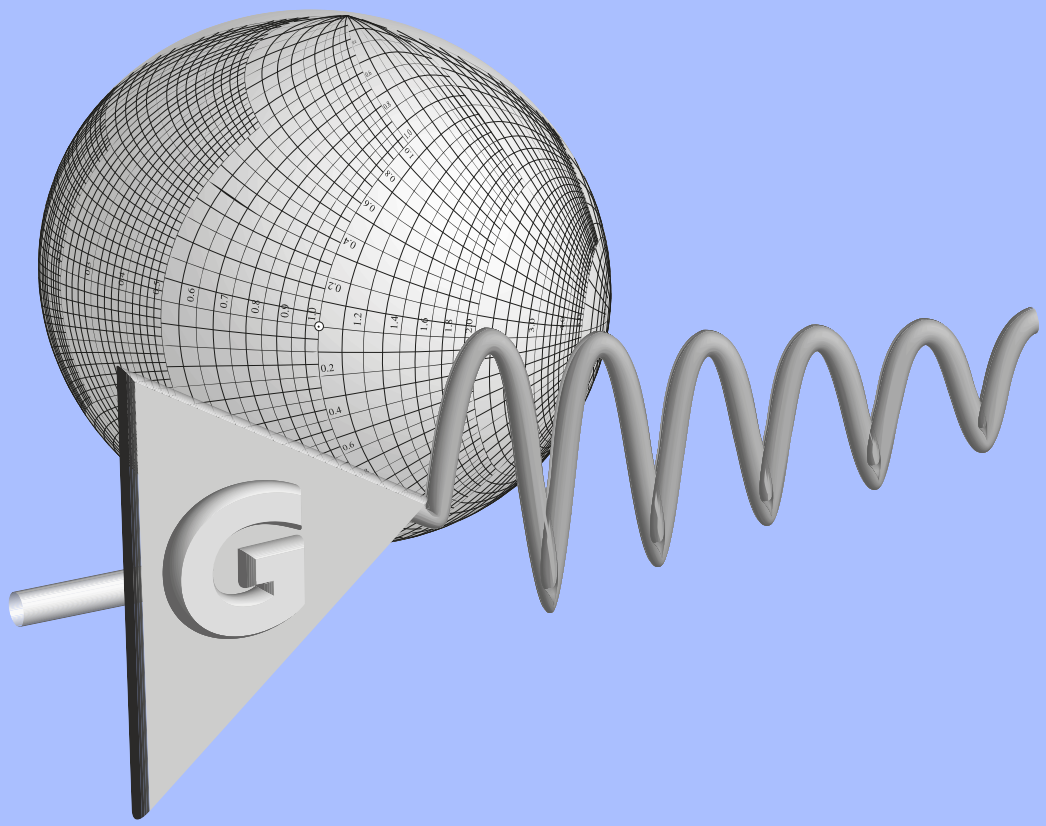
Radio Electronics

Sundström, Lars; Jönsson, Göran; Börjeson, Henrik

2014

Radio Electronics

L. Sundström
G. Jönsson
H. Börjeson

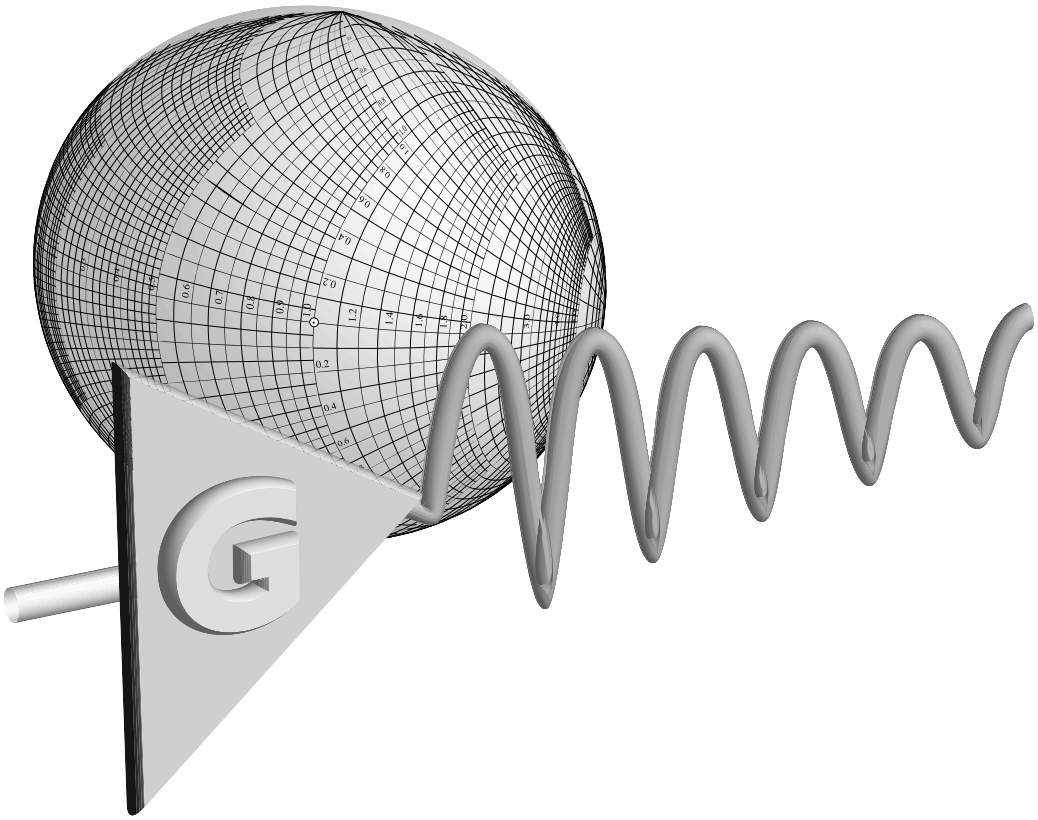


Electrical and Information Technology
Box 118
SE-221 00 Lund
Sweden

LUND UNIVERSITY

Radio Electronics

L. Sundström
G. Jönsson
H. Börjeson



Electrical and Information Technology
Box 118
SE-221 00 Lund
Sweden

LUND UNIVERSITY

Contents

Chapter 1

Resonant Circuits

In radio frequency (RF) applications reactive components, such as capacitors and inductors, are often combined with resistors to form *RLC* networks. They are of great value as they can be used to match impedances (important for efficient power transfer, for example), cancel transistor parasitics to provide high gain at high frequencies and filter out unwanted signals.

Furthermore, passive RF components can in general not be treated as ideal, but can reliably be modelled by *RLC* networks. The impedance of such a circuit is a complex function of frequency, and has therefore normally both a resistive (real) and reactive (imaginary) part. The condition where the reactive part vanishes and the impedance is purely resistive is called *resonance*. The frequency (or frequencies) where this occurs is called the *resonance frequency*. A circuit with one or more resonance frequencies is named a *resonant circuit*. The word resonance refers to the ability of reactive components to store energy. That is, energy can bounce back and forth, or resonate, between the magnetic field in the inductor and the electrical field in the capacitor.

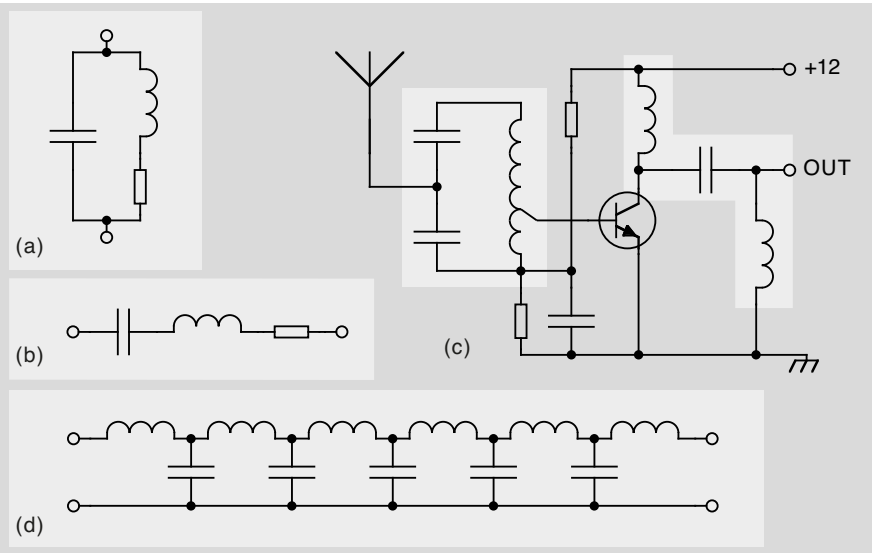


Figure 1.1 Example of resonant circuits: (a) parallel LC, (b) series LC, (c) receiver input stage and (d) multielement low-pass filter.

The magnitude of the resonant circuit impedance shows a more or less sharp maximum or minimum at the resonance frequency. The *bandwidth* or *selectivity* is often defined in terms of the width of this peak or notch.

Another important property is the loss of energy which is denoted by the *quality factor* or simply Q. As it will be shown in the succeeding sections the Q is a useful quantity that can be used to compute bandwidth, impedance transformation, voltage/current transformation among many other utilities.

This chapter gives a short briefing about the most important terms concerning resonant circuits that are necessary to understand in high-frequency design. The fundamental theory of network theory is covered in [1]. A more comprehensive discussion is found in [2] and [3].

1.1 Basic Terms

1.1.1 Series and Parallel Resonance

The model of the resonant circuits can be derived in two basic forms, the serial or the parallel circuit. Although most of the real networks are combinations of series and parallel connections, it is often preferable to characterise the circuit by a clean serial or parallel equivalent circuit. Useful methods for conversion between the basic forms will be described in section 1.2.3.

The series network is recognised when the current path I_{in} forms a path through all the elements in contrast to the parallel network where an equal voltage V_{in} is found on all the elements.

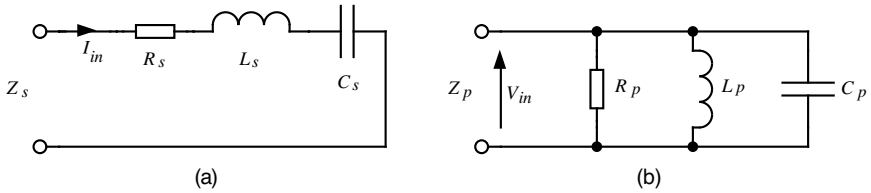


Figure 1.2 Basic resonant circuits, (a) series and (b) parallel network.

The serial form implies that impedance terms are the most convenient way to analyse the serial resonance whereas it is more handy to analyse the parallel resonance in admittance terms.

