223}$$

The oscillator core will therefore necessarily have to provide $|\Gamma_{osc}| > 1$. Because we have seen in the Smith chart discussion that for all $\mathrm{Re}(Z) = 0, \ldots, \infty$, $|\Gamma| \leq 1$, this means that the real part of the oscillator core input impedance will have to be negative.

Another way of determining the proper conditions for oscillation is to use the Barkhausen self-excitation criterion. The block diagram for this discussion is shown in Figure 5.90. The system with positive feedback has the transfer function:

$$s(\omega) = \frac{1}{1 - F(\omega)}, \tag{5.224}$$

where $F(\omega)$ is the open loop gain. Obviously, the transfer function grows beyond all bounds – becomes unstable – for

$$F(\omega) = 1. \tag{5.225}$$

As $F(\omega)$ is a complex function, two conditions need to be fulfilled:

$$\mathrm{Re}\{F(\omega)\} = 1 \tag{5.226}$$

$$\mathrm{Im}\{F(\omega)\} = 0. \tag{5.227}$$

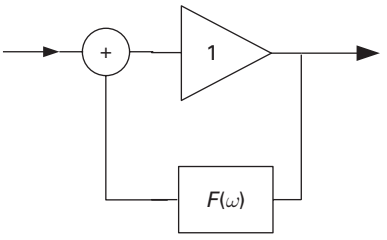

Fig. 5.90 Barkhausen self-excitation criterion: system model with positive feedback.

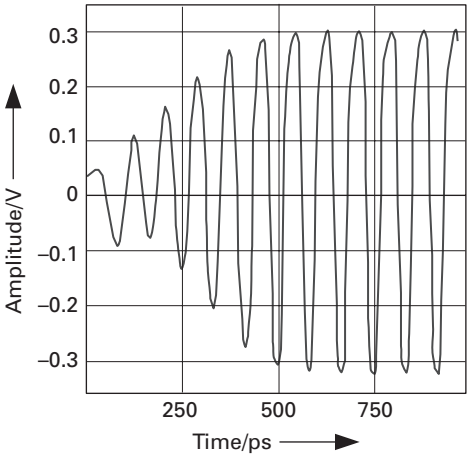

Fig. 5.91 Time-domain simulation of oscillator start-up behaviour.

5.5.3 Non-linearity in oscillators

The self-excitation criteria introduced so far assumed that the systems under investigation were all linear. In practice, however, this would not lead to the desired result of well-controlled sinusoidal signal generation.

Using the Γ criterion (Equation (5.222)), the initial $|\Gamma_{osc}|$ should be significantly (10–20%) higher than $|\Gamma_{res}|$ for the reliable onset of oscillation. However, the oscillation amplitude would then grow beyond all bounds, or in practice until it is limited by the supply voltage.

Fortunately, the active components we are using in the oscillator core to generate the negative resistance all exhibit gain saturation, i.e. the differential gain decreases with increasing signal amplitude. This leads to a self-stabilisation of the oscillation amplitude.

Figure 5.91 shows a time-domain simulation of oscillator start-up behaviour using a non-linear active device model (here, a Si/SiGe HBT). Note how the device linearity leads to a steady-state oscillatory behaviour after only a few cycles. Looking carefully, you will also notice that the initial frequency of oscillation is different from the one in steady state – this is caused by the reactive component of the oscillator core impedance,

which also shows a non-linear behaviour and varies as the amplitude increases. The effect, which is undesirable, is frequently referred to as *chirp*.

5.5.4 Oscillator topologies

The negative resistance necessary to fulfil the condition $|\Gamma_{\text{osc}}| > 1$ for the oscillator core can be generated in a variety of ways:

- The active devices in the oscillator core could have a negative differential resistance of its quasi-stationary I–V characteristics. Examples are tunnel diodes, Gunn diodes or exotic devices such as real-space transfer transistors. Except for Gunn diodes, which are still being used in inexpensive microwave modules (e.g. motion detectors), they have no longer (or never had) any commercial significance.
- The active device could incorporate a transit-time region which leads to a phase lag between the applied voltage and the current through the device. If this phase shift is larger than $\pi/2$ at a given frequency, the resulting impedance at that frequency has a negative real part. An example of such a device is the IMPact ionization Avalanche Transit-Time (IMPATT) diode, which is still being used extensively in millimetre-wave oscillators.
- The most common way to generate negative resistance is the use of positive feedback around an amplifying device.

The last item will be discussed in more detail here.

There are several ways of introducing positive feedback around an amplifier. Four of them are shown in Figure 5.92.

The configuration in Figure 5.92(a) uses magnetic coupling in a transformer around a common-source (or common-emitter) stage. As the common-source amplifier is inverting, a reversal of the winding sense in the transformer is necessary to generate the required positive feedback. Realised with vacuum tubes, this circuit was known very early in the history of radio and is called *Armstrong* (or Meissner) oscillator.

The *Hartley* topology (Figure 5.92(b)) uses a tapped-inductor feedback path around a non-inverting common-drain stage. The *Colpitts* oscillator (Figure 5.92(c)), uses a similar concept, but a capacitive voltage divider instead of the tapped inductor – it is therefore easier to realise in integrated form, and probably the most popular topology for MMIC implementations. When the inductor is replaced by a series LC circuit, the *Clapp* topology results (Figure 5.92(d)).

Due to its popularity, we will examine the Colpitts topology in more detail (Figure 5.93).

The current through capacitor C_2 is

$$i_{C2} = v_{\text{gs}} \left[g_{\text{m}} + \jmath\omega(C_1 + C_{\text{gs}}) \right], \tag{5.228}$$

which leads to the input voltage

$$v_{\text{in}} = v_{\text{gs}} + \frac{i_{C2}}{\jmath\omega C_2} = v_{\text{gs}} \left(1 + \frac{C_1 + C_{\text{gs}}}{C_2} - \jmath \frac{g_{\text{m}}}{\omega C_2} \right). \tag{5.229}$$

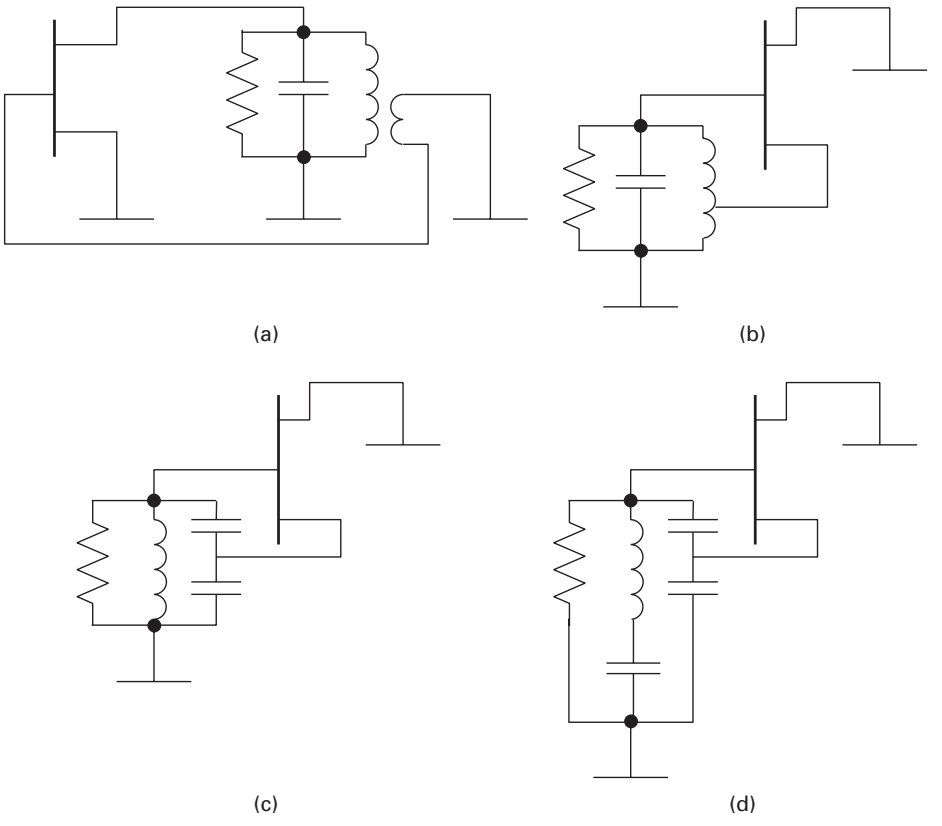

Fig. 5.92 Common oscillator topologies: (a) Armstrong or Meissner, (b) Hartley, (c) Colpitts and (d) Clapp.

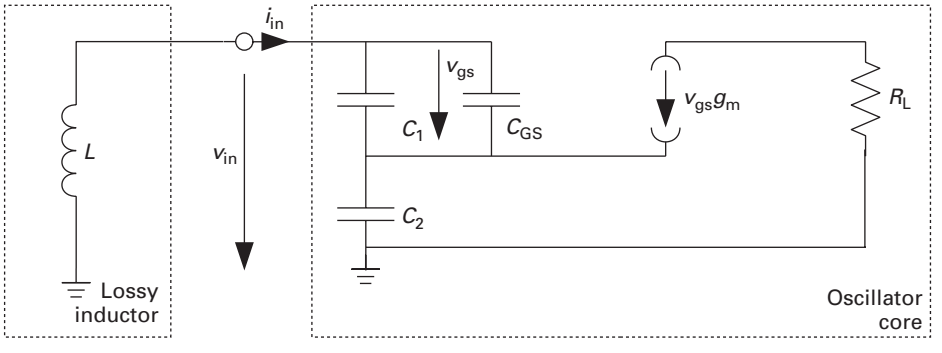

Fig. 5.93 Simple equivalent circuit of the Colpitts oscillator topology.

The input current is

$$i_{\text{in}} = v_{\text{gs}} \jmath\omega(C_1 + C_{\text{gs}}). \tag{5.230}$$

This allows us to calculate the input impedance of the oscillator core:

$$Z_{\text{in}} = \frac{v_{\text{in}}}{i_{\text{in}}} = -\frac{g_{\text{m}}}{\omega^2 C_2(C_1 + C_{\text{gs}})} + \frac{1}{\jmath\omega}\frac{C_2(C_1 + C_{\text{gs}})}{C_1 + C_2 + C_{\text{gs}}}. \tag{5.231}$$

The negative real part of the input impedance is obvious; the imaginary part is simply the series connection of the capacitances at the input.

Another aspect of the Colpitts oscillator is apparent: especially at high frequencies, the capacitor C_1 is not really necessary, the C_{gs} of the device suffices.

A very popular topology, especially for RF CMOS circuits, is the *cross-coupled pair*, shown in Figure 5.94(a). It is a differential pair with the gate of each transistor connected to the drain of the opposite transistor. Because of the 180° phase shift between both branches, this forces the small-signal gate voltages to be

$$v_{\text{gs1}} = -v_{\text{gs2}}. \tag{5.232}$$

Remember that in a differential pair under perfectly differential excitation, all nodes along the vertical centre plane are virtual grounds. Then the circuit behaviour can be completely described using the half-circuit in Figure 5.94(b). To maintain symmetry, C_0 has been replaced by the series connection of two capacitors of $2C_0$, while L_0 is divided into the series connection of two inductors with $L_0/2$.

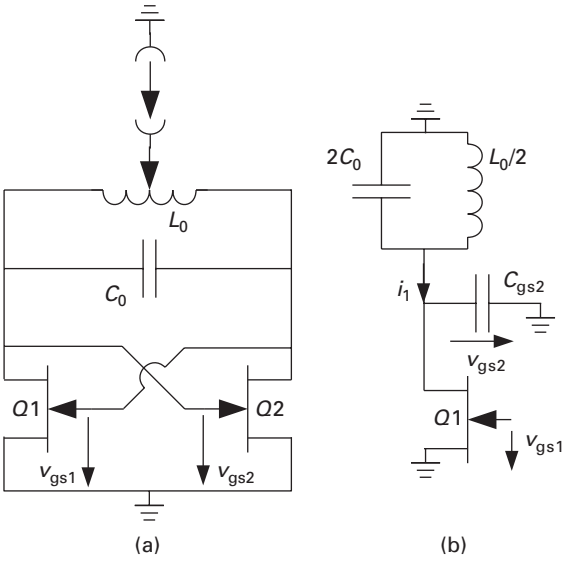

Fig. 5.94 (a) Simplified circuit of a cross-coupled pair oscillator and (b) its equivalent half-circuit.

Let us now calculate the admittance seen by the resonator towards the oscillator core:

$$\underline{Y}_1 = \frac{i_1}{v_{gs2}} = \frac{g_m v_{gs1} + v_{gs2} \jmath \omega C_{gs2}}{v_{gs2}} \\ = -g_m + \jmath \omega C_{gs2}, \tag{5.233}$$

using Equation (5.232). The negative real part is hence $-g_m$, and the oscillation frequency is

$$\omega_0 = \frac{1}{\sqrt{L_0(C_0 + C_{gs})}}, \tag{5.234}$$

assuming $C_{gs1} = C_{gs2} = C_{gs}$.

Finally, negative resistance can also be generated using common-gate (or common-base) configurations.

In the common-gate amplifier stage shown in Figure 5.95, note the inductance L_0 inserted into the gate lead. We will calculate the small-signal input impedance for this circuit. The input current is

$$i_1 = -v_{gs}(g_m + \jmath \omega C_{gs}). \tag{5.235}$$

The current through L_0 is $v_{gs} \jmath \omega C_{gs}$, so the input voltage v_1 is (after a short calculation)

$$v_1 = -v_{gs}(1 - \omega^2 L_0 C_{gs}). \tag{5.236}$$

The input impedance is then, after separation into its real and imaginary parts,

$$\underline{Z}_1 = \frac{v_1}{i_1} = \frac{g_m(1 - \omega^2 L_0 C_{gs})}{g_m^2 + \omega^2 C_{gs}^2} - \jmath \omega C_{gs} \frac{1 - \omega^2 L_0 C_{gs}}{g_m^2 + \omega^2 C_{gs}^2}. \tag{5.237}$$

Now remember that the transit frequency $\omega_T \approx g_m / C_{gs}$ and assume that

$$\frac{1}{L_0 C_{gs}} \ll \omega^2 \ll \omega_T^2.$$

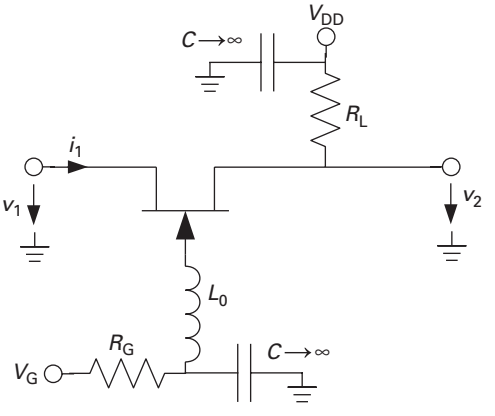

Fig. 5.95 Oscillator core using a common-gate stage.

Then, we obtain

$$\underline{Z}_1 = -\frac{\omega^2 L_0}{\omega_T} + \jmath\omega L_0 \frac{\omega^2}{\omega_T^2}. \tag{5.238}$$

Provided that the frequency is larger than the resonant frequency of L_0, C_{gs}, this circuit will, therefore, also generate a negative resistance. The imaginary part is inductive.

Voltage-controlled oscillators

Electronic control of the oscillation frequency is a very common requirement. While both variable inductance and variable capacitance concepts are in principle feasible, variable inductance approaches suffer from poor integrability. Attempts to realise integrated variable inductors using micro-electro-mechanical structures (MEMS) exist, but have not found practical use yet.

In practical applications, variable capacitors, in turn, are always realised using *varactor diodes*, which may build upon p–n junction diodes, Schottky diodes, or MOS diodes, depending on the underlying semiconductor technology. In each case, the capacitance of a blocking diode structure is varied by changing the voltage across the diode. Variable capacitors using MEMS have been investigated quite extensively, but again have not reached sufficient maturity for commercial applications so far.

The voltage-controlled oscillator (VCO) example in Figure 5.96 uses the cross-coupled differential pair topology introduced in Figure 5.94(a). The fixed capacitor has been replaced by two varactor diodes. The tuning voltage is connected to a virtual ground point, which facilitates decoupling between the RF and the DC control paths.

The series resistance of the varactor diodes should not be overlooked – it may substantially lower the overall resonator quality factor. The total quality factor can be shown to be

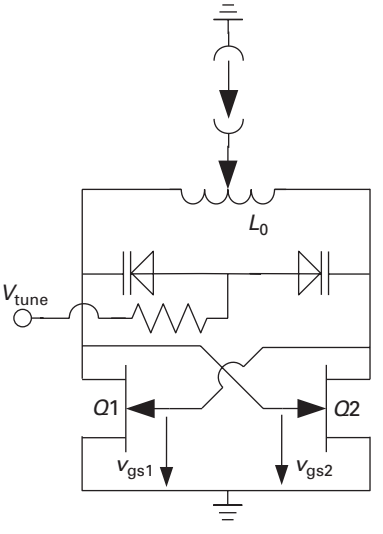

Fig. 5.96 VCO using a cross-coupled topology.

$$Q_{\text{total}} = \frac{Q_{\text{C}} Q_{\text{L}}}{Q_{\text{C}} + Q_{\text{L}}}, \tag{5.239}$$

where

$$Q_{\text{L}} = \frac{\omega_0 L}{R_{\text{L}}} = \sqrt{\frac{L}{C}} \frac{1}{R_{\text{L}}}, \; Q_{\text{C}} = \frac{1}{\omega_0 C R_{\text{C}}} = \sqrt{\frac{L}{C}} \frac{1}{R_{\text{C}}}$$

are the quality factors of an inductor with series resistance R_{L} and a capacitor with series resistance R_{C}, respectively, in a parallel resonant circuit. The assumption of weak losses has been made:

$$\frac{\omega_0^2 L^2}{R_{\text{L}}^2} \gg 1, \omega_0^2 R_{\text{C}}^2 C^2 \ll 1.$$

5.5.5 Noise in oscillators

The noise phenomena in active devices of course also affect oscillator performance. In principle, two things may happen:

- The oscillation amplitude may fluctuate randomly with time – *amplitude noise*.
- The phase of the sinusoidal signal may fluctuate randomly with time – *phase noise*.

Of the two, amplitude noise is the least critical. First of all, the gain compression effect which leads to a stable oscillation condition in the first place also reduces random amplitude fluctuations. Also, in frequency translation applications (frequency up- or down-conversion), oscillators typically work into switch-type mixers where the oscillator amplitude has little effect on the conversion efficiency, provided it is still sufficient for switching operation (see Section 5.6, p. 396).

For these reasons, we will restrict our discussions to phase noise, which has a much stronger impact on system performance.

Phase noise describes the random fluctuations of the oscillator phase with time. Mathematically,

$$s(t) = A \sin\left[\omega t + \phi(t)\right]. \tag{5.240}$$

In the frequency domain, these phase fluctuations are manifest in *noise sidebands* close to the carrier (see Figure 5.97). The spectral power density of the noise sidebands decreases with increasing distance to the carrier. They have a number of detrimental effects, for example

- in transmitters, they lead to interference in nearby channels;
- in receivers, phase noise increases the perceived in-channel noise due to a phenomenon known as *reciprocal mixing*, which is an intermodulation effect between a strong interferer and the local oscillator noise sidebands.

The customary figure of merit for phase noise suppression is the ratio between the carrier power and the power of the noise sidebands in a 1 Hz bandwidth at an offset Δf from the carrier. It is typically expressed in 'dBc/Hz'.

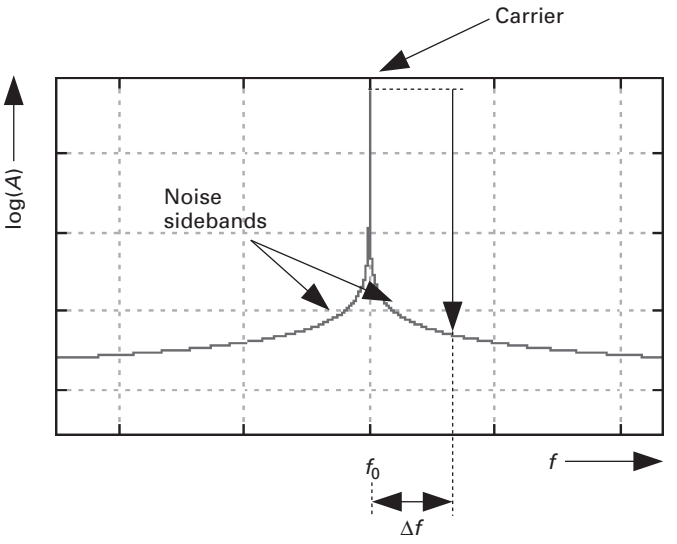

Fig. 5.97 Phase noise sidebands around a carrier in the frequency domain.

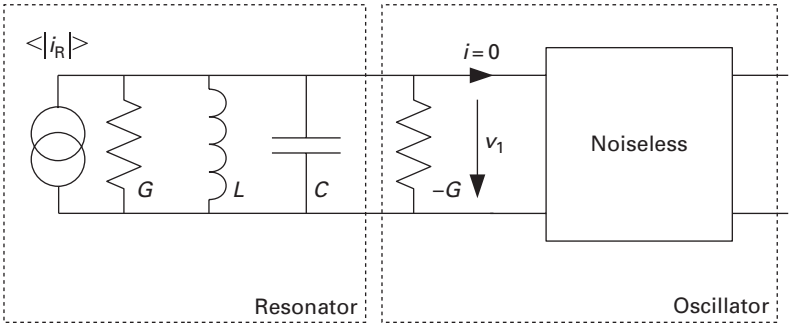

Fig. 5.98 Noise equivalent circuit of an ideal oscillator core with a lossy resonator.

To appreciate the effect of the resonator on the oscillator noise performance, consider Figure 5.98. The oscillator core itself is assumed to be noiseless (a strong over-simplification). The imaginary part of its input admittance is merged into the LC resonator; at the resonant frequency, the oscillator core presents a real part $-G$ which just offsets the resonator loss G.

The resonator dissipative element, G, creates a thermal noise current whose noise phasor is (see Equation (5.82))

$$\left\langle |i_R|^2 \right\rangle = 8kTBG,$$

where B is the measurement bandwidth.

The voltage v_1 is $i_R \cdot Z_{res}$, where Z_{res} is the resonator impedance. Because the resonator losses are exactly compensated, it is the impedance of an ideal parallel LC resonator:

$$\underline{Z}_{\text{res}} = \frac{\jmath\omega L}{1-\omega^2 LC} = \frac{\jmath\omega L}{1-\frac{\omega^2}{\omega_0^2}}. \tag{5.241}$$

Taylor series expansion of the denominator, aborted after the linear term, leads to an approximate impedance for small deviations $\delta\omega$ from the resonance ω_0:

$$\underline{Z}_{\text{res}}(\omega_0 \pm \Delta\omega) \approx \mp\frac{\jmath\omega_0 L}{\frac{2\Delta\omega}{\omega_0}} = \mp\jmath\frac{\omega_0}{2\Delta\omega}\frac{1}{QG}, \tag{5.242}$$

using $Q = R/\omega_0 L = 1/(\omega_0 LG)$.

The phasor of the noise-generated voltage v_1 is then

$$\left\langle |v_1|^2 \right\rangle = \left\langle |i_{\text{R}}|^2 \right\rangle |\underline{Z}_{\text{res}}|^2 \approx \frac{8kTB}{GQ^2\left(\frac{2\Delta\omega}{\omega_0}\right)^2}, \tag{5.243}$$

using the approximation in Equation (5.242).

Comparing the noise power in bandwidth B to the signal power P_{S}, finally we obtain the phase noise suppression:

$$L(\Delta\omega) = \frac{2kTB}{P_{\text{S}}}\left(\frac{\omega_0}{2\Delta\omega Q}\right)^2. \tag{5.244}$$

Note the quadratic dependence of the phase noise suppression on the resonator Q.

In practical cases, the oscillator core is not noiseless. Furthermore, we need to take low-frequency noise phenomena into account, which lead to a stronger increase in phase noise close to the carrier. Leeson (1966) [26] introduced the following semi-empirical formula, which builds upon Equation (5.244):

$$L(\Delta\omega) = \frac{2FkTB}{P_{\text{S}}}\left[1+\left(\frac{\omega_0}{2\Delta\omega Q}\right)^2\right]\left(1+\frac{2\pi f_{\text{c}}}{|\Delta\omega|}\right). \tag{5.245}$$

The additional factors are F – accounting for the additional noise in the oscillator core – and f_{c}, the cutoff frequency for low-frequency ('$1/f$') noise.

Figure 5.99 shows an example calculation using Equation (5.245). We clearly distinguish three different regions:

(i) Close to the carrier, the noise power drops with −30 dB/decade. Here, the low-frequency noise increase in the oscillator core's active devices dominate.
(ii) Further out, the decay is −20 dB/decade, corresponding to the earlier calculations. Here, white noise sources (such as the thermal noise provided by the lossy resonator) dominate.
(iii) Far away from the carrier, a noise floor is visible, but this is generally not very relevant, unless F is very large.

If low phase noise oscillators are a requirement, both high quality factor resonators and active devices with low-frequency noise are a must. In general, bipolar devices (including HBTs) will fare much better than FETs. On-chip resonators generally have much lower Q than off-chip resonators can achieve.