The impedance and admittance of the circuits in figure 1.2 are simply:

$$Z_s(\omega) = R_s + jX_s = R_s + j\left(\omega L_s - \frac{1}{\omega C_s}\right) \tag{1.1}$$

$$Y_p(\omega) = \frac{1}{Z_p(\omega)} = G_p + jB_p = \frac{1}{R_p} + j\left(\omega C_p - \frac{1}{\omega L_p}\right) \tag{1.2}$$

The circuits will be examined in the succeeding section, but first some other parameters have to be explained.

1.1.2 Resonance Frequency

From inspection of the equations (1.1) and (1.2) it is clearly seen that impedance respectively admittance quantities goes to infinity both at DC and at infinitely high frequency. What divides "low" from "high" is the frequency at which the inductive and capacitive parts cancel. Known as the resonant frequency, this is given by

$$\begin{aligned} \left(\omega L_s - \frac{1}{\omega C_s}\right) = 0 \quad \text{or} \quad \left(\omega C_p - \frac{1}{\omega L_p}\right) = 0 \\ \Rightarrow \omega_0 = \frac{1}{\sqrt{LC}} \quad \text{or} \quad f_0 = \frac{1}{2\pi\sqrt{LC}} \end{aligned} \tag{1.3}$$

To say that the reactive terms cancel is certainly true, but a little careless. As it will be shown shortly, the individual voltages or currents in the reactive elements can be surprisingly large, although they cancel each other as far as the external world is concerned. It will also be shown that these augmented quantities is a sign that an impedance transformation has taken place. However, to explore these behaviours entirely and describe them in the most generally useful way, the quality factor has to be introduced.

1.1.3 Quality Factor, Q

Contrary to resistors perfectly ideal reactive elements can not, according to the basic theory, cause any loss of power. Seeing that impedance of most of the components and circuits contains both a resistive and a reactive part the loss of energy has to be regarded. This quality is described by a parameter named the *quality factor*, or just Q.

There are numerous definitions of Q, but the most fundamental one is

$$Q \equiv 2\pi \cdot \frac{\text{maximum instantaneous energy stored in the network}}{\text{energy dissipated per cycle}} \tag{1.4}$$

Note that Q is dimensionless, and that it is proportional to the ratio of energy stored to the average power dissipated per unit time. The definition is fundamental because it does not care about what stores or dissipates the energy. Consequently it applies perfectly well even to distributed systems, such as transmission lines, where individual inductances, capacitances and resistances hardly can be identified. There is no restriction for Q to solely be used in electrical circuits, it can excellently be applied to mechanical systems etc. It should also be clear that the notion of Q is relevant both to resonant and nonresonant systems. Therefore it is appropriate to characterise an RC circuit, or even a single component, by the Q.

The diagram in figure 1.3 gives a sense of the magnitude of Q for several, both electrical and non electrical systems. It is clearly seen that inductors tend to be significantly lossier than capacitors, which is useful to keep in mind when the design work is about to start.

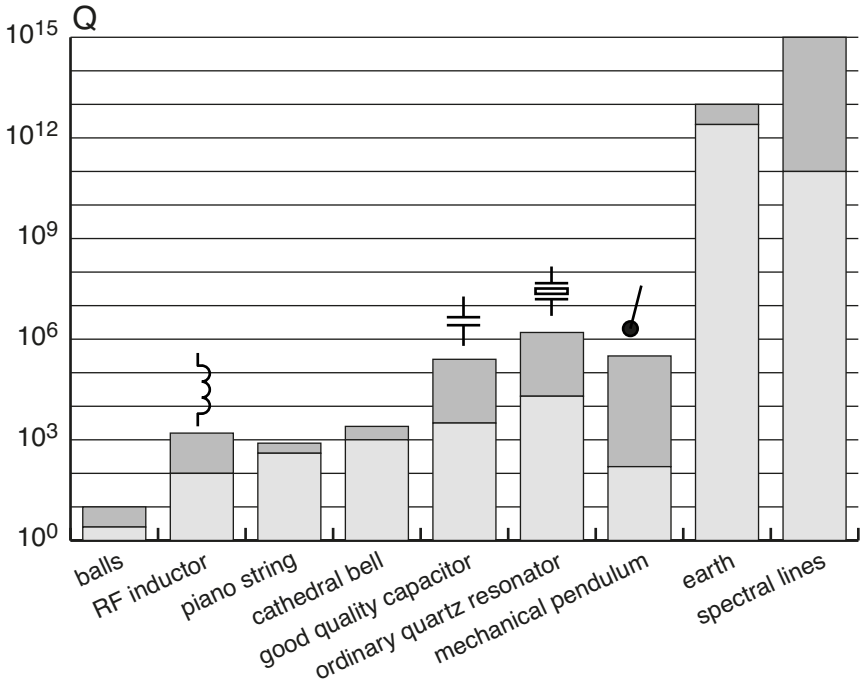


Figure 1.3 The order of Q shown for some various examples.

To get some more useful expressions the definition is used to derive the Q of the parallel circuit in figure 1.2b at resonance. The resonant frequency is denoted $\omega_0 = 2\pi f_0$, the cycle time $T_0 = 1/f_0$ and the peak voltage across the network is denoted U_{pk} .

Recall that energy in such a network bounces back and forth between the inductor and the capacitor, with a constant sum at resonance. As a consequence, the peak energy stored in either the capacitor or inductor is equal to the total energy stored in the network at any given time. Since the peak capacitor voltage is known, it is convenient to use it to compute the network energy:

$$E_{tot} = \frac{1}{2} C_p U_{pk}^2 \tag{1.5}$$

The next step is to derive the amount of energy dissipated per cycle which is equal to the dissipated power multiplied by the cycle time. Again, this is easy to carry out, since the network degenerates to a simple resistance at resonance.

$$E_{diss} = \left(\frac{U_{pk}}{\sqrt{2}}\right)^2 \cdot \frac{1}{R_p} \cdot T_0 \tag{1.6}$$

Finally the Q of the network can be expressed:

$$Q = 2\pi \frac{E_{tot}}{E_{diss}} = \frac{2\pi}{T_o} \cdot \frac{\frac{1}{2} C_p U_{pk}^2}{\frac{1}{2} U_{pk}^2 \cdot \frac{1}{R_p}} = \omega_0 C_p \cdot R_p = \frac{R_p}{\omega_0 L_p} \tag{1.7}$$

It is easy to check the validity of the derived expression. As the parallel resistance goes to infinity, Q does as well. This behaviour seems reasonable since, in the limit of infinite resistance, the network degenerates to a pure LC system. With only purely reactive elements in the network, there is no way for energy to dissipate, and Q should go to infinity, just as the equation says it should.

If ω_0 is substituted in equation (1.7) the expression for Q will be

$$Q = \frac{R_p}{\omega_0 L_p} = \frac{R_p}{\sqrt{L_p / C_p}} \tag{1.8}$$

The quantity $\sqrt{L/C}$ has the dimension of resistance, and is sometimes called the *characteristic impedance* Z_0 of the network. This term is usually applied to transmission lines, but has a certain importance even in lumped networks. It is important because it is equal to the magnitude of the capacitive and inductive reactances at resonance, as it is easily shown:

$$Z_0 = |Z_C| = |Z_L| = \omega_0 L = \frac{L}{\sqrt{LC}} = \sqrt{\frac{L}{C}} \tag{1.9}$$

This quantity will frequently be used in Chapter 2 where the characteristic impedance of a transmission line is given by the same expression, but L and C are interpreted as the inductance and capacitance per unit length.

1.1.4 Bandwidth

The behaviour of the network has been discussed in the previous sections at frequencies far from resonance as well as exactly at the resonance frequency. In this section the behaviour of the resonant circuit will be explored at frequencies slightly shifted from resonance.

The parallel circuit will be used to examine the admittance when the frequency is displaced $\Delta\omega$ from the resonance. The first step is to rewrite equation (1.2) as

$$Y(\omega) = G + j\left(\omega C - \frac{1}{\omega L}\right) = G + \frac{j}{\omega L}(\omega^2 LC - 1) \tag{1.10}$$

Then let $\omega = \omega_0 + \Delta\omega$ and $\omega_0 = 1/(\sqrt{LC})$ will give

$$\begin{aligned} Y(\omega) &= G + \frac{j}{\omega L}\left(\left(\frac{1}{LC} + 2\Delta\omega\omega_0 + (\Delta\omega)^2\right)LC - 1\right) \\ &= G + \frac{j}{\omega L}(2\Delta\omega\omega_0 + (\Delta\omega)^2)LC \end{aligned} \tag{1.11}$$

For values of $\Delta\omega$ that are small relative to ω_0 this expression simplifies to

$$Y(\Delta\omega) \approx G + j2\Delta\omega C \tag{1.12}$$

Thus, this admittance behaviour is exactly the same as that of a resistor of value $G = 1/R$ in parallel with a capacitor of value $2C$, except with $\Delta\omega$ replacing ω. Hence, the shape of the admittance curve for small positive displacements above ω_0 is the same as that of a parallel RC network. The single-sided bandwidth may therefore be defined as simply $1/(2RC)$ by analogy with the RC case. From symmetry of the admittance function, the shape for displacements below ω_0 will be the mirror image of that above resonance, so that the total -3 dB bandwidth, see figure 1.4, is just

$$B_{3dB} = \omega_2 - \omega_1 = \frac{1}{RC} \tag{1.13}$$

If the bandwidth is normalised to the resonant frequency, a highly useful expression is derived:

$$\frac{B_{3dB}}{\omega_0} = \frac{1}{R\omega_o C} = \frac{1}{Q} \tag{1.14}$$

The fractional bandwidth is simply $1/Q$. That is, for a given resonant frequency higher Q implies narrower bandwidth.

Keep in mind that either frequency or angular frequency may be used to specify the bandwidth:

$$B_{3dB}[\text{Hz}] = \frac{f_0}{Q} \qquad \text{or} \quad B_{3dB}[\text{rad/s}] = \frac{\omega_0}{Q} \tag{1.15}$$

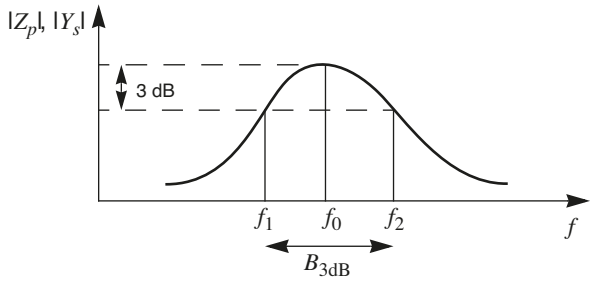


Figure 1.4 Definition of the 3 dB bandwidth.