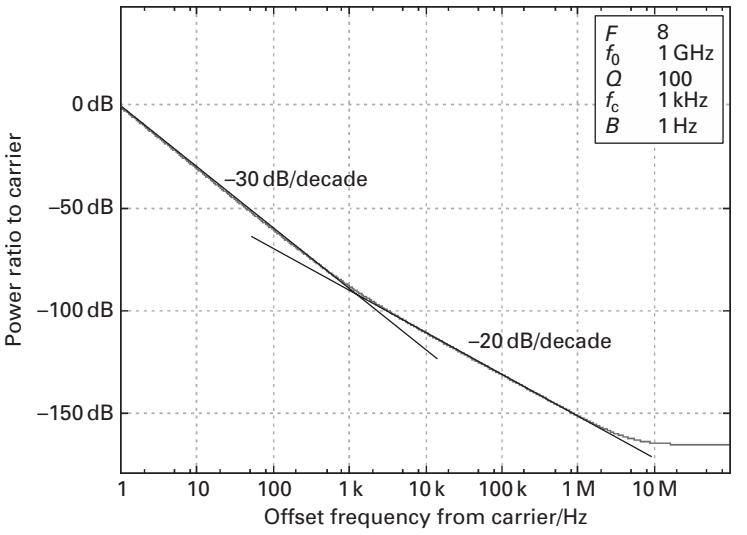

Fig. 5.99 Simulated phase noise sideband of an oscillator using the Leeson equation.

5.6 Mixers

Mixers are generally frequency-translation components, with a variety of applications in analogue signal processing, such as frequency shifting of signals (up and down conversion), phase shifting, modulation and demodulation. A special class of mixers – four-quadrant multipliers – can also be used in correlators, for example in impulse-radio ultra-wideband receivers.

The mathematics behind mixer operation has been reviewed already in the context of non-linear amplification (p. 370 and following). Recall that if we take a signal consisting of two sinusoidal components of different frequencies ω_1 and ω_2 and feed it into a non-linear two-port, the output $h(t)$ can be described by a Taylor series expansion:

$$h(t) = k_1 \left[a \sin(\omega_1 t) + b \sin(\omega_2 t)\right] + k_2 \left[a \sin(\omega_2 t) + b \sin(\omega_2 t)\right]^2 + \ldots \quad (5.246)$$

The quadratic term expands to

$$\begin{aligned} &k_2 \left[a \sin(\omega_2 t) + b \sin(\omega_2 t)\right]^2 \quad (5.247)\\ &= k_2 \left[a^2 \sin^2(\omega_1 t) + b^2 \sin^2(\omega_2 t) + 2ab \sin(\omega_1 t) \sin(\omega_2 t)\right]. \end{aligned}$$

The product term can be expressed as

$$2ab \sin(\omega_1 t) \sin(\omega_2 t) = ab \left\{\cos[(\omega_1 - \omega_2)t] - \cos[(\omega_1 + \omega_2)t]\right\}. \quad (5.248)$$

Any non-linear system under two-tone excitation will therefore produce spectral components at the sum and difference of the input signals. We also conclude that an *analogue multiplier* would be the ideal mixer, because it *only* produces the sum and difference of the input spectral components.

Incidentally, the interaction used here for frequency mixing purposes, was called *second-order intermodulation* in the context of non-linear amplifiers.

5.6.1 Transconductance multiplier

A simple analogue multiplier can be realised using a bipolar differential amplifier. It utilises the fact that in bipolar transistors, the small-signal transconductance is linearly dependent on the large-signal collector current in the operating point. Now take a differential pair Q_1, Q_2 with a common current source transistor Q_3 (Figure 5.100). The RF input voltage v_1 is fed differentially into the top transistor pair (superimposed on the bias voltage V_0), while the oscillator current signal, I_{LO}, is fed single-endedly into the base of Q_3. Let Q_3 have a current gain B. Then, the differential output signal is

$$v_2 = v_1 g_m R_L = v_1 I_{LO} \frac{B R_L}{2V_T}, \tag{5.249}$$

because

$$g_m = \frac{B I_{LO}}{2V_T}.$$

The output voltage is therefore proportional to the product of the input voltage and the local oscillator current.

This simple circuit, however, has a number of drawbacks. First of all, it will only work for $I_{LO} > 0$. Secondly, the input voltage has to be much smaller than V_T (or 26 mV at room temperature) to fulfil the small-signal assumption. Finally, it will only work with bipolar transistors.

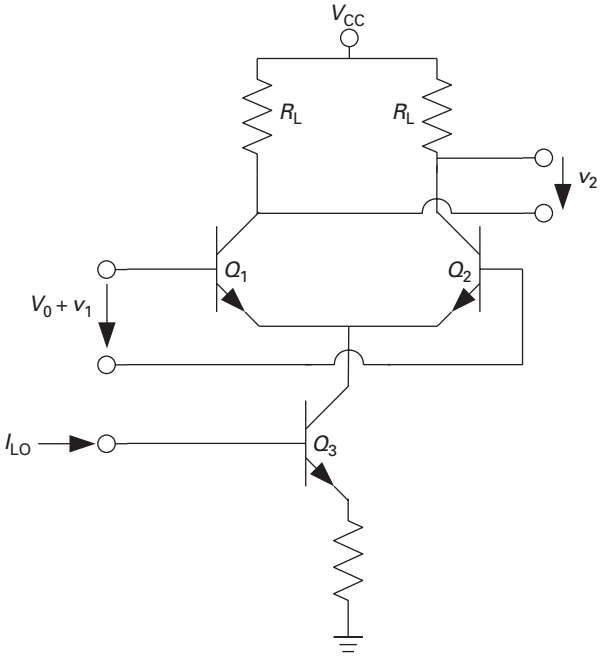

Fig. 5.100 Transconductance multiplier circuit.

5.6.2 Resistive mixer

After looking at a mixer circuit which works only with bipolar transistors, let us briefly consider a circuit which will work only with FETs. To appreciate its mode of operation, examine the output I–V characteristics of a FET (see Figure 5.101).

At low V_{DS}, the relationship between I_D and V_{DS} is an approximately linear one and can therefore be accurately described by the channel conductance g_{DS}. Assuming a simple Statz–Curtice model for the FET:

$$I_D(V_{GS}, V_{DS}) = \frac{\beta(V_{GS} - V_P)^2}{1 + \alpha(V_{GS} - V_P)} \tanh(\gamma V_{DS}),$$

where V_P is the pinch-off or threshold voltage and α, β and γ are model parameters, we find

$$g_{DS} = \gamma \frac{\beta(V_{GS} - V_P)^2}{1 + \alpha(V_{GS} - V_P)} \left[1 - \tanh^2(\gamma V_{DS})\right] \approx \gamma \frac{\beta(V_{GS} - V_P)^2}{1 + \alpha(V_{GS} - V_P)}, \qquad (5.250)$$

provided that $\tanh^2(\gamma V_{DS}) \ll 1$ or $\gamma V_{DS} < 0.3$.

Now the FET is placed in a circuit as shown in Figure 5.102. The inductor L enforces a steady-state bias point $V_{DS,0} = 0$, but is invisible at the LO, RF or IF frequencies. The gate bias can be set to a suitable V_{GS} for optimum mixer operation. The FET, together with the generator resistance of the RF port R_{RF}, forms a resistive voltage divider whose division ratio is modulated by the gate potential:

$$V_{DS}(t) = \frac{V_{RF}(t)}{1 + g_{DS}(t) R_{RF}}. \qquad (5.251)$$

The periodic variation of g_{DS} by the gate potential $V_{GS} = V_{GS,0} + V_{LO} \sin(\omega_{LO} t)$:

$$g_{DS} = \beta\gamma \left[V_{GS,0} - V_P + V_{LO} \sin(\omega_{LO} t)\right]^2 \qquad (5.252)$$

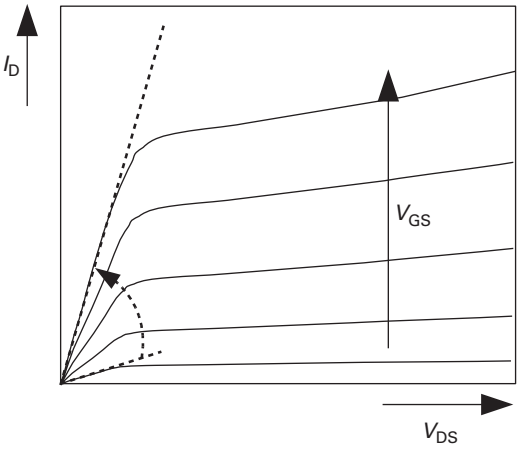

Fig. 5.101 FET output I–V characteristics with indication of variable resistor operation.

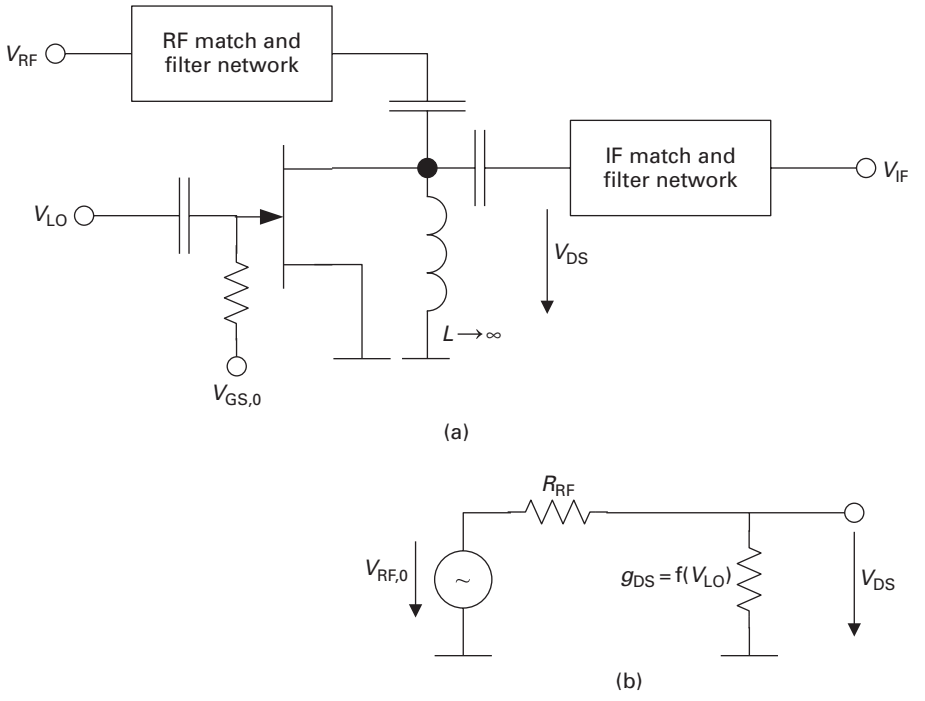

Fig. 5.102 (a) Resistive mixer configuration and (b) its simplified equivalent circuit.

results in the desired mixing action (pure square law behaviour, i.e. $\alpha = 0$, is assumed here, see Equation (5.250)). Note that Equation (5.252) only holds for $V_{LO} < V_{GS,0} - V_P$.

Frequently, the local oscillator voltage will be chosen such that $V_{LO} \geq V_{GS,0} - V_P$ – the channel conductance is then switched between two saturated states, $g_{DS} = 0$ and a high state essentially limited by the source and drain series resistances, which were initially omitted in the simplified discussion. Under this condition, and assuming $V_{RF}(t) = V_{RF} \sin(\omega_{RF} t)$, the time-dependent drain-source voltage becomes

$$V_{DS}(t) = \frac{V_{RF} \sin(\omega_{RF} t)}{R_S + R_D + R_{RF}} \left[R_S + R_D + R_{RF} \mathrm{rect}(\omega_{LO} t) \right]. \tag{5.253}$$

Since the Fourier series of the rect function is

$$\mathrm{rect}(\omega_{LO} t) = \frac{1}{2} + \frac{2}{\pi} \left[\sin(\omega_{LO} t) + \frac{\sin(3\omega_{LO} t)}{3} + \ldots \right], \tag{5.254}$$

this switching mode operation leads to the desired multiplication with the fundamental frequency of the local oscillator, but also with higher-order harmonics – the latter operation is referred to as *sub-harmonic pumping*.

Figure 5.103 shows as an example the measured conversion gain of a resistive mixer, as a function of the local oscillator power. At low P_{LO}, the conversion gain

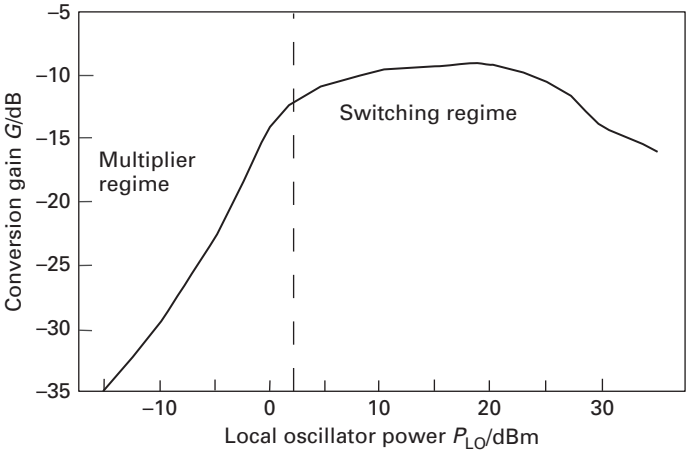

Fig. 5.103 Measured conversion gain of a resistive mixer.

increases proportionally with the local oscillator power – the mixer acts like a multiplier. For higher P_{LO}, the mixer enters the switching regime – the conversion gain is roughly independent of the local oscillator power. This behaviour can be seen, in variations, for any of the mixer circuits discussed here. In standard wireless systems, mixers are operated in the switching regime, because independence from local oscillator power fluctuations is highly desirable. For example, it reduces the effect of oscillator amplitude noise. There are applications, however, where operation in the multiplier regime is wanted. An excellent example are correlation receivers for impulse-radio ultra-wideband systems, where the formation of a cross-correlation between the received pulse and a template pulse in the receiver calls for a true multiplier.