The use of the ratio 3 dB is a widely used convention to define the bandwidth of resonant circuits. 3 dB means that the quantity of interest has changed to half the value relative the maximum (due to the identity $P = V^2/R = RI^2$, the ratio is $1/\sqrt{2}$ when currents or voltages are calculated). The 3 dB bandwidth is sometimes denoted *half-power bandwidth* and consequently f_1 and f_2 are called *half-power frequencies*.

However, in many applications, such as filters, it may be suitable to choose different ratios than 3 dB.

1.2 Serial and Parallel Models

The aim of this section is to place together the behaviour and the most important expressions of the two basic ways to model a resonant circuit. In the successive text it is assumed that all of the components are ideal. All losses are represented by the resistor.

1.2.1 Series Resonance

In the pure series model all of the components are connected in such a way that the current I_{in} flows through all the elements in a single path. The index "*s*" at each component denotes series connection.

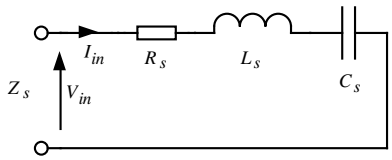


Figure 1.5 Series *RLC* circuit.

The input impedance of the series circuit is

$$\begin{aligned} Z_s(\omega) &= R_s + jX_s = R_s + j\left(\omega L_s - \frac{1}{\omega C_s}\right) \\ &= R_s\left[1 + jQ_s\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right] \end{aligned} \tag{1.16}$$

where $\omega_0 = 1/\sqrt{L_s C_s}$ is the resonant frequency. A very distinct attribute is that the serial resonance always shows *minimum impedance*, with a value R_s, at the resonance.

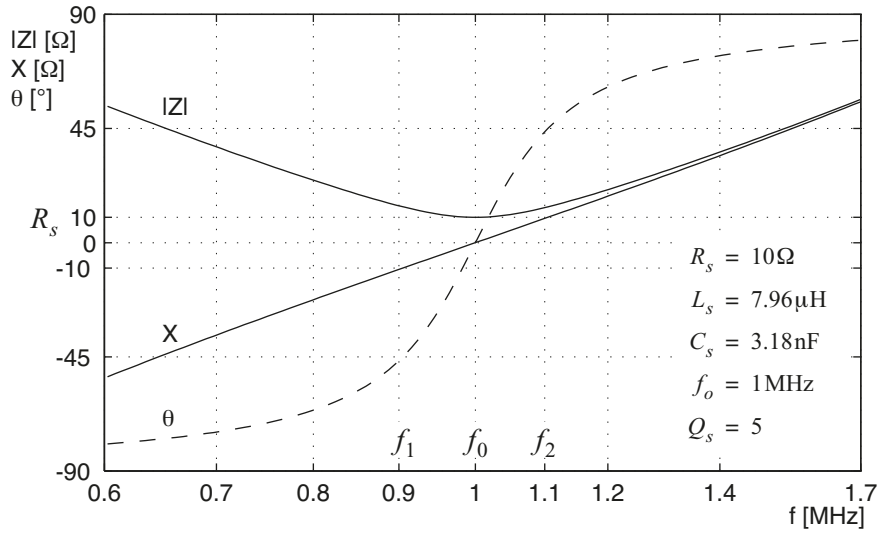


Figure 1.6 Frequency response of the series *RLC* circuit.

It can be noticed that at the half-power frequencies, f_1 and f_2, the magnitude of the reactance is equal to the resistance ($|X| = R$), the magnitude of the impedance $|Z| = \sqrt{2}R$, and the phase angle of the impedance $\theta = \pm 45°$.

The corresponding admittance, even if it is rarely needed, may be written

$$\begin{aligned} Y_s(\omega) &= \frac{1}{Z_s(\omega)} = \frac{R_s}{R_s^2 + X_s^2} - j\frac{X_s}{R_s^2 + X_s^2} \\ &= \frac{1}{R_s}\left(\frac{1}{1+Q^2}\right) - j\frac{1}{X_s}\left(\frac{Q^2}{1+Q^2}\right) \end{aligned} \tag{1.17}$$

The quality factor is given by

$$Q_s = \frac{\omega_0 L_s}{R_s} = \frac{1}{R_s \cdot \omega_0 C_s} = \frac{f_0}{f_2 - f_1} = \frac{f_0}{B_{3dB}} \tag{1.18}$$

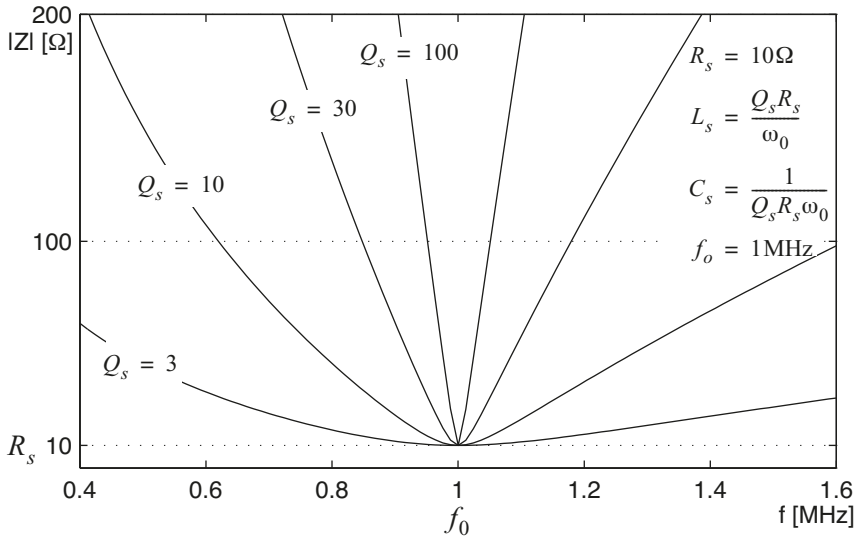


Figure 1.7 Plots of $Z(\omega)$ versus f for a series resonant circuit with several values of Q_s.

The voltage at the capacitor and inductor can differ significantly from the network input voltage. At the resonance frequency the network input current is $I_{in} = U_{in}/R_s$. Since $X_L = -X_C$ at resonance, the inductor and capacitor voltages will be equal in magnitude and opposite in phase.

$$|V_I| = |V_C| = Z|I| = \omega_0 L_s \frac{U_{in}}{R_s} = Q_s U_{in} \tag{1.19}$$

That is, the voltage across the reactive elements is Q times as large as the overall network voltage. Hence, if $Q_s = 200$ and the network is driven at resonance with an one-volt voltage source, the voltage at the inductor and the capacitor will each reach 200 volts. This is a useful experience to keep in mind when choosing components.

1.2.2 Parallel Resonance

In a pure parallel model components are connected in such a way that an equal voltage V_{in} is found on all the elements. The index "p" at each variable denotes parallel connection.

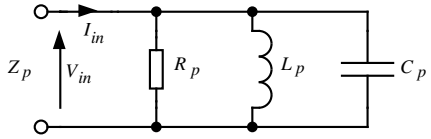


Figure 1.8 Parallel *RLC* circuit.

The input admittance of the parallel circuit is

$$\begin{aligned} Y_p(\omega) &= G_p + jB_p = \frac{1}{R_p} + j\left(\omega C_p - \frac{1}{\omega L_p}\right) \\ &= \frac{1}{R_p}\left[1 + jQ_p\left(\frac{\omega}{\omega_0} - \frac{\omega_0}{\omega}\right)\right] \end{aligned} \tag{1.20}$$

where $\omega_0 = 1/\sqrt{L_p C_p}$ is the resonant frequency. The expression is similar to equation (1.16) except that all admittance quantities are substituted by impedance quantities. In order to make a comparison between the serial and parallel circuits the equivalent impedance is derived and plotted:

$$\begin{aligned} Z_p(\omega) &= \frac{1}{Y_p(\omega)} = \frac{G_p}{G_p^2 + B_p^2} - j\frac{B_p}{G_p^2 + B_p^2} \\ &= R_p\left(\frac{1}{1+Q^2}\right) - j\frac{1}{B_p}\left(\frac{Q^2}{1+Q^2}\right) \end{aligned} \tag{1.21}$$

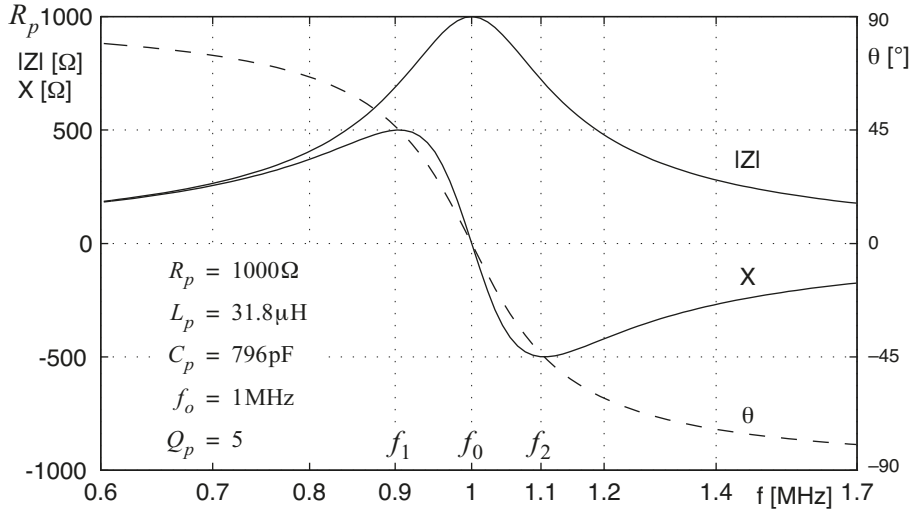


Figure 1.9 Frequency response of the parallel *RLC* circuit.

The equation for Q involves the same terms as for the serial case, but in reciprocal form:

$$Q_p = \frac{R_P}{\omega_0 L_p} = R_p \omega_0 C_p = \frac{f_0}{f_2 - f_1} = \frac{f_0}{B_{3dB}} \tag{1.22}$$

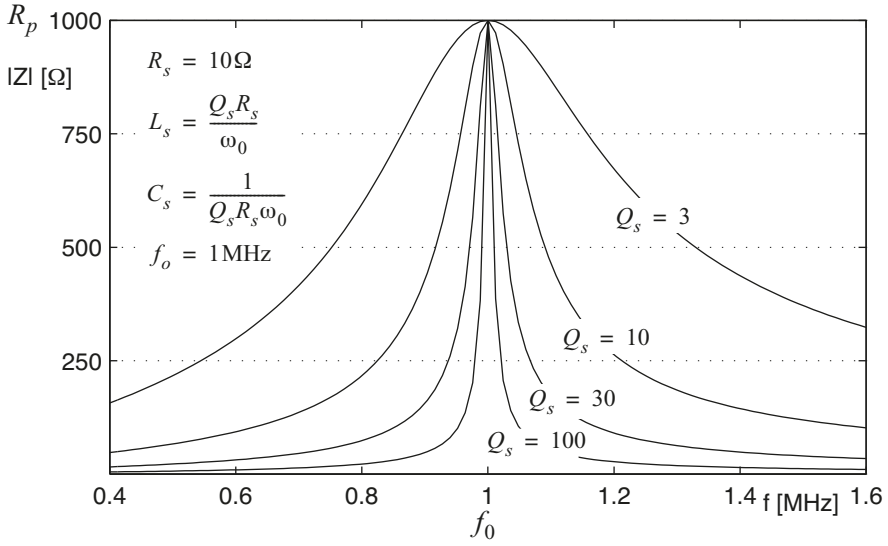


Figure 1.10 Plots of $Z(\omega)$ versus f for a parallel resonant circuit with several values of Q_p.

The inductive and capacitive branch currents at resonance can depart significantly from the network input current. At resonance the voltage across the network is $I_{in}R_p$. Since the inductive and capacitive reactances are equal in magnitude at the resonance frequency, the inductive and capacitive branch currents will be equal in magnitude:

$$|I_L| = |I_C| = \frac{|V|}{Z} = \frac{I_{in}R_p}{\omega_0 L_p} = Q|I_{in}| \tag{1.23}$$

Accordingly the current flowing in the reactive branches is Q times as large as the net current. Hence, if $Q_p = 200$ and the network is driven at resonance with a one-ampere source, that is one-ampere will flow through the resistor, but 200A will flow through the inductor and the capacitor (until it vaporise).

Again, we have seen that it is of utmost importance to consider the maximum voltages and current that can arise in high-Q resonant circuits, particularly in high-power applications.

1.2.3 Series-to-Parallel Conversion

In practice purely series or parallel resonant circuits rarely exist. Therefore it is essential to master conversion between the serial and parallel shape. Consider, for example, the circuit sketched in figure 1.11. As has been indicated

previously, inductors tend to be significantly lossier than capacitors. Consequently the model shown in the figure is often a more realistic approximation to typical parallel *RLC* circuits.

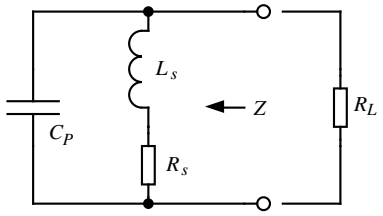


Figure 1.11 A practical parallel *LC* circuit with series resistance in the inductive branch.

The trouble starts when an external load, R_L, is connected to the circuit. For example, without any ingenious tools the only way to calculate the Q is through lengthy algebraic $j\omega$ -calculations.

However, since the purely parallel RLC network has been analysed in detail, it would have been a wastage not to re-use as much as possible of this work. The idea is to convert the series *LR* branch into a parallel equivalent. Clearly, such a substitution cannot be valid in general, but over a suitably restricted frequency range around the resonance the precision is fairly reasonable.

To show this formally, the impedances of the series and parallel LR sections in figure 1.12 are equated.

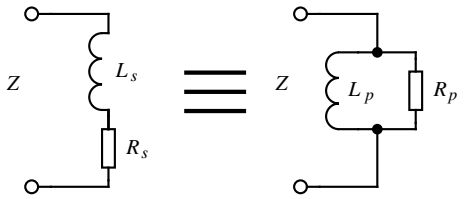


Figure 1.12 Series and parallel equivalent.

$$R_s + j\omega_o L_s = R_p \,\|\, j\omega_o L_p = \frac{(\omega_o L_p)^2 R_p + j\omega_o L_p R_p^2}{R_p^2 + (\omega_o L_p)^2} \tag{1.24}$$

If the serial and parallel sections are supposed to be equivalent, their Q's must certainly be equivalent and

$$Q = Q_s = \frac{\omega_0 L_s}{R_s} = Q_p = \frac{R_P}{\omega_0 L_p} \tag{1.25}$$

When the real and imaginary parts in equation (1.24) are separated and combined with equation (1.25) the result is

$$R_p = R_s(1 + Q^2) \quad \text{and} \tag{1.26}$$

$$L_p = L_s\left(\frac{1+Q^2}{Q^2}\right) \tag{1.27}$$

In the same manner similar sets of equations for computing series and parallel *RC* equivalents can be derived. The resistive part will be identical to equation (1.26) and the capacitive part will be

$$C_p = C_s\left(\frac{Q^2}{1+Q^2}\right) \tag{1.28}$$

Note the similarity with the equations (1.17) and (1.21) when impedances are converted to admittance or vice versa. The expressions (1.26) to (1.28) may be written in a universal form that applies to both RL and RC networks or to impedance-admittance conversion:

$$R_p = R_s(1+Q^2) \quad \text{where } R \text{ is the resistive part, and} \tag{1.29}$$

$$X_p = X_s\left(\frac{1+Q^2}{Q^2}\right) \quad \text{where } X \text{ is the reactive part.} \tag{1.30}$$

Accordingly, any circuit or component can be characterised either by a series or parallel equivalent. Independently of the used configuration, the general circuit properties (Z, f_0, Q) are assumed to be equal. The formulas above are very handy tools to convert any mixed *RLC* network into a pure parallel (or series) one that is straightforward to analyse. However, it is vitally important to remember that the equivalencies are valid only in a narrow span around the resonant frequency.

1.3 Tapped Resonant Circuits

The *RLC* networks discussed so far lacks flexibility because the bandwidth and Q are fixed once the values of R and any load resistance are specified. In order to get an extra degree of freedom to independently choose the bandwidth and the load resistance, an additional circuit element is required. A common way to do this is to divide either the capacitor or inductor into two series components. The low-resistance load is then assumed to be connected across one of them as in figure 1.13. Alternatively a “tap” can be joined to a continuous coil as shown in figure 1.18. Tapped circuits are frequently used in intermediate-frequency amplifiers and oscillators since it combines a resonator with impedance transformation that allows coupling to the circuit without degrading Q excessively. The remaining part of this chapter will deal with voltage and impedance transformation in tapped resonant circuits and later on the topic of impedance matching and network design will be extensively discussed in Chapter 5.

1.3.1 Tapped Capacitor

The operation of the tap is best understood as a consequence of the capacitive voltage divider. A voltage transformation in a perfectly lossless network must be accompanied by an impedance transformation proportional to the square of the voltage ratio if power is to be conserved. As an additional feature the tap provides DC-isolation between the source and the resonant circuit. It is to be designed for specified values for the source (or load) resistance R_g, resonance frequency f_0, bandwidth $B_{3\,\mathrm{dB}}$ and the values of L, C_1 and C_2 are to be found. For the sake of clarity, the reactances are assumed to be ideal in this section. To avoid confusion with the serial resistance R_s, the source is indexed "g" in this section.

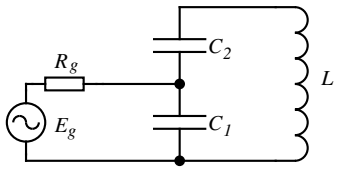


Figure 1.13 A tapped-capacitor circuit with a connected source.

1.3.1.1 Unloaded Tap

When the tap is considered to be unloaded, i.e. $R_g \gg X_{C_1}$ (for practical work $R_g \geq 10X_{C_1}$ is good enough), the following expressions may be used:

$$E_g' = E_g\left(\frac{X_{C_1} + X_{C_2}}{X_{C_1}}\right) = E_g\left(\frac{C_1 + C_2}{C_2}\right) \tag{1.31}$$

Note the similarity of equation (1.31) to resistive voltage dividers. As there is no additional loss of power,

$$R_g' = R_g\left(\frac{C_1 + C_2}{C_2}\right)^2 \tag{1.32}$$

$$C_{eq} = \frac{C_1 C_2}{C_1 + C_2} \tag{1.33}$$

Note that R_g must be less than R'_g because of the analogy with a voltage divider. The equivalent circuit, when the source is transformed into a pure parallel circuit, is shown in figure 1.14. Now the already familiar expressions for f_0, Q and $B_{3\,\mathrm{dB}}$ can be used to solve the task.

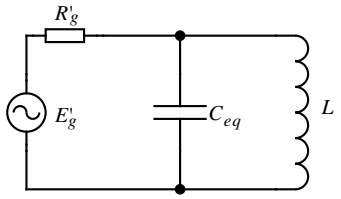


Figure 1.14 The source resistance transformed into the resonant circuit.

1.3.1.2 Loaded Tap

If the tap can not be regarded as unloaded as above, the design process becomes more complicated. In general, the impedance transformation can be calculated by successive parallel-to-series conversion as shown in the following section.

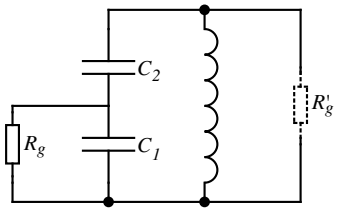


Figure 1.15 Transformation of R_g into R'_g when the tap is heavily loaded i.e. $R_g \ll X_{C_1}$.

The deduction is performed in two steps:

1. Parallel-to-series conversion of R_g and C_1

   Equations (1.29) and (1.30) and the quality factor Q_1 corresponding to the sub-circuit R_g and C_1 are used to transform these elements into the equivalent series elements R'_s and C'_1 in the capacitive branch.

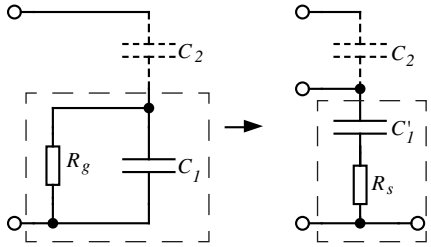


Figure 1.16 Parallel-to-series conversion.

$$Q_1 = R_g \cdot \omega_0 C_1 \tag{1.34}$$

$$R_s = \frac{R_g}{1 + Q_1^2} \tag{1.35}$$

$$C'_1 = C_1\left(\frac{1 + Q_1^{\,2}}{Q_1^{\,2}}\right) \tag{1.36}$$

If Q_1 is not reasonably high, it is obvious that the resonance frequency will be slightly decreased. However, for simplicity it is here assumed that the frequency shift can be neglected.

2. Series-to-parallel conversion of R_s and C'_{eq}

   In the final step R_s is transformed into an equivalent parallel resistor R_p across C''_1 and C_2. For this purpose, a new Q_2 from the sub-circuit R_s and C_{eq} has to be calculated. If Q_2 is sufficiently large, which is the normal case at this end, $C''_{eq} \approx C'_{eq}$ and there is no need to calculate a final value for C''_1.