The resistive mixer has the advantage of high simplicity, and, most prominently, very high linearity with respect to the RF port. The latter is especially true when local oscillator power leaking through the gate-drain capacitance is short-circuited to ground, so that the drain potential is not modulated at the LO frequency [27].

Disadvantages are the significant required local oscillator voltage swing and the conversion loss inherent to the voltage divider principle.

An important issue for practical mixers is *port isolation* – ideally, power fed into the LO port should not be present at the RF and IF ports, and power fed into the RF port should not be present at the IF port in downconverters, while power fed into the IF port should not leak to the RF port in upconverters. In the resistive mixer, the isolation between the RF and IF ports is realised by filters only, which is a significant disadvantage. The LO-to-RF and LO-to-IF isolations are somewhat better because the leakage path is via the gate-drain capacitance, but it may still be too high.

Mixer concepts where port isolation is assisted by destructive interference are much better, in this respect. They will be treated next.

5.6.3 Single-balanced mixer

Consider the circuit in Figure 5.104. Transistor Q_1 is a common-source amplifier stage fed by the RF signal. Its load is formed by the differential pair Q_2, Q_3 with load resistances R_L. R_2 and R_3 are for biasing only. Q_2 and Q_3 are driven by the local oscillator – but due to the balun transformer,[7] their gate signals are exactly 180° out of phase.

Assume that the local oscillator signal is large enough so that the Q_2 and Q_3 are being switched off alternately. Then, the differential voltage V_{IF} can be written as

$$V_{IF} = R_L I_{D1} \left[2 \cdot \mathrm{rect}(\omega_{LO} t) - 1\right], \tag{5.255}$$

where I_{D1} is the drain current of Q_1. Provided that Q_1 is operated in the small-signal regime, the small-signal IF output voltage is, considering the fundamental frequency component of the local oscillator signal only,

$$V_{IF} = \hat{V}_{RF} \frac{4 g_{m,1} R_L}{\pi} \sin \omega_{LO} \sin \omega_{RF} t, \tag{5.256}$$

using Equation (5.254) and $V_{RF} = \hat{V}_{LO} \sin(\omega_{LO} t)$.

The desired multiplication is again visible, producing spectral components at $\omega_{RF} \pm \omega_{LO}$.

The mixer uses a special property already discussed in the context of differential amplifiers: the common-source connection of transistors Q_1 and Q_2 is a virtual

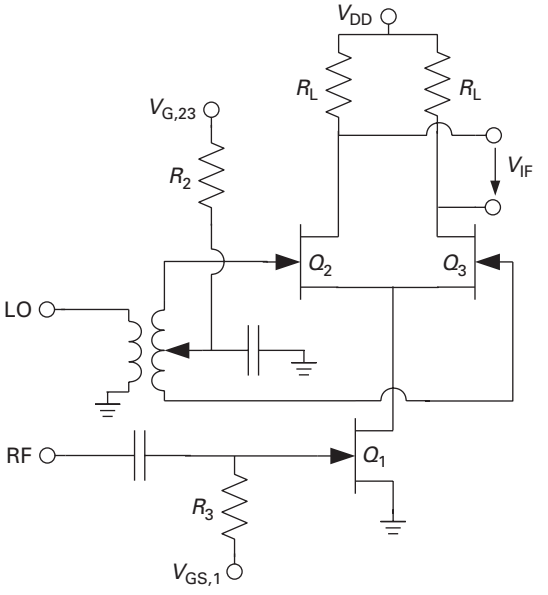

Fig. 5.104 Single-balanced mixer circuit.

[7] Note that baluns are drawn as transformers in this and the following circuit diagrams. At micro- and millimetre-wave frequencies, they are rarely transformers, but may be realised using transmission line segments, or as active baluns.

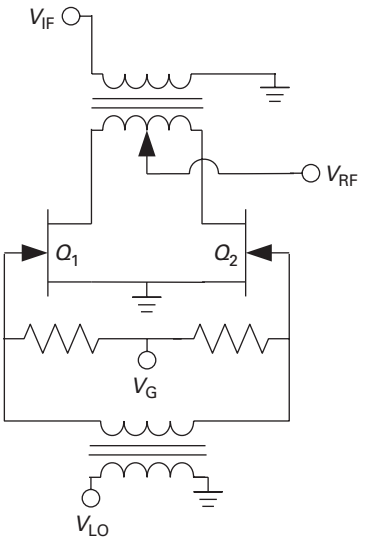

Fig. 5.105 Single-balanced upconversion mixer with FETs in shunt configuration.

ground for purely differential excitation. Therefore, the local oscillator and RF ports are highly decoupled by virtue of destructive interference. LO-to-RF leakage is very critical, because the local oscillator signal must not radiate via the antenna.

The circuit shown in Figure 5.105 is also a single-balanced mixer, used to upconvert a signal at an intermediate frequency to a higher frequency (RF). As opposed to the circuit shown in Figure 5.104, where the switches were in series configuration, they are in shunt here, alternately connecting the two ends of the top balun to ground, periodically changing the sign of the IF signal at the RF port and thus producing the desired multiplication of the LO signal with IF. As upconversion is what we desire, a subsequent filter will have to suppress the difference frequency and pass only the sum. Because the RF port is connected to the virtual ground connection with respect to fully differential excitation of Q_1 and Q_2, destructive interference of potential LO leakage at the IF port will again ensure a very high LO-to-RF isolation.

5.6.4 Double-balanced mixer

The configuration shown in Figure 5.104 is a building block of the double-balanced mixer, probably the most popular active mixer configuration in MMIC design today. These topologies are commonly referred to as *Gilbert cells* [12].

In the double-balanced mixer (Figure 5.106), all signals have to be applied in a differential fashion – this is indicated by the presence of three baluns. Transistors Q_1 and Q_2 form a differential amplifier, connected to another pair of differential amplifiers Q_3-Q_6. In the original publication by Gilbert, the configuration is operated as a true four-quadrant multiplier, while in most applications, the top four transistors are switched by the LO signal between two states – they are jointly referred to as the *switching quad*. Only the latter mode shall be discussed here. In this case, the double-balanced

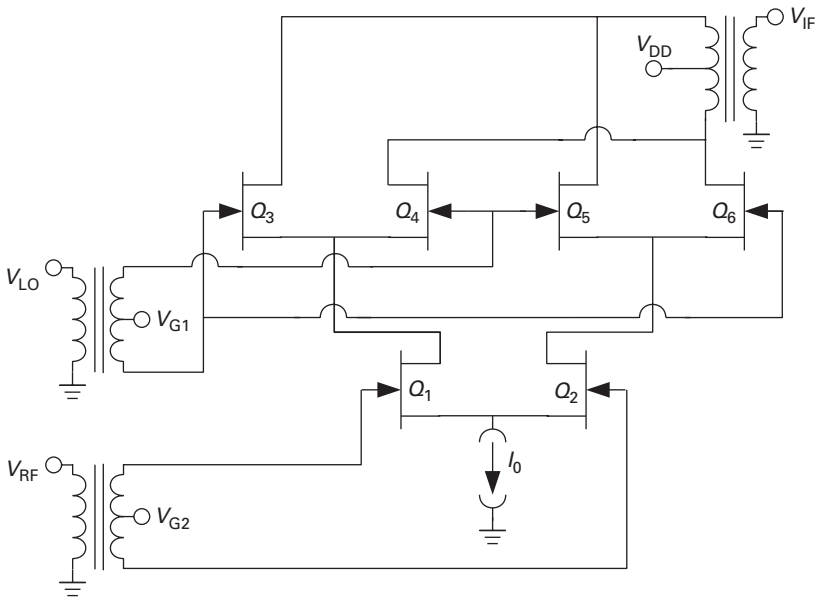

Fig. 5.106 Double-balanced mixer with Gilbert cell topology.

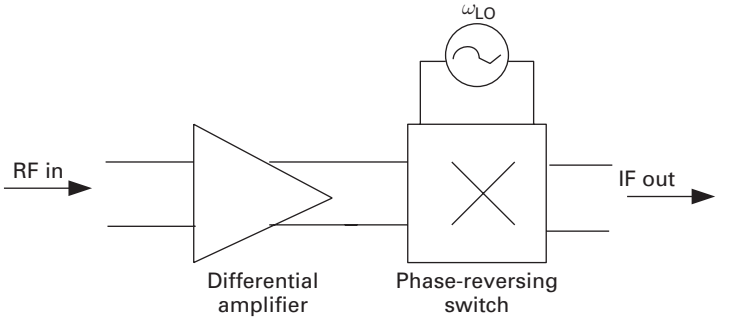

Fig. 5.107 Block diagram of a double-balanced mixer of the Gilbert cell type, partitioned into a differential amplifier and a phase-reversing switch.

mixer can be viewed as the cascade of a differential amplifier and a phase-reversing switch, such that the amplified RF signal changes its phase by 180° with the period of the LO signal. This is shown schematically in Figure 5.107.

The big advantage of the double-balanced mixer is that all ports are now decoupled by destructive interference. For example, the differential RF signal leakage cancels out at both joint gate connections of the switching quad. Likewise, the LO leakage cancels at the source connections of Q_3, Q_4 and Q_5, Q_6, respectively.

A disadvantage of the circuit shown in Figure 5.106 is the rather large voltage headroom required. On top of the required minimum drain-source voltage of the switching quad and the differential amplifier Q_1, Q_2, additional headroom is required for the current source I_0. The latter is frequently replaced by a parallel resonant circuit, tuned to

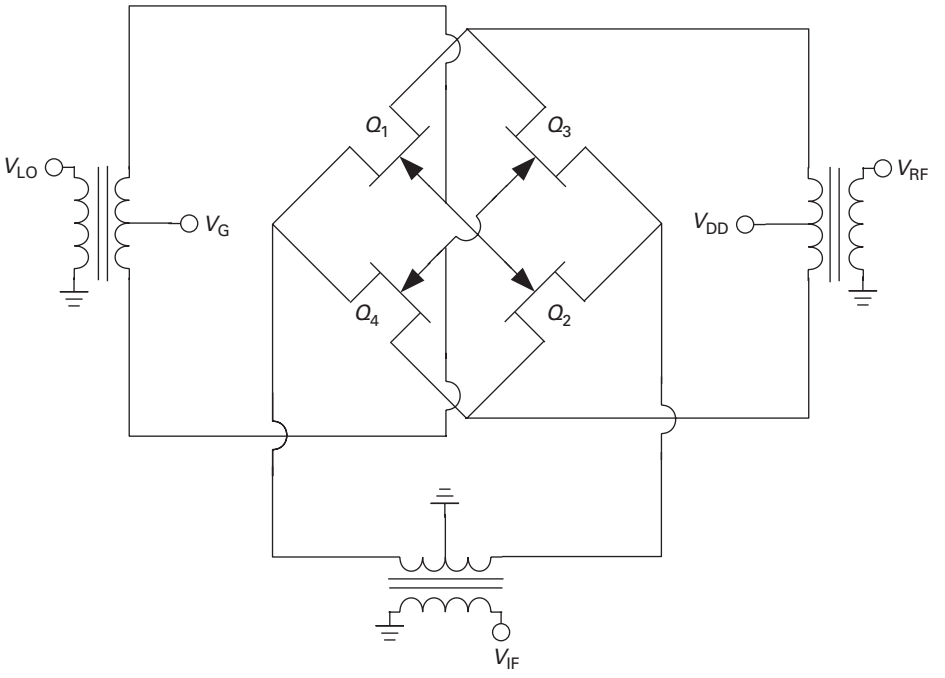

Fig. 5.108 Double-balanced mixer with a ring topology.

ω_{RF}. It provides the required high impedance for the RF signal, but provides a zero DC resistance, lowering the supply voltage requirements.

Double-balanced mixers can also be realised using resistive mixer principles (see Section 5.6.2). This has the advantage of very low power consumption, but of course the circuit will not be able to produce any conversion gain.

Figure 5.108 shows an example. Depending on the choice of the drain potential V_{DD}, this circuit can be operated as a resistive mixer (V_{DS} low, in the linear regime), or the transistors may act as current switches (V_{DS} in the saturated region). The RF signal is applied to the drains of the transistors. The LO signal alternately turns on transistors Q_1, Q_2 and Q_3, Q_4, leading to a periodic phase reversal of RF signal at the IF port.