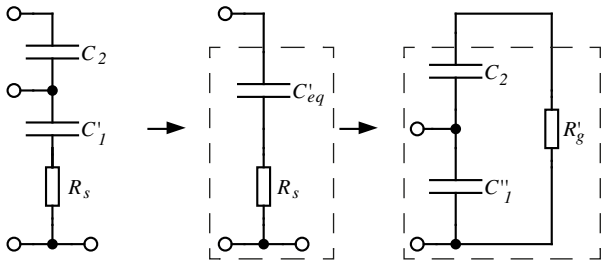


Figure 1.17 Series-to-parallel conversion.

$$C_{eq} = \frac{C'_1 \cdot C_2}{C'_1 + C_2} \tag{1.37}$$

$$Q_2 = \frac{1}{R_s \cdot \omega_0 C'_{eq}} \text{ (assuming unshifted } \omega_0\text{)} \tag{1.38}$$

$$R'_g = R_s(1 + Q_2^2) \tag{1.39}$$

$$C''_{eq} = C'_{eq}\left(\frac{Q_2^2}{1 + Q_2^2}\right) \approx C'_{eq} \text{ if } Q_2 \text{ large} \tag{1.40}$$

These calculations may see, tedious, but the ideas behind them are very simple. After some experience about the serial-parallel conversions, the computations are quite reasonable to solve.

1.3.2 Tapped Inductor, Transformer

An alternative to the capacitive tap is to achieve the coupling by a tapped-inductor circuit or autotransformer. If the coil is wound upon a ferrite core such that the coefficient of coupling is close to unity, it behaves like an ideal transformer and the results are readily predictable. With air-cored coils that are often used in the higher-frequency ranges, the coupling coefficient k may be on the order of 0.1, and the ideal transformer approximation does not hold in all cases. Still the resonance frequency can be calculated and eventually R'_g can be neglected in many practical situations, but for a more accurate computation it is necessary to deal with a more elaborate analysis.

Non-ideal transformers will not be treated in this text. However, it is essential to keep in mind that losses in the windings and iron cores will rapidly increase with frequency.

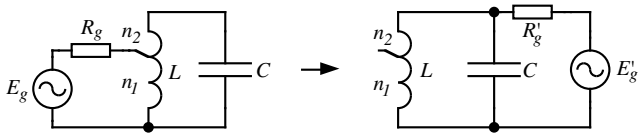


Figure 1.18 A tapped-inductor circuit with a connected source. The total number of turns is $n_1 + n_2$.

If there is need for DC-isolation an ordinary transformer can be used.

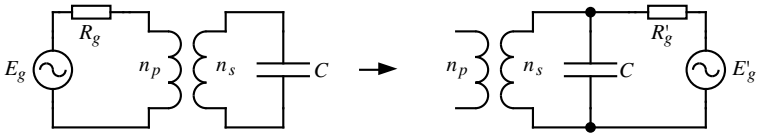


Figure 1.19 Transformer with separate windings.

The derivations are almost the same as for the tapped-capacitor case. Therefore only the transformation equations are given. If the transformer can be considered as unloaded, i.e. $Q \geq 10$, and with unity coupling (k=1), the following equations are useful:

$$E'_g = E_g\left(\frac{n_1 + n_2}{n_1}\right) = E_g\left(\frac{n_s}{n_p}\right) \tag{1.41}$$

$$R'_g = R_g\left(\frac{n_1 + n_2}{n_1}\right)^2 = R_g\left(\frac{n_s}{n_p}\right)^2 \tag{1.42}$$

Note that R_g must be less than R'_g because of the analogy to a voltage divider.

1.4 References

[1] J. Nilsson, S. Riedel, *Electric Circuits*, 5th edition, Addison-Wesley.

[2] H. L. Krauss, C. W. Bostian and F. H. Raab, *Solid state radio engineering*, John Wiley, 1980.

[3] P. H. Young, *Electronic Communication Techniques*, Charles E. Merill Publishing Company, 1994

Chapter 2

Transmission Lines

The previous chapter discussed resonant circuits which may be referred to as lumped circuits with components like inductors, capacitors and resistors. The term lumped is used because the components (and whatever is connecting them together) can be regarded as infinitesimally small, that this their physical size do not influence the electrical behaviour of the circuit. Therefore, when such circuits are analysed the connection between two components can be regarded as a node which has a certain node voltage. In other words, whether the connection is a small piece of copper on a printed circuit board or a long cable is not considered.

When a time-varying signal is present in a circuit, say, a sinusoid, it cannot be assumed that the circuit may be considered as lumped. Consider the simple circuit in figure 2.1 where a voltage source with an output impedance Z_S is connected to a load Z_L with a pair wires of equal length ℓ. If the length of the wires is disregarded each wire is assumed to be part of one node. Then, the voltage at the load is easily calculated as the voltage division between Z_S and Z_L.

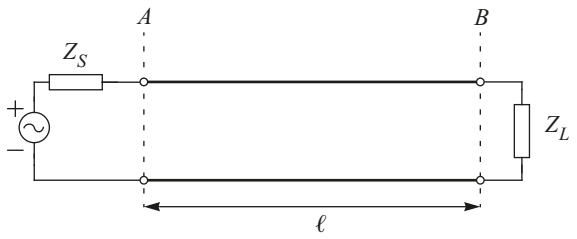


Figure 2.1 Voltage source connected to a load using wires with length ℓ.

It may be conceived by the reader that a signal appearing at A will propagate with a finite velocity to B. However, if the propagation time is much smaller than the time period of the signal it is obvious that the signal at A and B will be the same and therefore the circuit can be regarded as a *lumped circuit*.

When the time period of the signal is on the same order of magnitude as the propagation time or less it is equally obvious that the signal in A and B will not be the same any more. The length of the wires will influence the electrical behaviour of the circuit which now should be considered as a *distributed circuit*. In this case, the wire might be considered as a medium for transmission of a signal from a source to a load and the term *transmission line* or line for short is used. In its simplest form, a transmission line can be considered to be a pair of guiding conductors or wires as shown in figure 2.1.

Because the signal propagates from one point to another through a transmission line it is natural to use the term wave or as is common in electronics, *travelling wave*. A wavelength λ can be associated with the wave as it propagates in the line (λ_0 will be used to denote the wavelength in free space). The wavelength of the signal is defined as $\lambda = v_p \cdot T = v_p/f$, where v_p is the velocity of propagation, or more commonly *phase velocity*, and T is the time period of the signal. As the physical size is a prime parameter rather than the propagation time through the line, the wavelength is more convenient to use than the time period of the signal. That is, a circuit should be considered as distributed if the wavelength is on the same order of magnitude as the physical size of the circuit or less. The case where the wavelength is a fraction of the transmission line length is shown in figure 2.2 which illustrates the signal voltage along the transmission line at one time instance and also the signal voltage as a function of time at the left end of the transmission line.

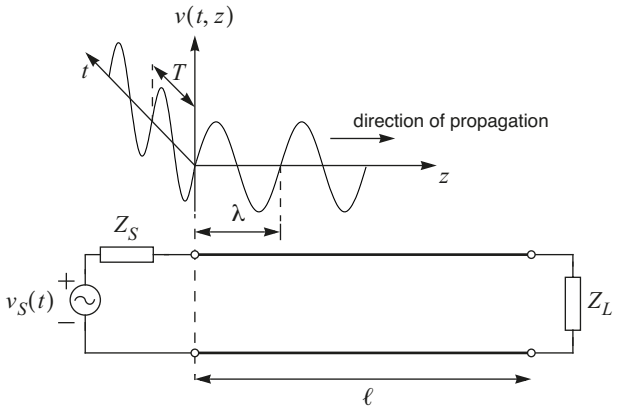


Figure 2.2 Travelling wave on a transmission line.

This chapter will start by presenting a theoretical basis for transmission lines and travelling waves. This in turn will be used to define the reflection coefficient, the standing wave ratio and some other useful quantities. Finally, we will discuss how the transmission line can be used for realising impedance transformation and resonant circuits.

2.1 Modelling

To understand transmission lines a mathematical representation of the wave is needed as well as an electrical model of the line. Below, the travelling wave will be defined first without introducing any parameters that directly relates

to the physical realisation of a transmission line. Secondly, the transmission line will be modelled using an equivalent lumped circuit and finally these two models will be tied together.

2.1.1 Modelling Waves on Transmission Lines

In figure 2.3 a piece of transmission line is assumed to be excited by a sinusoidal signal from the left. The signal will propagate along the line in the positive (forward) direction along the z-axis. The voltage at a given point of time and position along the line can then be written as

$$v^+(t,z) = \left|V_0^+\right|\cos(\omega t - \beta z + \phi_0^+) = \mathrm{Re}[V_0^+ e^{j(\omega t - \beta z)}] \tag{2.1}$$

where $V_0^+ = \left|V_0^+\right| e^{j\phi_0^+}$ is the *complex amplitude* or *phasor* of $v^+(t,z)$ at $z = 0$, β is the *phase constant* of the wave and ϕ_0 is an auxiliary phase. The index "$^+$" indicates positive direction of propagation.

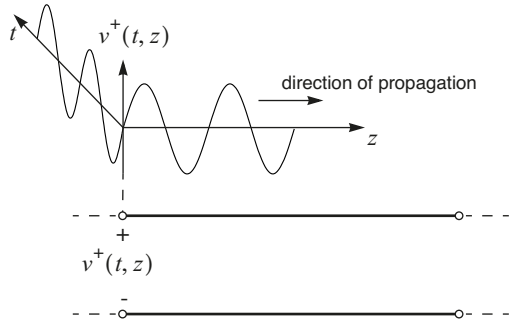


Figure 2.3 Wave travelling in a positive direction along a transmission line.

Amplitude maxima for $v^+(t,z)$ occurs for $\omega t - \beta z + \phi_0^+ = 2n\pi$ (n = integer). The phase of a specific point of the wave front shifts by 2π if z increases by one wavelength (λ) or if the time changes by one period (T), that is

$$2\pi = \omega T = \beta\lambda \Rightarrow$$

$$\lambda = \frac{2\pi}{\beta} \tag{2.2}$$

$$T = \frac{2\pi}{\omega} \tag{2.3}$$

The phase velocity of the wave is obtained from (2.2) and (2.3)

$$v_p = \frac{\lambda}{T} = \lambda f = \frac{\omega}{\beta} \tag{2.4}$$

In the general case, waves may travel in both direction in a transmission line and the wave propagating in the negative (reverse) direction, i.e. towards negative values of z, can be written as

$$v^-(t,z) = \left|V_0^-\right| \cos(\omega t + \beta z + \phi_0^-) = Re[V_0^- e^{j(\omega t + \beta z)}] \tag{2.5}$$

Since a sinusoidal signal with a fixed frequency ω is assumed the time dependency of the signals are of little importance and is therefore commonly omitted. Instead the wave phasor as a function of the position along the line is used, that is

$$V^+(z) = V_0^+ e^{-j\beta z} \tag{2.6}$$

$$V^-(z) = V_0^- e^{j\beta z} \tag{2.7}$$

Equally, we may define the complex amplitude of the total voltage at a given position when waves are travelling in both directions:

$$V(z) = V^+(z) + V^-(z) \tag{2.8}$$

So far, only the voltage of a wave has been discussed. It goes without saying that there must be an accompanying current wave that is related to the voltage wave and therefore

$$I^+(z) = I_0^+ e^{-j\beta z} \tag{2.9}$$

$$I^-(z) = I_0^- e^{j\beta z} \tag{2.10}$$

At a given position or node of the line the total voltage appear as the sum of the two waves as given in (2.8) and similarly the net current through the node or the position is simply given by

$$I(z) = I^+(z) - I^-(z) \tag{2.11}$$

The relation between the voltage wave and the current wave will be discussed later when the electrical properties of the transmission line have been dealt with.