5.6.5 Micromixer

The micromixer concept, shown in Figure 5.109, evolved from the Gilbert multiplier topology and was also published by B. Gilbert [13]. The transistors $Q_4 - Q_7$ are the switching quad, as before – the circuit is drawn using bipolar transistors here, but would work as well with FETs. The topology is somewhat simplified, for example the baluns are not included.

The local oscillator signal is again applied in differential format. The RF signal, however, is single-ended and applied to the amplifier structure formed by the transistors $Q_1 - Q_3$. Q_1 is a common-base amplifier stage and Q_3 a common-emitter stage. Because the former provides a non-inverting and the latter an inverting voltage gain,

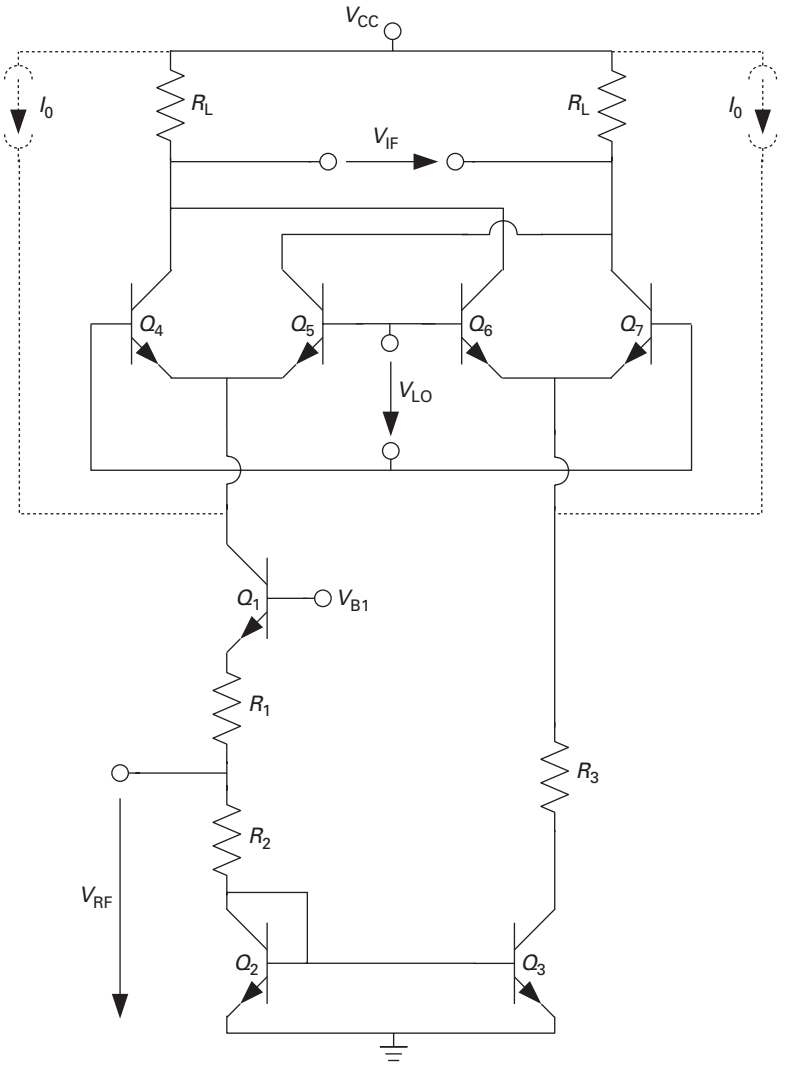

Fig. 5.109 Simplified micromixer topology using bipolar transistors.

the signals fed into the two branches of the switching quad are 180° out of phase – the circuit doubles as an single-ended-to-differential converter. Q_2 forms a current mirror with Q_3 such that the currents in both branches are equal.

Frequently, the gain of the common-base/common-source amplifier/balun is increased by injecting additional current into the two branches. This is indicated in Figure 5.109 by the two dashed current sources I_0.

Compared to the original Gilbert multiplier, the input stage of the micromixer can be made more linear, and can be designed to achieve a broadband match, using the resistors R_1 and R_2.

5.6.6 Diode-based mixers

While the circuit design descriptions in this chapter (and, for that matter, the whole book) emphasise concepts with active components, the use of diodes in mixers must be mentioned here because they are still very commonly employed, especially at millimetre-wave frequencies where transistors lose their usefulness. The diode of choice is the Schottky (metal–semiconductor) diode due to the absence of carrier storage effects when switching from forward to reverse bias.

As in the case of operating a FET as a resistive or switching mixer, the principle is the change of the differential resistance. For a diode, this is shown in Figure 5.110. Clearly, the differential diode conductance $g_D = dI_D/dV_D$ is very low for bias point A (or equally for $V_D < 0$) and very high for bias point B. We will now use the local oscillator again to periodically change the diode between these two states.

Consider Figure 5.111, which depicts the very popular *diode ring mixer*. The circuit is very similar to the FET ring mixer shown earlier (Figure 5.108). Assume that the RF signal is always much smaller than the local oscillator (LO) signal. Then, the diode state will depend on the applied LO signal only. Either the left or the right diode pair may conduct and exhibit a high differential conductance. The diode ring acts as a phase-reversing switch, which connects the RF signal to the intermediate frequency (IF) load with periodically alternating polarity.

Similarly, single-balanced and single-ended (unbalanced) mixer topologies can be realised using diodes. This shall not be further expanded here.

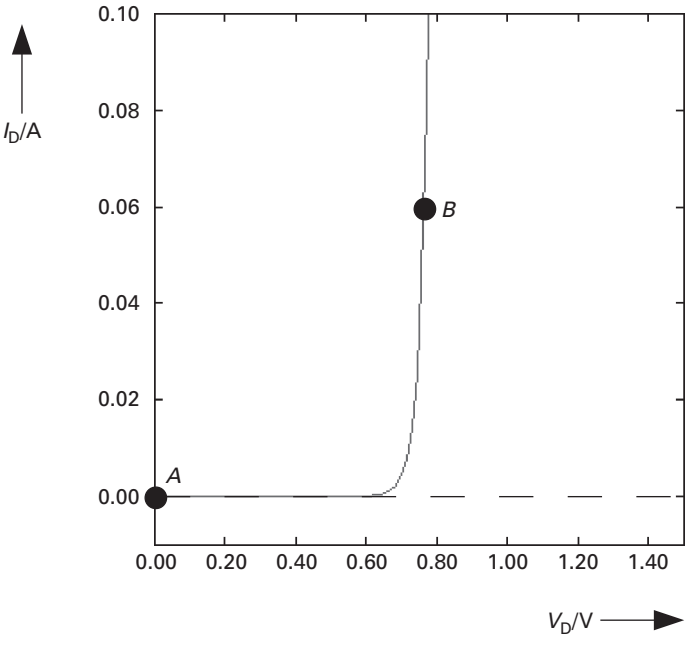

Fig. 5.110 Example Schottky diode I–V characteristics, with bias point indicated for small (A) and high (B) differential conductance.

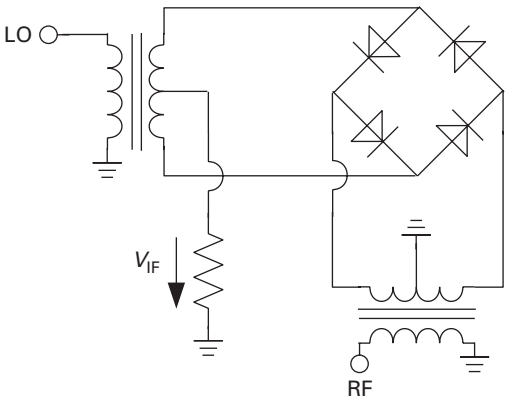

Fig. 5.111 Diode ring mixer.

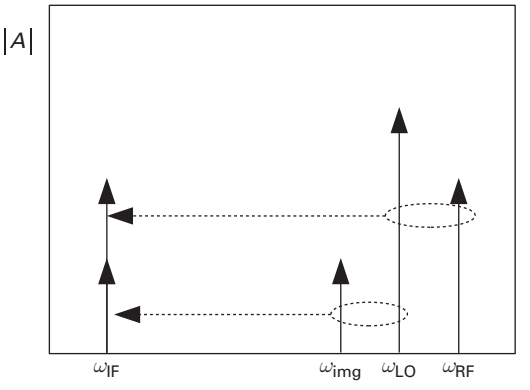

Fig. 5.112 The image frequency problem in downconversion mixing: ω_{RF} and ω_{img} both produce components at ω_{IF} when mixed with ω_{LO}.

5.6.7 Image-rejection mixer topologies

One problem inherent to any of the mixers discussed here still needs to be mentioned. When operated as a downconverter (converting an RF signal at a higher frequency to a lower IF), it uses the fact that the multiplication of sinusoidal signals produces a spectral component at the difference of the two initial frequencies, as was discussed in Equation (5.248).

$$2\sin(\omega_1 t)\sin(\omega_2 t) = \cos\left[(\omega_1 - \omega_2)t\right] - \cos\left[(\omega_1 + \omega_2)t\right].$$

Consider now a case where $\omega_1 = \omega_{RF}$ is the RF signal and $\omega_2 = \omega_{LO}$ is the LO frequency, which is lower than ω_{RF}. The difference is the IF: $\omega_{RF} - \omega_{LO} = \omega_{IF}$. If, however, there is another signal present at the RF input, with a frequency $\omega_{img} = \omega_{RF} - 2\omega_{IF}$, it will *also* mix with ω_{LO} to produce a spectral component at the right IF: $\omega_{LO} - \omega_{img} = \omega_{IF}$. The problem is schematically shown in Figure 5.112. ω_{img} is called the *image frequency*.

The signal at the image frequency will now overlay the downconverted RF signal and in most cases cause unacceptable interference. The most common way to avoid this is to make sure, by appropriate filtering, that no signal is in fact present at ω_{img}.

In upconversion mixers, the image problem also applies. Upconverting a signal at ω_{IF} by means of an LO at a higher frequency ω_{RF} results in two spectral components at $\omega_{\text{LO}} \pm \omega_{\text{IF}}$. Again, one of these components will have to be suppressed, classically by filtering.

However, in modern receiver and transmitter concepts, there is a trend to lower IFs where digital/analogue conversion can be achieved very inexpensively. Consequently, the frequency distance $2\omega_{\text{IF}}$ between the desired signal and the image is getting smaller, requiring very rigid filtering. The necessary filter qualities are rarely possible on chip, which is another significant drawback, as off-chip filters are costly in production and assembly.

A more convenient solution is the use of an *image-rejection* mixer. The image-rejection mixer topology can use any of the fundamental mixer circuits we discussed.

The block diagram shown in Figure 5.113 shows a possible implementation for a downconversion mixer. At the input, two signals shall be present, at the RF and image frequencies:

$$\omega_{\text{RF}} = \omega_{\text{LO}} + \omega_{\text{IF}}, \omega_{\text{img}} = \omega_{\text{LO}} - \omega_{\text{IF}}$$

The signal is split into two signals with equal amplitude, but 90° phase shift (*in quadrature*). We assume that this can be done equally for the RF and the image signals – this is a good assumption because the concept is especially important if RF and image signals are located in close spectral proximity and cannot be separated easily by filtering. For simplicity, we assume also that both RF and image signals are sinusoidal with an amplitude of 1.

The oscillator signal, at ω_{LO}, shall also have an amplitude of 1. It is split into two signals with equal amplitude and phase.

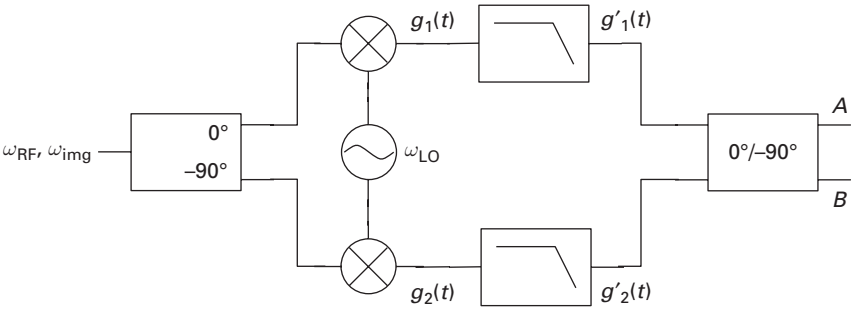

Fig. 5.113 Block diagram of an image reject mixer.

Let us first consider the RF signal only (the image signal is turned off). The signal at the output of the top mixer is

$$g_1(t) = \frac{1}{4}\sin(\omega_{\mathrm{RF}}t)\sin(\omega_{\mathrm{LO}}t)$$
$$= \frac{1}{8}\left\{\cos[(\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})t] - \cos[(\omega_{\mathrm{RF}} + \omega_{\mathrm{LO}})t]\right\}.$$

The high frequency component at $\omega_{\mathrm{RF}} + \omega_{\mathrm{LO}}$ is easily suppressed in the low-pass filter (LP). After the low-pass filter in the upper branch,

$$g_1'(t) = \frac{1}{8}\cos[(\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})t] = \frac{1}{8}\cos(\omega_{\mathrm{IF}}t) = \frac{1}{8}\sin\left(\omega_{\mathrm{IF}}t + \frac{\pi}{2}\right). \tag{5.257}$$

In the lower branch, the RF signal is delayed by $-\pi/2$.

Using $\sin(\omega_{\mathrm{RF}} - \pi/2) = -\cos(\omega_{\mathrm{RF}}t)$, we obtain at the output of the bottom mixer:

$$g_2(t) = -\frac{1}{4}\cos(\omega_{\mathrm{RF}}t)\sin(\omega_{\mathrm{LO}}t)$$
$$= -\frac{1}{8}\left[\sin(\omega_{\mathrm{RF}} + \omega_{\mathrm{LO}})t - \sin(\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})t\right].$$

The high frequency component is removed in the low-pass filter:

$$g_2'(t) = \frac{1}{8}\sin(\omega_{\mathrm{IF}}t). \tag{5.258}$$

The four-port at the right of the block diagram in Figure 5.113 is a 90° *hybrid coupler*. Signals fed into the inputs emerge at outputs A and B with equal amplitude, but the signal entering at the top experiences an additional $-\pi/2$ phase shift at output B, while the signal entering at the bottom experiences an additional phase shift of $-\pi/2$ at output B.

Using Equations (5.257) and (5.258), we then find at A:

$$\frac{1}{8}\left[\sin\left(\omega_{\mathrm{IF}}t + \frac{\pi}{2}\right) + \sin\left(\omega_{\mathrm{IF}}t - \frac{\pi}{2}\right)\right] = 0.$$

At B:

$$\frac{1}{8}\left[\sin(\omega_{\mathrm{IF}}t) + \sin(\omega_{\mathrm{IF}}t)\right] = \frac{1}{4}\sin(\omega_{\mathrm{IF}}t).$$

The IF signal due to the wanted (RF) signal hence only appears at output B.

Let us consider the image frequency, which is below ω_{LO} such that $\omega_{\mathrm{IF}} = \omega_{\mathrm{LO}} - \omega_{\mathrm{img}}$. Using $\sin(-\alpha) = -\sin\alpha$ and $\cos(-\alpha) = \cos\alpha$, we find at the output of the upper low-pass filter:

$$g_1'(t) = -\frac{1}{8}\sin\left(\omega_{\mathrm{IF}}t - \frac{\pi}{2}\right), \tag{5.259}$$

while at the output of the lower low-pass filter:

$$g_2'(t) = -\frac{1}{8}\sin(\omega_{\mathrm{IF}}t). \tag{5.260}$$

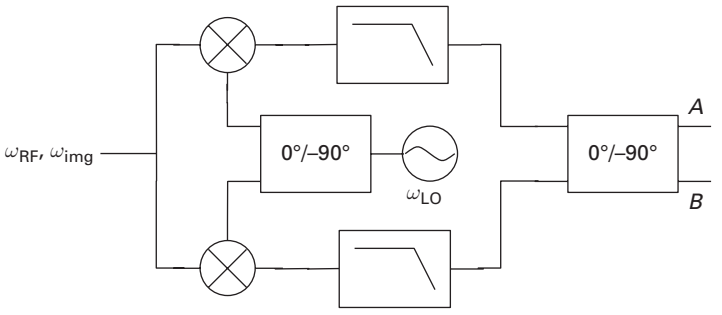

Fig. 5.114 Image reject mixer topology with quadrature LO.

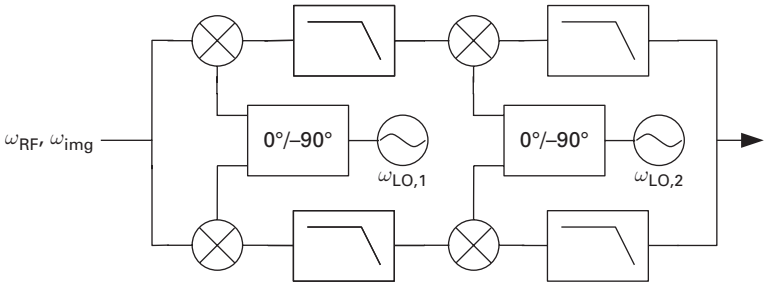

Fig. 5.115 A 'third method' image-rejection mixer topology using two mixing steps.

These signals combine at output A as

$$-\frac{1}{8}\left[\sin\left(\omega_{IF}t-\frac{\pi}{2}\right)+\sin\left(\omega_{IF}t-\frac{\pi}{2}\right)\right]=-\frac{1}{4}\sin\left(\omega_{IF}t-\frac{\pi}{2}\right).$$

At B:

$$-\frac{1}{8}\left[\sin(\omega_{IF}t-\pi)+\sin(\omega_{IF}t)\right]=0.$$

The IF signal due to the image frequency only appears at port A. The circuit in Figure 5.113 hence separates the IF components due to the RF and image signals at the input.

The topology in Figure 5.114 serves the same purpose and was described by Hartley already in 1928 [18]. It has the advantage that oscillators can be constructed so that they directly generate quadrature output signals, which maintain 90° phase shift over a wide frequency band.

The final topology, introduced by Weaver in 1956, eliminates the 90° hybrid at the IF, but uses a second down-conversion step [39]. It is shown in Figure 5.115.

While it requires two additional mixers and low-pass filters plus an extra LO, it eliminates the need for the IF hybrid which, for low IFs, is very difficult to integrate.

5.6.8 Mixer noise figure

The existence of an image frequency also complicates the noise figure assessment of mixers; while the intended signal is present at only one frequency (ω_{RF}), noise present at the image frequency ω_{img} is also converted to the IF and decreases the signal-to-noise ratio at the mixer output.

In principle, noise figure definitions for mixers use Friis' definition of the two-port noise figure (Equation (5.71)); the noise figure is calculated as the quotient of the signal-to-noise ratios at the input and the output of the two-port, assuming that the noise temperature of the input is $T_0 = 290\,\mathrm{K}$.

The issue here is how to calculate the signal-to-noise ratio at the input.

We may consider the signal-to-noise ratio only for the RF frequency. If the signal power is S, the signal-to-noise ratio before the mixer is

$$\mathrm{SNR}_{\mathrm{inp}} = \frac{S}{kT_0\Delta f}, \tag{5.261}$$

where Δf is the measurement bandwidth.

The mixer noise sources are combined in an equivalent noise source $kT_{\mathrm{n}}\Delta f$, placed at the input. T_{n} is the noise temperature of the mixer.

Generally, the mixer's gain at the RF and the image frequency can have different values, which we call G_{RF} and G_{img}, respectively.[8]

The image frequency also contributes a noise power of $kT_0\Delta f$. The signal-to-noise ratio at the output is then

$$\mathrm{SNR}_{\mathrm{out}} = \frac{G_{\mathrm{RF}}S}{k\Delta f[G_{\mathrm{RF}}(T_0 + T_{\mathrm{n}}) + G_{\mathrm{img}}T_0]}. \tag{5.262}$$

The ratio of the expressions (5.261) and (5.262) is called the *single-sideband noise figure*:

$$F_{\mathrm{SSB}} = 1 + \frac{T_{\mathrm{n}}}{T_0} + \frac{G_{\mathrm{img}}}{G_{\mathrm{RF}}} \tag{5.263}$$

We may also simply apply Equation (5.73), developed originally for two-ports, to the mixer. The result is the *IEEE single-sideband noise figure*:

$$F_{\mathrm{SSB,IEEE}} = 1 + \frac{T_{\mathrm{n}}}{T_0} \tag{5.264}$$

The IEEE definition is similar to (5.263), provided that $G_{\mathrm{img}} \ll G_{\mathrm{RF}}$. However, we can also calculate the signal-to-noise ratio before the mixer taking the thermal noise at the image frequency into account. Then,

$$\mathrm{SNR}_{\mathrm{inp}} = \frac{S}{2kT_0\Delta f}. \tag{5.265}$$

[8] Note that all gains are available gains, i.e. power match is assumed for all ports.

Now using (5.265) and (5.262) to calculate the noise figure, we obtain the *double-sideband noise figure*:

$$F_{\mathrm{DSB}} = \frac{F_{\mathrm{SSB}}}{2} = \frac{F_{\mathrm{SSB,IEEE}} + \frac{G_{\mathrm{img}}}{G_{\mathrm{RF}}}}{2}. \tag{5.266}$$

On a logarithmic scale, F_{SSB} is thus 3 dB larger than F_{DSB}.

For the assessment of receiver systems, F_{SSB} is the appropriate entity because we need to compare signal-to-noise ratios at the intended frequency only. However, measurements of mixer noise figure invariably yield F_{DSB}, unless the image frequency is suppressed at the input of the mixer by appropriate filtering. When using mixer noise figure data, this has to be carefully observed.

5.7 Baluns, unbals and hybrids

In the preceding sections, reference was frequently made to mysterious building blocks which convert signals from single-end to differential (two signals of equal amplitude, but 180° out of phase), or splitting a signal into two parts 'in quadrature' (90° out of phase). Due to their importance, they also deserve a brief section of their own.

The term *balun* is a contraction of *balanced to unbalanced*, describing the function this component performs; they convert a balanced (differential) signal, which can be understood as two ground-referenced signals of equal amplitude, but 180° phase difference, to an unbalanced signal, which is a single ground-referenced form. The *unbal* does exactly the opposite – it converts an *unbalanced* signal to a *balanced* (differential) one.

5.7.1 Passive baluns and unbals

Passive baluns and unbals are essentially the same components, operated in different directions. Therefore, it is common to call them by the name 'balun' only, irrespective of the actual role.

A very common balun/unbal at lower RF frequencies is the centre-tapped transformer (Figure 5.116). Due to the symmetric centre tap of the secondary winding, the balanced output voltage is formed by two voltages v_1 and $\bar{v}_1$ of equal amplitude, but opposite phase. At lower frequencies, the centre-tapped transformer balun is frequently constructed as a *toroidal* transformer, using ring-shaped magnetic cores. As the frequency of operation increases, this will rapidly not be possible anymore as the useful frequency

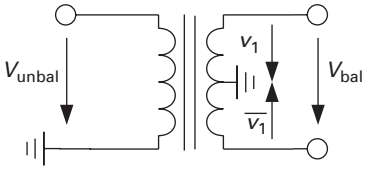

Fig. 5.116 Center-tapped transformer as a balun/unbal.

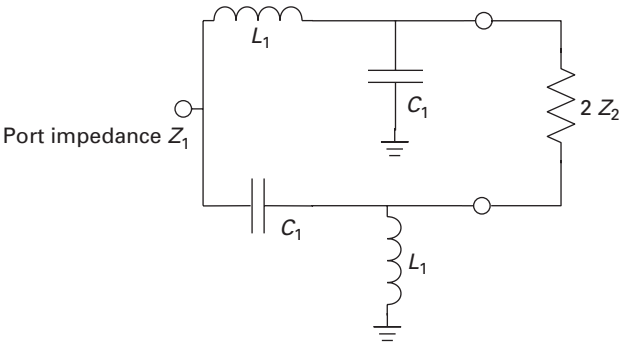

Fig. 5.117 Lumped-element balun circuit.

range of the core material is exceeded and also the parasitic capacitances of the windings become excessive.

A rather straightforward balun implementation is shown in Figure 5.117. It uses a combination of LC low-pass and high-pass filters to provide phase shifts of $-90°$ and $+90°$, respectively, at the design frequency ω_0. It can provide impedance transformation at the same time. Both high-pass and low-pass filters provide the 90° phase shifts if

$$L_1 C_1 = \frac{1}{\omega_0^2}. \tag{5.267}$$

At this frequency, they need to additionally fulfil:

$$\frac{L_1}{C_1} = 2Z_1 Z_2 \tag{5.268}$$

to transform between the single-ended port impedance Z_1 and the differential port impedance $2Z_2$. Solving Equations (5.267) and (5.268) yields

$$L_1 = \frac{\sqrt{Z_1 Z_2}}{\omega_0} \tag{5.269}$$

$$C_1 = \frac{L_1}{2Z_1 Z_2}. \tag{5.270}$$

As all lumped-element transformation circuits, the lumped-element balun will be quite narrow band.

At microwave frequencies, a rather large number of possible implementations of *transmission line* baluns exist. We will restrict our discussion to one example.