When propagation takes place in a transmission line where the losses are noticeable, the wave is gradually attenuated, see figure 2.4. The losses mainly arise from the line itself but if a line has a leaky shield or no shield at all, energy will radiate from the line and cause signal loss. In order to take the

attenuation into consideration, $j\beta$ in the previous expressions is replaced by the *propagation constant,* $\gamma = \alpha + j\beta$, where α is the *attenuation constant* and β as earlier defined, the phase constant. The attenuation constant is expressed in *neper* per metre. It can be converted into dB/m by

$$1 \text{ neper} = 10 \log e^2 = 8.69 \text{ dB} \tag{2.12}$$

By substituting $j\beta$ by γ in (2.6) and (2.7) the expressions for waves on a lossy transmission lines are obtained

$$V^+(z) = V_0^+ e^{-\gamma z} = V_0^+ e^{-\alpha z} e^{-j\beta z} \tag{2.13}$$

$$V^-(z) = V_0^- e^{\gamma z} = V_0^- e^{\alpha z} e^{j\beta z} \tag{2.14}$$

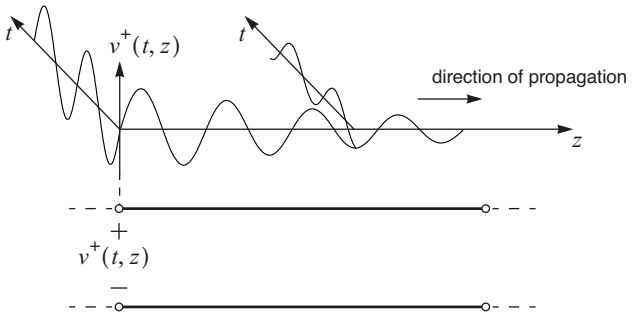


Figure 2.4 Wave travelling in a positive direction along a lossy transmission line.

This section has presented a model for the wave with a propagation coefficient γ to describe how the wave propagates along the line, e.g., in terms of velocity and attenuation. Although voltages and currents were assumed to constitute the wave, the model for the wave could just as well be used e.g. for an acoustic wave. Therefore, the next step involves relating the propagation coefficient to the electrical properties of a transmission line and also the relation between the voltage wave and the current wave must be defined to make the picture complete.

2.1.2 Modelling Transmission Lines

A transmission line is a conductor where the electrical properties are uniform with distance along the line. The analysis can be done by solving the Maxwell's field equations which involves, in addition to the time variable, three space variables. This is, however, an excessive approach as the result would

contain much more information than what is necessary here. Also, hand calculation would be out of the question. Instead, distributed-circuit theory will be used which involves only one space variable besides the time dependency.

A transmission line can be modelled using a lumped T-network circuit as shown in figure 2.5 with the following elements

- series resistor to model loss in the conductor
- series inductor to represent the current and stored magnetic energy in the line.
- shunt resistor (represented by its conductance) to model loss in the material supporting the two conductors (e.g. substrate on a printed circuit board)
- shunt capacitor to represent the charge and the stored electric energy in the line.

Because the properties of the line are distributed the section of length Δz should be regarded as an infinitesimal section of the line and consequently the four parameters (R, L, G, C) are given as "per unit length" and the circuit element values for one section are therefore obtained as these values multiplied by the length Δz.

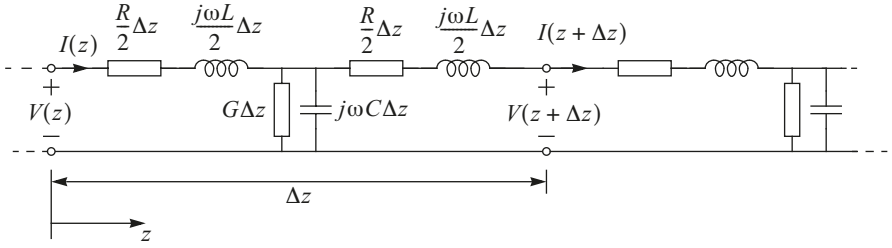


Figure 2.5 A lumped circuit model for a transmission line.

The voltage and current drops in one section may be written as

$$\frac{dV(z)}{dz}\Delta z = V(z+\Delta z) - V(z) = -I(z)(R+j\omega L)\Delta z \Rightarrow$$
$$\frac{dV(z)}{dz} = -I(z)(R+j\omega L) \tag{2.15}$$

$$\frac{dI(z)}{dz}\Delta z = I(z+\Delta z) - I(z) = -V(z)(G+j\omega C)\Delta z \Rightarrow$$
$$\frac{dI(z)}{dz} = -V(z)(G+j\omega C) \tag{2.16}$$

Dividing these two equations by Δz and taking the limit as $\Delta z \to 0$ the *transmission-line equations* (or *telegrapher equations*) are obtained as

$$\frac{d^2V(z)}{dz^2} = -\frac{dI(z)}{dz}(R+j\omega L) = (R+j\omega L)(G+j\omega C)V(z) \tag{2.17}$$

$$\frac{d^2I(z)}{dz^2} = -\frac{dV(z)}{dz}(G+j\omega C) = (R+j\omega L)(G+j\omega C)I(z) \tag{2.18}$$

These equations are given using the phasor notation of voltages and currents and it may be suitable at this point to remark the connection with the time domain representation of the signals which are easily obtained from

$$v(z,t) = \mathrm{Re}[V(z)e^{j\omega t}] \tag{2.19}$$

$$i(z,t) = \mathrm{Re}[I(z)e^{j\omega t}] \tag{2.20}$$

Remembering the relations between the wave voltages (currents) and the total cross-section voltage (current) at a given position along the z-axis:

$$V(z) = V_0^{+}e^{-\gamma z} + V_0^{-}e^{\gamma z} \tag{2.21}$$

$$I(z) = I_0^{+}e^{-\gamma z} - I_0^{-}e^{\gamma z} \tag{2.22}$$

The relations between the lumped circuit model of the transmission line and the propagation coefficient defined earlier can be derived by differentiating (2.21) and (2.22) two times with respect to z (compare with (2.17) and (2.18)):

$$\begin{aligned}\frac{d^2}{dz^2}(V(z)) &= \frac{d}{dz}(-\gamma\cdot(V_0^{+}e^{-\gamma z} - V_0^{-}e^{\gamma z})) \\ &= \gamma^2\cdot(V_0^{+}e^{-\gamma z} + V_0^{-}e^{\gamma z}) = \gamma^2 V(z)\end{aligned} \tag{2.23}$$

$$\begin{aligned}\frac{d^2}{dz^2}(I(z)) &= \frac{d}{dz}(-\gamma\cdot(I_0^{+}e^{-\gamma z} + I_0^{-}e^{\gamma z})) \\ &= \gamma^2\cdot(I_0^{+}e^{-\gamma z} - I_0^{-}e^{\gamma z}) = \gamma^2 I(z)\end{aligned} \tag{2.24}$$

If this result is compared with (2.17) and (2.18) we obtain

$$\gamma^2 = (R+j\omega L)(G+j\omega C) \tag{2.25}$$

Thus, there is now a direct connection between the propagation coefficient and the electrical properties of the line. To be able to relate the current wave and voltage wave (2.15) and (2.23) can be exploited:

$$\begin{aligned}\frac{dV(z)}{dz} &= -I(z)(R+j\omega L) \Rightarrow \\ I(z) &= I_0^{+}e^{-\gamma z} - I_0^{-}e^{\gamma z} = -\frac{1}{R+j\omega L}\cdot\frac{dV(z)}{dz} \\ &= -\frac{1}{R+j\omega L}\cdot(-\gamma\cdot(V_0^{+}e^{-\gamma z} - V_0^{-}e^{\gamma z})) \\ &= \sqrt{\frac{G+j\omega C}{R+j\omega L}}\cdot(V_0^{+}e^{-\gamma z} - V_0^{-}e^{\gamma z})\end{aligned} \tag{2.26}$$

The ratio of the voltage wave and the current wave is easily obtained from (2.26) and is given by

$$Z_0 = \frac{V_0^{+}}{I_0^{+}} = \frac{V_0^{-}}{I_0^{-}} = \sqrt{\frac{R+j\omega L}{G+j\omega C}} \tag{2.27}$$

where Z_0 is referred to as the *characteristic impedance* of the transmission line. Note that this is not an impedance in the normal sense as it applies to waves rather than node voltages and currents. Also, the characteristic admittance may be defined as $Y_0 = 1/Z_0$.

2.1.3 Low-Loss Transmission Lines

For practical transmission lines the loss in rather low such that $R \ll \omega L$ and $G \ll \omega C$. Under these assumptions the propagation constant can be approximated as

$$\begin{aligned}\gamma &= \alpha + j\beta = \sqrt{(R+j\omega L)(G+j\omega C)} \\ &= \sqrt{(j\omega)^2}\cdot\sqrt{\left(1+\frac{R}{j\omega L}\right)\left(1+\frac{G}{j\omega C}\right)} \\ &\approx j\omega\sqrt{LC}\cdot\left[\left(1+\frac{1}{2}\frac{R}{j\omega L}\right)\left(1+\frac{1}{2}\frac{G}{j\omega C}\right)\right] \\ &\approx j\omega\sqrt{LC}\cdot\left[1+\frac{1}{2}\left(\frac{R}{j\omega L}+\frac{G}{j\omega C}\right)\right] \\ &= \frac{1}{2}\left(R\sqrt{\frac{C}{L}}+G\sqrt{\frac{L}{C}}\right)+j\omega\sqrt{LC}\end{aligned} \tag{2.28}$$

Therefore the attenuation and phase constants are given by

$$\alpha \approx \frac{1}{2}\left(R\sqrt{\frac{C}{L}}+G\sqrt{\frac{L}{C}}\right) \tag{2.29}$$

$$\beta \approx \omega\sqrt{LC} \tag{2.30}$$

Similarly, the characteristic impedance is approximated by

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \approx \sqrt{\frac{L}{C}} \tag{2.31}$$

Note that (2.31) shows that the characteristic impedance is real-valued if the loss is small. Using (2.31), the propagation constant can also be written as

$$\gamma = \alpha + j\beta = \frac{1}{2}(RY_0 + GZ_0) + j\omega\sqrt{LC} \tag{2.32}$$

From equation (2.30) the phase velocity is

$$v_p = \frac{\omega}{\beta} = \frac{1}{\sqrt{LC}} \tag{2.33}$$

It is important to note that the product LC is independent of the geometry of the transmission line and depends only on the permeability μ and permittivity ε of the insulating medium in the transmission line (e.g. the substrate of a printed circuit board or the supporting material between the conductor and the shield in a cable). In the special case with air as the insulator free-space parameters can be assumed. Thus, the phase velocity is approximately equal to the velocity of light in free space:

$$v_p = \frac{1}{\sqrt{LC}} = \frac{1}{\sqrt{\mu_0\varepsilon_0}} = c = 3 \cdot 10^8 \text{m/s} \tag{2.34}$$

Again, if the supporting material has a permeability and/or permittivity different from their free-space counterparts we have

$$v_p = \frac{1}{\sqrt{\mu\varepsilon}} = \frac{1}{\sqrt{\mu_r\varepsilon_r\mu_0\varepsilon_0}} = \frac{c}{\sqrt{\mu_r\varepsilon_r}} \tag{2.35}$$

The relative phase velocity can be defined as

$$\text{Velocity factor} = \frac{\text{phase velocity in given medium}}{\text{velocity of light in free space}}$$

$$v_r = \frac{v_e}{c} = \frac{1}{\sqrt{\mu_r\varepsilon_r}} \tag{2.36}$$

2.1.4 Reflection Coefficient

When a wave (let it be e.g. a travelling wave or an acoustic wave) reaches a discontinuity in the propagation medium some amount of the wave will be reflected. Therefore, in the general case, the voltage or current at just any position along a transmission line is a sum of two waves: one wave travelling in the forward (positive) direction and the other travelling in the reverse (negative) direction. For the transmission line, the discontinuity might be an impedance terminating the line as shown in figure 2.6.

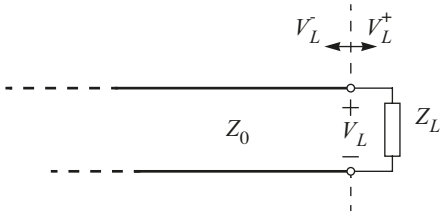


Figure 2.6 Transmission line terminated by an impedance.

In the interface between the line and the terminating impedance we may define an incident wave travelling in the forward direction, V_L^+. The ratio between this wave and its associated current I_L^+ is given by the characteristic impedance of the line Z_0. However, the ratio of the voltage across the load $\boldsymbol{V}_L$ and the current through the load $\boldsymbol{I}_L$ is of course given by Z_L. This mismatch in impedance levels results in that the incident wave cannot fully be absorbed by the load and this results in a reflection. Noting that

$$V_L = I_L Z_L = (I_L^+ - I_L^-) \cdot Z_L \tag{2.37}$$

$$V_L = V_L^+ + V_L^- = (I_L^+ + I_L^-) \cdot Z_0 \tag{2.38}$$

gives

$$(I_L^+ - I_L^-) \cdot Z_L = (I_L^+ + I_L^-) \cdot Z_0 \tag{2.39}$$

and solving for I_L^- yields

$$I_L^- = I_L^+ \cdot \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.40}$$

Consequently,

$$V_L^- = V_L^+ \cdot \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.41}$$

Here we note that there will be no reflection only when the impedance of the load will be equal to the characteristic impedance of the line.

The ratio between the reflected and incident wave is a commonly used quantity and is termed *reflection coefficient*, denoted by Γ. In other words,

$$\Gamma \equiv \frac{\text{reflected voltage}}{\text{incident voltage}} = \frac{\text{reflected current}}{\text{incident current}} = \frac{\bar{V}}{V^+} = \frac{\bar{I}}{I^+} \tag{2.42}$$

Thus, for the specific case discussed above where the reflection in the interface between a transmission line and a load was considered the reflection coefficient becomes (also using (2.41))

$$\Gamma_L = \frac{V_L^-}{V_L^+} = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.43}$$

or if solved for Z_L

$$Z_L = Z_0 \frac{1 + \Gamma_L}{1 - \Gamma_L} \tag{2.44}$$

While the reflection coefficient is a convenient quantity to describe the mismatch between a transmission line and a terminating impedance it is by no means limited to that case. The use of the reflection coefficient can be generalised to describe the ratio between the reverse and forward waves at any position along a transmission line. Starting from the load and moving from it we will move against the propagation direction of the forward wave but follow the propagation direction of the reverse wave. Thus, the forward and reverse waves will change by a certain amount of phase but with opposite signs and therefore the reflection coefficient will change accordingly.

Figure 2.7 illustrates a line terminated with a load as discussed above but now with the extension that the waves should be investigated at a distance d from the load. Note that d is measured from the load in the negative direction with reference to the z-axis.

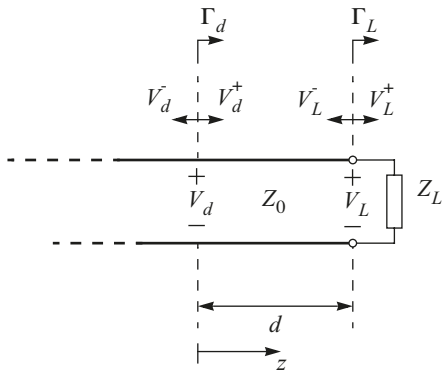


Figure 2.7 Transformation of the reflection coefficient.

Equations (2.13) and (2.14) can be used to investigate how the waves are changing with distance (in the positive z direction) starting at a distance d from the load:

$$V^+(z) = V_d^+ e^{-\gamma z} \tag{2.45}$$

$$V^-(z) = V_d^- e^{\gamma z} \tag{2.46}$$

Setting $z = d$ gives

$$V^+(d) = V_L^+ = V_d^+ e^{-\gamma d} \tag{2.47}$$

$$V^-(d) = V_L^- = V_d^- e^{\gamma d} \tag{2.48}$$

and consequently

$$\Gamma_L = \frac{V_L^-}{V_L^+} = \frac{V_d^- e^{\gamma d}}{V_d^+ e^{-\gamma d}} = \Gamma_d \cdot e^{2\gamma d} = \Gamma_d \cdot e^{2\alpha d} \cdot e^{j2\beta d} \text{ or} \tag{2.49}$$

$$\Gamma_d = \Gamma_L \cdot e^{-2\gamma d} = \Gamma_L \cdot e^{-2\alpha d} \cdot e^{-j2\beta d} \tag{2.50}$$

At this point it goes without saying that the reflection coefficient is, in general, a complex-valued quantity. As such we may write

$$\Gamma = |\Gamma| e^{j\theta} \tag{2.51}$$

where the magnitude $|\Gamma|$ is usually less than unity. Note that θ is the phase angle between the forward and reverse voltages and is referred to as the phase angle of the reflection coefficient. Returning to equation (2.50) it can be seen that phase constant β results in a clockwise rotation of Γ_d with increasing distance d from the load whereas the attenuation constant α results in an exponential decay of the magnitude $|\Gamma_d|$. This is exemplified in figure 2.8.

For a lossless line the magnitude remains constant and only the phase of Γ is shifted with an angle of $-2\beta d$ as shown in figure 2.9.

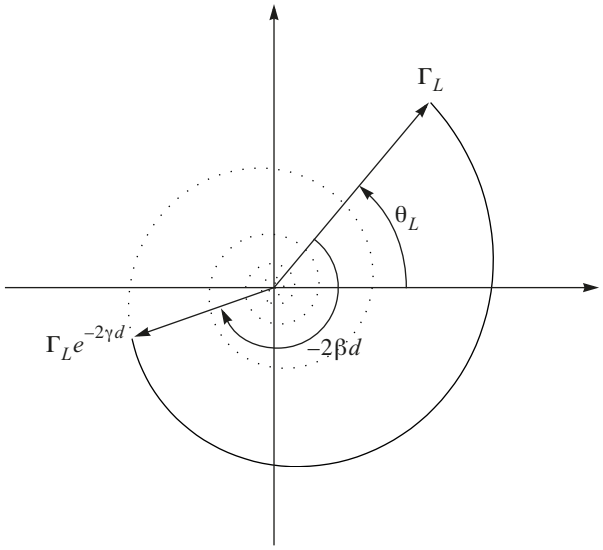


Figure 2.8 Reflection coefficient for a lossy transmission line.

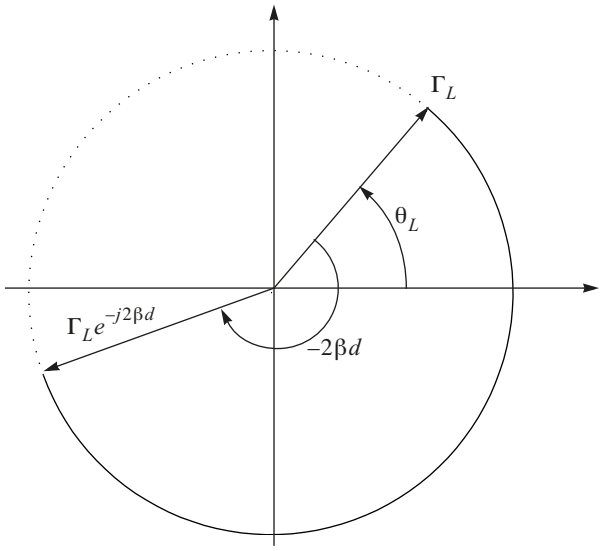


Figure 2.9 Reflection coefficient for a lossless transmission line.

2.1.5 Impedance and Admittance Transformation

The reflection coefficient was first introduced to represent the mismatch between the characteristic impedance of a transmission line and a terminating load impedance. As the reflection coefficient is given by the ratio between the reverse and forward waves, the definition of the reflection coefficient could be extended to represent the ratio of these waves at any position along a transmission line. Furthermore, an important relationship between the reflection coefficient at the load and at a distance d from the load, respectively, was found to be (refer to figure 2.9)

$$\Gamma_d = \Gamma_L \cdot e^{-2\gamma d} \tag{2.52}$$

Equation (2.52) can be viewed as a transformation of the reflection coefficient at the load. Similarly, the load impedance will also be transformed by the transmission line. Referring to figure 2.10, the impedance Z_d is given by

$$Z_d = \frac{V_d}{I_d} \tag{2.53}$$

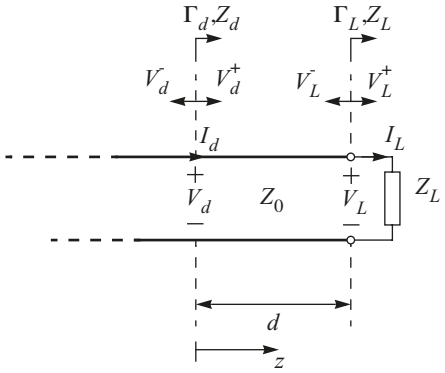


Figure 2.10 Transformation of load reflection coefficient and impedance using a transmission line.