A very popular transmission line balun, whose operation is also quite easy to understand, is the *rat-race* coupler structure shown in Figure 5.118. It consists of a transmission line ring with a circumference of 1.5 times the wavelength λ, with four ports arranged as shown in Figure 5.118. If equal power distribution to the coupled ports is desired, the transmission line ring must have a characteristic impedance of $\sqrt{2}Z_0$, where Z_0 is the port impedance. If we feed a signal into port 1, it will split into two partial signals, which travel clockwise and counter-clockwise through the ring. They will interfere constructively at port 3, 2 and 3, while the phase difference is π at port 4, leading to destructive interference there. The signal at port 3 lags port 1 by $\pi/2$, while the

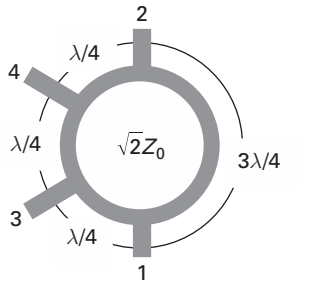

Fig. 5.118 A 180° hybrid ('rat race') coupler structure.

signal at port 2 leads port 1 by $\pi/2$ (well, it lags by $3\pi/2$, but that is the same). Because port 4 is decoupled, and we have equal power distribution, the scattering parameters with respect to port 1 are then:

$$S_{21} = +j\frac{1}{\sqrt{2}}$$
$$S_{31} = -j\frac{1}{\sqrt{2}}$$
$$S_{41} = 0.$$

The phase difference between ports 2 and 3 is π or 180° – a single-ended signal into port 1 is converted into a differential signal between ports 2 and 3. The component is reciprocal, of course – a differential signal applied between ports 2 and 3 will combine into a single-ended signal out of port 1.

The rat-race coupler has another interesting property – a signal inserted into port 4 will split into in-phase components out of ports 2 and 3, while port 1 is isolated. In scattering matrix terms,

$$S_{24} = -j\frac{1}{\sqrt{2}}$$
$$S_{34} = -j\frac{1}{\sqrt{2}}$$
$$S_{14} = 0.$$

5.7.2 Active baluns and unbals

A major disadvantage of the passive baluns discussed so far is their narrow bandwidth of operation, and at lower frequencies their potentially large chip area consumption, due to either the size of the transmission line segments or the size of the necessary reactances.

Active circuits can provide balanced-to-unbalanced and unbalanced-to-balanced-conversions over a wide operational bandwidth, and often in a very small chip area. They are, therefore, frequently used in IC implementations of microwave circuits. Disadvantages are the additional noise due to the active components, potential non-linearities and the added power consumption, which can be considerable if high linearity is required.

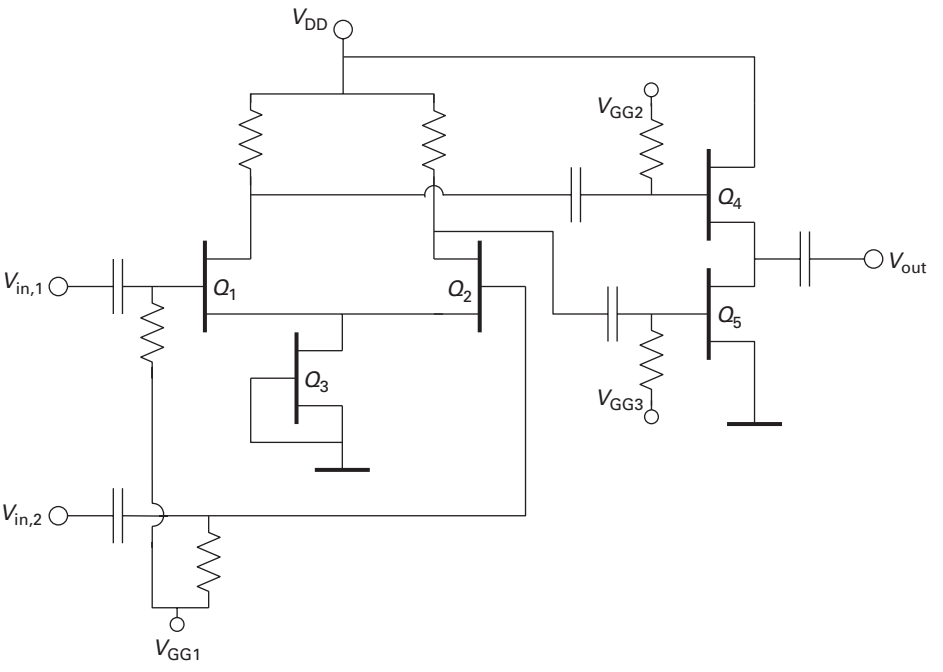

Fig. 5.119 Active balun circuit.

Figure 5.119 shows a typical active balun conversion circuit. Transistors $Q_1 - Q_3$ form a differential amplifier, which is used here to suppress any common-mode components between the input voltages $V_{\text{in},1}$ and $V_{\text{in},2}$. Transistors Q_4 and Q_5 form the balun proper – Q_4 acts in common-drain (source follower) configuration, while Q_5 is in common-source configuration, both working against the common single-ended output.

The simplest active unbal (Figure 5.120), uses a transistor which is simultaneously configured in common-source and common-drain topology. Output 1 is at the source, hence the voltage gain is in good approximation $+1$. Output 2 is at the drain, and the resistors R_1 and R_2 must be chosen such that the voltage gain is -1, leading to the desired balanced output signal V_{out}. Quasistatically, this is very simple – the small-signal output voltage $v_{\text{out},2}$ is

$$v_{\text{out},2} = -R_1 \frac{v_{\text{out},1}}{R_2},$$

so that for equal magnitudes of $v_{\text{out},1}$ and $v_{\text{out},2}$,

$$R_1 = R_2.$$

For higher frequencies, the performance deteriorates, especially due to leakage through the transistor's gate-drain capacitance. The useful range can be extended by cascading a differential amplifier with good common-mode rejection.

Figure 5.121 shows an alternative active unbal, which uses common-source and common-gate amplifiers in parallel. Q_1 is a common-gate topology, with a quasi-static

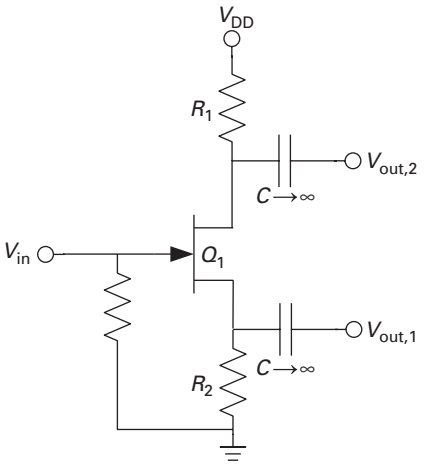

Fig. 5.120 Active unbal using a transistor in common-source/common-drain configuration.

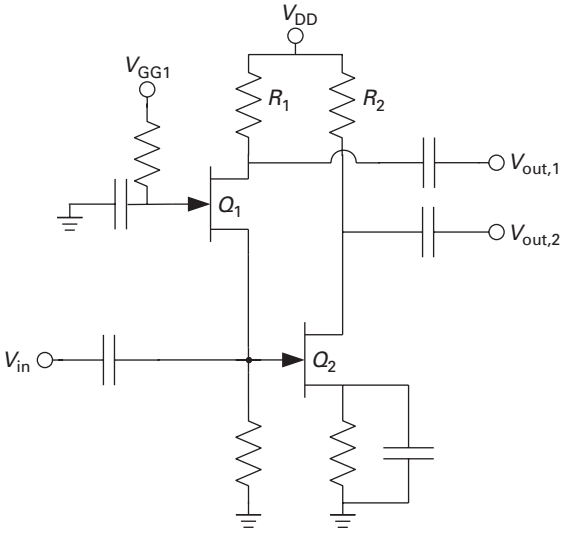

Fig. 5.121 Active unbal using a combination of common-source and common-gate topologies.

small-signal gain of $g_m R_1$,[9] while Q_2 is a common-source amplifier with a voltage gain of $-g_m R_2$. Provided $R_1 = R_2$ and equal transistor transconductances g_m, both output voltages will have the same magnitude, but opposite phase: $V_{out,1} = V_{out,2}$.

Compared to the earlier circuit (Figure 5.120), this circuit has the advantage that the input impedance is significantly reduced, leading to a higher bandwidth, while the output impedances are equal for both ports.

[9] Neglecting the transistor output conductance.

5.7.3 Quadrature generation

The generation of two signals with 90° phase shift is another very frequent task, as we have seen in the discussion of image-rejection mixer topologies.

The simplest generation of quadrature signals is realised using the simple RC network shown in Figure 5.122. The output voltages are

$$V_{\mathrm{Q}} = V_0 \frac{1}{1 + \jmath\omega R_1 C_1} \tag{5.271}$$

$$V_{\mathrm{I}} = V_0 \frac{\jmath\omega R_1 C_1}{1 + \jmath\omega R_1 C_1}. \tag{5.272}$$

At $\omega_0 = 1/(R_1 C_1)$, V_{Q} lags V_0 by 45°, while V_{I} leads V_0 by 45° – V_{Q} and V_{I} are hence in quadrature. The 90° phase difference between V_{Q} and V_{I} is maintained over a wide frequency range, but the amplitudes are equal only at ω_0. Note also that this is a lossy network – at ω_0, $|V_{\mathrm{I}}| = |V_{\mathrm{Q}}| = V_0/\sqrt{2}$.

The polyphase filter family, an example of which is shown in Figure 5.123, uses also passive RC elements to generate quadrature output signals from a differential input

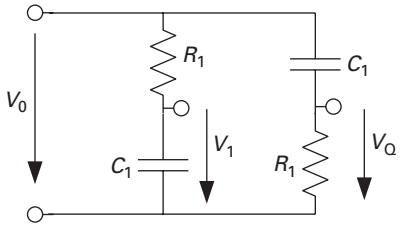

Fig. 5.122 Quadrature generation using a simple RC network.

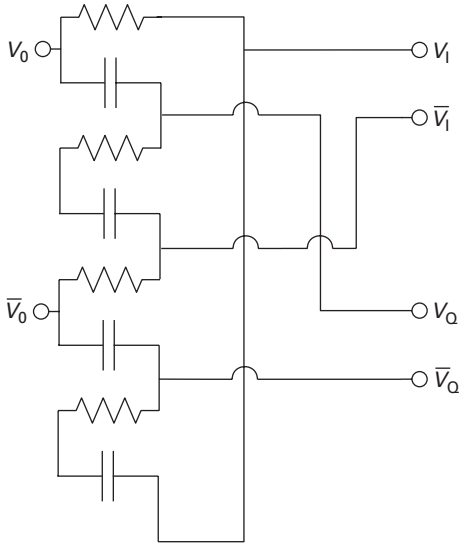

Fig. 5.123 RC polyphase network for differential quadrature generation.

signal. It is very frequently used in fully differential receiver concepts. All RC filters have significant loss and should therefore not be used in noise-critical signal paths – a common use is for quadrature generation from differential LOs, or in quadrature IF combiners at the output of image-rejection mixers such as the one shown in Figure 5.114. An in-depth description of polyphase filter operation can be found in [3].

Another common quadrature generator, this time using transmission lines, is the 90° hybrid, shown in Figure 5.124.

All transmission line segments are electrically a quarter wavelength long. A signal fed into port 1 will emerge at 2 with 90° phase lag, and with 180° at port 3. The signals at 2 and 3 are hence in quadrature. The signal from port 1 will interfere destructively at port 4, which is hence ideally decoupled, provided that 2 and 3 are terminated with the proper impedance Z_0.

Quadrature signals can also be generated digitally where linear operation is not mandated – following an oscillator.

The circuit depicted in Figure 5.125(a) needs a clock frequency at four times the intended output frequency. As state changes occur only on the rising edge of the clock, it

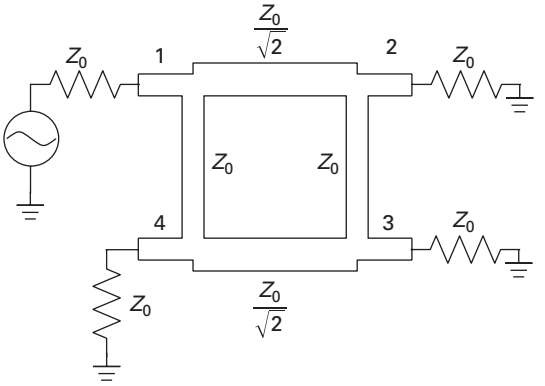

Fig. 5.124 Transmission line 90° hybrid.

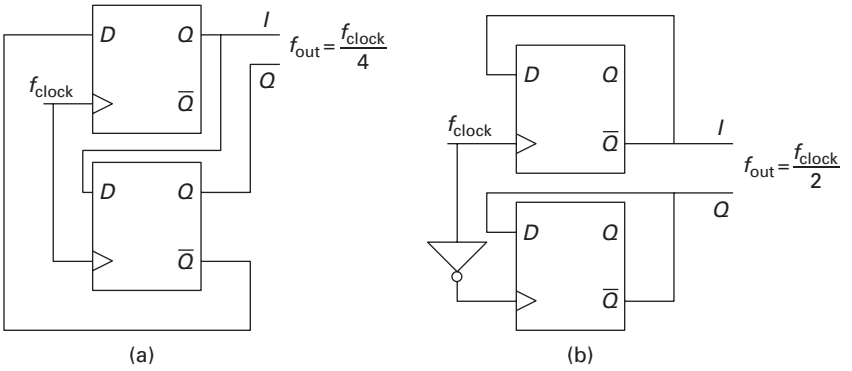

Fig. 5.125 Quadrature signal generation using flipflops.

is insensitive to the clock's duty cycle. However, the very high clock frequency requirement makes it unsuitable for many applications. The circuit in Figure 5.125(b) needs only twice the output frequency as a clock, but as it triggers on both the rising and the falling edges of the clock, it is very sensitive to the clock's duty cycle. The time delay in the inverter may lead to additional problems as the clock frequency increases.

Finally, quadrature clock signals may also be generated in special oscillator circuits. This, however, shall be beyond the scope of this book.

5.8 Problems

(1) Demonstrate that the power delivered to a load of arbitrary impedance can be expressed as the difference in the squared magnitudes of the incident and reflected normalised power waves – see Equation (5.3).

(2) Consider a two-port whose scattering matrix is known. Calculate the power delivered to an arbitrary load Z_L as a function of the available power of the generator, whose source impedance shall be equal to the normalising impedance Z_0.

(3) You have to design a common-source amplifier with a FET technology whose transit frequency f_T is 50 GHz. The load resistance of the amplifier is given to be 100 Ω, the voltage gain shall be $A_V = -10$. Calculate the input capacitance – hint: use the common rule of thumb that $C_{GD} \approx 0.1 \cdot C_{GS}$. You may also neglect g_{DS}.

(4) Consider the circuit in Figure 5.126. The transistors Q_1 and Q_2 shall have a transconductance of $g_m = 20$ mS and a transit frequency $f_T = 50$ GHz. Calculate the input admittance of this circuit.

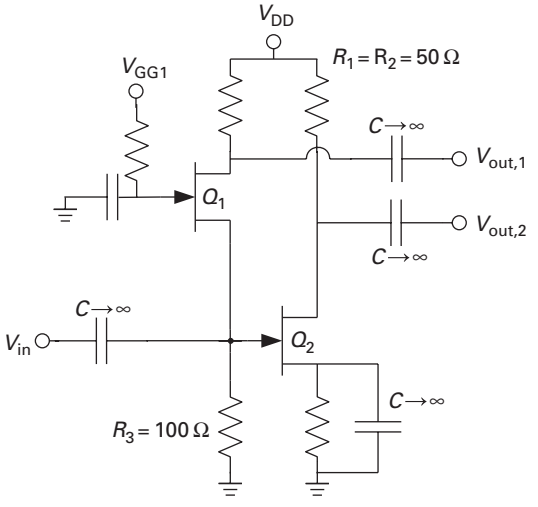

Fig. 5.126 Active unbal circuit for Problem 4.

(5) Suggest a simple one-transistor circuit capable of creating an impedance (referenced to ground) with a real part $\mathrm{Re}\{Z_1\} = -400\,\Omega$ at $f = 10\,\mathrm{GHz}$. The FET technology you are using has an f_T of 100 GHz and the transistor transconductance shall be $g_\mathrm{m} = 50\,\mathrm{mS}$.

(6) The circuit in Figure 5.127 shall be used to realise a wideband voltage amplifier stage. The transistor Q_1 has an $f_\mathrm{T} = 10\,\mathrm{GHz}$ and transconductance of $g_\mathrm{m} = 100\,\mathrm{mS}$ in a bias point $I_\mathrm{D} = 10\,\mathrm{mA}$, $V_\mathrm{GS} = -0.5\,\mathrm{V}$ and V_DD is 10 V. The voltage gain shall be $a_\mathrm{V} = -10$.
   (a) Calculate R_3.
   (b) What is the optimum choice for C_1 with respect to maximum bandwidth, if the transistor can be modelled using the simple equivalent circuit in Figure 5.20?
   (c) Calculate R_2.
   (d) What is V_DS under these circumstances?

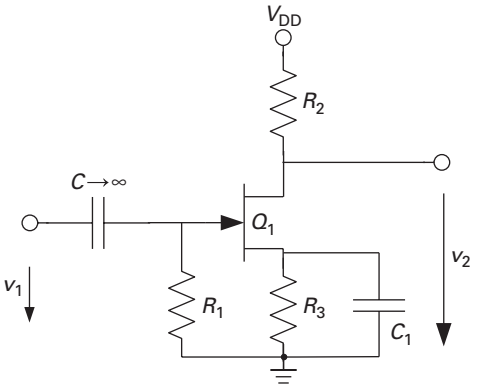

Fig. 5.127 Simple wideband amplifier.

(7) Consider the circuit in Figure 5.32. Given that all transistors have the same size, why do Q_1 and Q_2 have the same collector current, provided that the current gain $\beta \gg 1$? Let now $I_\mathrm{C} = 5\,\mathrm{mA}$ for all transistors, $\beta = 100$, $f_\mathrm{T} = 50\,\mathrm{GHz}$, $R_\mathrm{L} = 100\,\Omega$. Draw the small-signal equivalent circuit and calculate the input impedance and the voltage gain.

(8) Discuss three different ways of eliminating the feedback capacitance in amplifiers.

(9) The impedance seen in the input of a common-source amplifier stage at $f = 10\,\mathrm{GHz}$ is $Z_1 = (5 - \jmath 159)\,\Omega$. The output admittance of the preceding stage is $Y_2 = 10 + \jmath 3\,\mathrm{mS}$. Suggest a suitable matching network so that the available power is transferred from the first to the second stage. Hint: the Smith chart is very helpful here.

(10) You have the task to design a simple distributed amplifier. The transistors to use have a transconductance $g_\mathrm{m} = 400\,\mathrm{mS}$ and $f_\mathrm{T} = 70\,\mathrm{GHz}$. The gain cell shall be a simple common-source stage. The gate width of each transistor is $W_\mathrm{G} = 100\,\mu\mathrm{m}$. For the drain-gate capacitance, use $C_\mathrm{DG} = 0.1 C_\mathrm{GS}$.

The unloaded gate line has a characteristic impedance $Z_u = 80\Omega$ and an inductance per unit length of $L' = 0.8\,\text{nH}\,\text{mm}^{-1}$.

Calculate the necessary length of the gate line segments so that the characteristic impedance of the loaded gate line is 50Ω. The drain line shall equally have this impedance.

(11) In the schematic in Figure 5.63, identify the function of each of the transistors shown. What is the purpose of the RC combination attached to the emitters of two of the transistors?

(12) A communications system can tolerate a minimum third-order intermodulation distance of 60 dB; the maximum input power is −20 dBm. What is the necessary input-referred third-order intercept point?

(13) Explain why a high resonator quality factor is especially important in oscillator using active devices with little low-frequency noise, such as Si/SiGe HBTs.

(14) Show how the Hartley image reject topology (Figure 5.114) achieves separation of the signal and image frequencies to output A and B.

References

[1] Abele P., Kallfass I., Zeuner M. *et al.* (2003). 32 GHz MMIC distributed amplifier based on n-channel SiGe MODFETs. *Electron. Lett. 39*, 1448–1449.

[2] Battjes C. R. (1980). Monolithic wideband amplifier. USA Patent 4, 236, 119.

[3] Behbahani F., Kishigami Y., Leete J., Abidi A. (2001). CMOS mixers and polyphase filters for large image rejection. *IEEE J. Solid-State Circ. 36*, 873–887.

[4] Beyer J., Prasad S. N., Becker R., Nordman J., Hohenwarter G. (1984). MESFET distributed amplifier design guidelines. *IEEE Trans. Microw. Theory Tech. MTT-32*, 268–275.

[5] Chartier S., Schleicher B., Korndörfer F., Glisic S., Fischer G., Schumacher H. (2007). A fully integrated fully differential low-noise amplifier for short range automotive radar using a SiGe:C BiCMOS technology. In *Proc. 2nd European Microwave IC Conference (EuMIC), Munich, Germany, 8–10 October, 2007*.

[6] Chartier S., Sönmez E., Schumacher H. (2006). Millimeter-wave amplifiers using a 0.8 µm Si/SiGe HBT technology. In *Proc. Silicon Monolithic Integrated Circuits in RF Systems, San Diego, CA, 18–20 January 2006*. IEEE.

[7] Darlington S. (1953). Semiconductor translating device. USA Patent 2, 663, 806.

[8] Ellinger F. (2007). *Radio Frequency Integrated Circuits and Technologies*, 1st edn. Springer.

[9] Fleming J. A. (1905). Instrument for converting alternating electric currents into continuous currents. USA Patent 803, 684.

[10] Forest L. D. (1908). Space telegraphy. USA Patent 879, 532.

[11] Friis H. T. (1944). Noise figure of radio receivers. *Proc. IRE 32*, 7, 419–422.

[12] Gilbert B. (1968). A precise four-quadrant multiplier with subnanosecond response. *IEEE J. Solid-State Circ. 3*, 365–373.

[13] Gilbert B. (1997). The Micromixer: a highly linear variant of the Gilbert mixer using a bisymmetric class-AB input stage. *IEEE J. Solid-State Circ. 32*, 1412–1423.

[14] Gonzalez G. (1997). *Microwave Transistor Amplifiers, Analysis and Design*. Prentice Hall.

[15] Gupta M. S. (1992). Power gain in feedback amplifiers, a classic revisited. *IEEE Trans. Microw. Theory Tech. MTT-40*, 5 (May), 864–879.
[16] Häfele M., Trasser A., Beilenhoff K., Schumacher H. (2005). A GaAs distributed amplifier with an output voltage of 8.5 V_{pp} for 40 Gb/s modulators. *Proc. 13th GAAS Symposium, Paris, France, 3–4 October 2005,* 345–348.
[17] Hagen J. B. (1996). *Radio-Frequency Electronics*, 1st edn. Cambridge University Press.
[18] Hartley R. V. L. (1928). Modulation system. USA Patent 1, 666, 206.
[19] Haus H., Adler R. (1958). Optimum noise performance of linear amplifiers. *Proc. IRE 46*, 1519–1533.
[20] Hoffmann M. (1997). *Hochfrequenztechnik*, 1st edn. Springer Berlin Heidelberg.
[21] Hunt F., Hickman R. (1939). On electronic voltage stabilizers. *Rev. Sci. Instrum. 10*, 1 (January), 6–21.
[22] Kline R. (1993). Harold Black and the negative-feedback amplifier. *IEEE Contr. Syst. Mag. 13*, 4, 82–85.
[23] Ku W. H. (1966). Unilateral gain and stability criteria of active two ports in terms of scattering parameters. *Proc. IEEE 54*, 11 (November), 1617–1618.
[24] Kurokawa K. (1965). Power waves and the scattering matrix. *IEEE Trans. Microw. Theory Tech. MTT-13*, 3 (March), 194–202.
[25] Lee T. H. (1997). *The Design of CMOS Radio-Frequency Integrated Circuits*, 1st edn. Cambridge University Press, 181.
[26] Leeson D. (1966). A simple model for feedback oscillator noise spectrum. *Proc. IEEE 54*, 329–330.
[27] Maas S. A. (1987). A GaAs MESFET mixer with very low intermodulation. *IEEE Trans. Microw. Theory Tech. MTT-35*, 425–429.
[28] Mason S. J. (1954). Power gain in feedback amplifiers. *Trans. IRE, Prof. Group on Circ. Theory CT-1*, 2 (June), 20–25.
[29] Miller J. (1920). Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit. *Scientific Papers of the Bureau of Standards 15* (351), 367–385.
[30] Percival W. S. (1937). Thermionic valve circuits. Britain Patent 460, 562.
[31] Rothe H., Dahlke W. (1955). Theorie rauschender Vierpole. *AEÜ 9*, 117–121.
[32] Rothe H., Dahlke W. (1956). Theory of noisy fourpoles. *Proc. IRE 44*, 811–815.
[33] Schick C., Feger T., Sönmez E., Schad K., Trasser A., Schumacher H. (2005). Broadband Si/SiGe HBT amplifier concepts for 40 GBit/s fibreoptic communication systems. In *Proc. 35th European Microwave Conf., Paris, France, October 2005*.
[34] Schmitt O. (1937). A simple differential amplifier. *Rev. Sci. Inst. 8*, 4 (April), 126–127.
[35] Schumacher H., Abele P., Sönmez E., Schad K., Trasser A. (2003). Low-cost circuit solutions for micro- and millimeter-wave systems using commercially available SiGe technologies. In *Proc. 1st International SiGe Technology and Device Meeting, Nagoya, Japan, 14–17 January 2003*.
[36] Smith P. H. (1931). Transmission line calculator. *Electronics 12*, 1 (January), 29–31.
[37] Smith P. H. (1944). An improved transmission line calculator. *Electronics 17*, 1 (January), 130.
[38] Sokal N., Sokal A. (1975). Class E – a new class of high-efficiency tuned single-ended switching power amplifiers. *IEEE J. Solid-State Circ. 10*, 3 (June), 168–176.
[39] Weaver D. (1956). A third method of generation and detection of single-sideband signals. *Proc. IRE 44*, 1703–1705.