To derive an expression for Z_d the associated voltage $\boldsymbol{V}_d$ and current $\boldsymbol{I}_d$ are expressed in terms of waves which in turn can be transformed to waves at the load, i.e.

$$V_d = V_d^+ + V_d^- = V_L^+ e^{\gamma d} + V_L^- e^{-\gamma d} \tag{2.54}$$

$$I_d = I_d^+ - I_d^- = I_L^+ e^{\gamma d} - I_L^- e^{-\gamma d} = Y_0 (V_L^+ e^{\gamma d} - V_L^- e^{-\gamma d}) \tag{2.55}$$

At the load end ($d = 0$) equations (2.54) and (2.55) become

$$I_L Z_L = V_L^+ + V_L^- \tag{2.56}$$

$$I_L Z_0 = V_L^+ - V_L^- \tag{2.57}$$

By solving these two equations for V_L^+ and V_L^- we obtain

$$V_L^+ = \frac{I_L}{2}(Z_L + Z_0) \tag{2.58}$$

$$V_L^- = \frac{I_L}{2}(Z_L - Z_0) \tag{2.59}$$

Substitution of V_L^+ and V_L^- in equations (2.54) and (2.55) yields

$$V_d = \frac{I_d}{2}[(Z_L + Z_0)e^{\gamma d} + (Z_L - Z_0)e^{-\gamma d}] \tag{2.60}$$

$$I_d = \frac{I_d}{2Z_0}[(Z_L + Z_0)e^{\gamma d} - (Z_L - Z_0)e^{-\gamma d}] \tag{2.61}$$

and finally the ratio of equations (2.60) and (2.61) gives

$$Z_d = Z_0 \frac{(Z_L + Z_0)e^{\gamma d} + (Z_L - Z_0)e^{-\gamma d}}{(Z_L + Z_0)e^{\gamma d} - (Z_L - Z_0)e^{-\gamma d}} \tag{2.62}$$

Some simplifications are possible by using

$$e^{\pm\gamma d} = \cosh\gamma d \pm \sinh\gamma d \tag{2.63}$$

which yields

$$Z_d = Z_0 \frac{Z_L \cosh\gamma d + Z_0 \sinh\gamma d}{Z_0 \cosh\gamma d + Z_L \sinh\gamma d} = Z_0 \frac{Z_L + Z_0 \tanh\gamma d}{Z_0 + Z_L \tanh\gamma d} \tag{2.64}$$

For the special but common case of a line with no or negligible loss the propagation constant is reduced to $\gamma = j\beta$, and using the following relationships,

$$\sinh j\beta d = j \sin\beta d \quad \text{and} \quad \cosh j\beta d = \cos\beta d\,, \tag{2.65}$$

yield

$$Z_d = Z_0 \frac{Z_L \cos\beta d + jZ_0 \sin\beta d}{Z_0 \cos\beta d + jZ_L \sin\beta d} = Z_0 \frac{Z_L + jZ_0 \tan\beta d}{Z_0 + jZ_L \tan\beta d} \tag{2.66}$$

The transformation of impedances using transmission lines is extensively used in radio and microwave electronics. Two special but important cases will be presented below where a line of a given length ℓ is terminated with a short-circuit and an open-circuit, respectively, see figure 2.11.

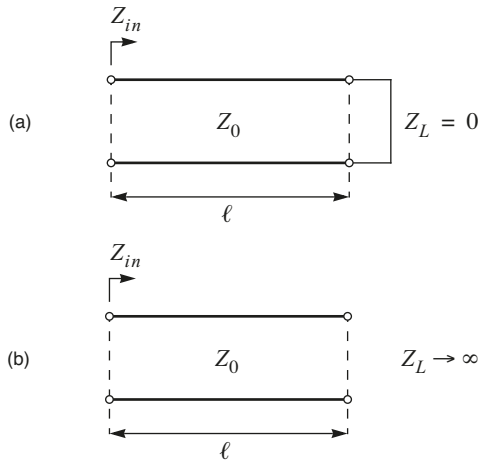


Figure 2.11 Transmission line with (a) short-circuit load and (b) open-circuit load.

The frequency behaviour of these circuits are particularly interesting and can be derived from (2.66) assuming that the lines are lossless. For the short-circuit load the input impedance is given by

$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan\beta\ell}{Z_0 + jZ_L \tan\beta\ell} = jZ_0 \tan\beta\ell = jZ_0 \tan\frac{2\pi f\ell}{v_p} \tag{2.67}$$

If the frequency is sufficiently low the above expression can be linearised as

$$Z_{in} = jZ_0 \tan\frac{2\pi f\ell}{v_p} \approx jZ_0 \frac{2\pi f\ell}{v_p} = j2\pi f \cdot \frac{Z_0\ell}{v_p} \tag{2.68}$$

which shows that for low frequencies the short-circuited transmission line behaves like an inductor with a value $L = Z_0\ell / v_p$. This is exemplified in figure 2.12 where the input reactance has been plotted according to (2.67) and the low-frequency approximation (2.68).

Similarly, with an open-circuit load the input impedance is given by

$$Z_{in} = Z_0 \frac{Z_L + jZ_0 \tan\beta\ell}{Z_0 + jZ_L \tan\beta\ell} = \frac{Z_0}{j\tan\beta\ell} = \frac{Z_0}{j\tan\frac{2\pi f\ell}{v_p}} \tag{2.69}$$

Again, if the frequency is sufficiently low the impedance is given by

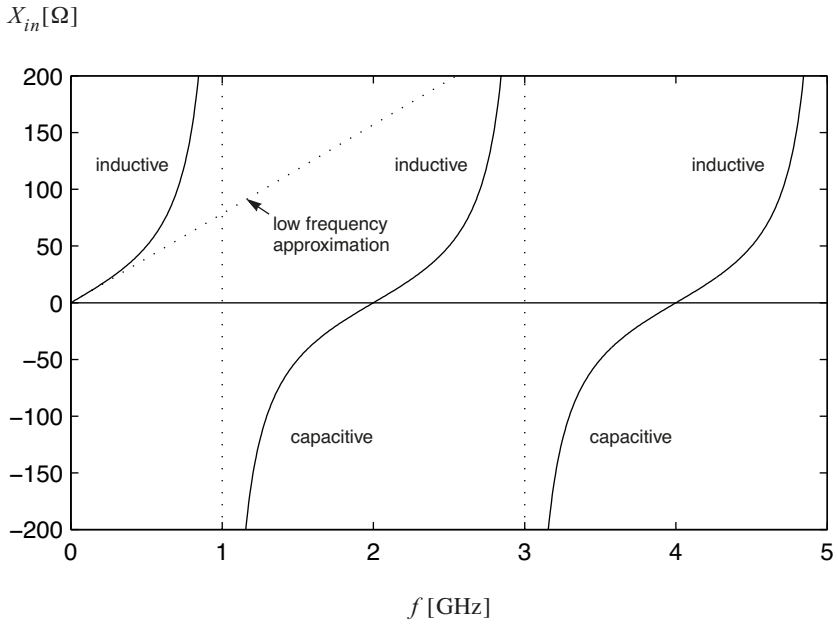


Figure 2.12 Input reactance of a line with short-circuit termination.
($\ell = 0.075\,\text{m}$, $Z_0 = 50\Omega$, $v_p = c$)

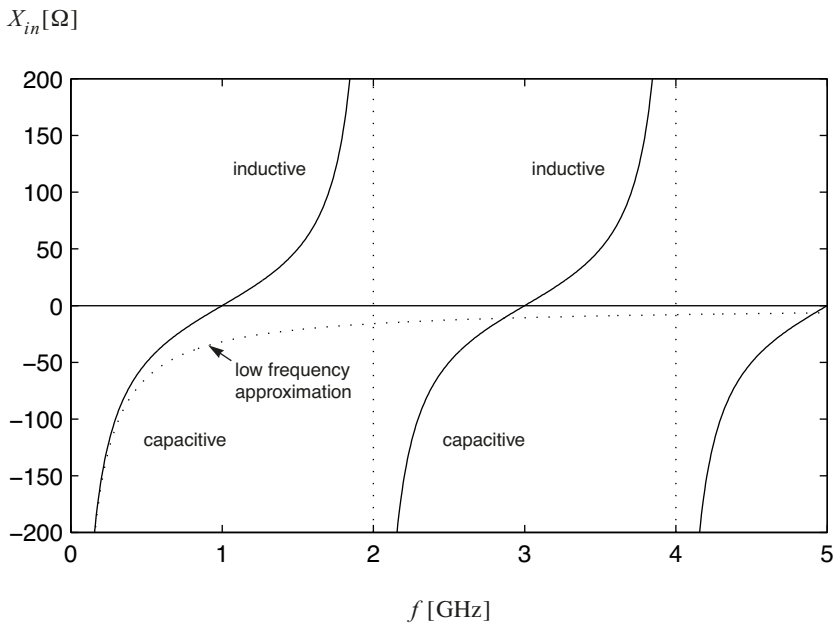


Figure 2.13 Input reactance of a line with open-circuit termination.
($\ell = 0.075\,\text{m}$, $Z_0 = 50\Omega$ and $v_p = c$)

$$Z_{in} = \frac{Z_0}{j\tan\dfrac{2\pi f\ell}{v_p}} \approx \frac{-jZ_0 v_p}{2\pi f\ell} = \frac{1}{j2\pi f \cdot \dfrac{\ell}{v_p Z_0}} \tag{2.70}$$

which shows that for low frequencies the open-circuit transmission line behaves like an capacitor with a value $C = \ell/(v_p Z_0)$. This illustrated in figure 2.13 which shows the input reactance based on (2.69) as well as on the low-frequency approximation of (2.70).

For these two special cases we can conclude that a short-circuit and an open-circuit transmission line can be used to emulate the behaviour of an inductor and capacitor, respectively. The frequency range is limited though, the curves in figures 2.12 and 2.13 degenerates completely at 1GHz (with the given line parameters) and the approximation is valid to about 500MHz. The length used for the lines corresponds to a quarter of a wavelength at 1GHz and thus the approximation is valid for $\ell < \lambda/8$.

2.1.5.1 Impedance and Reflection Coefficient

So far, expressions have been derived for the transformation of the reflection coefficient and the impedance, respectively, along a transmission line. However, with the introduction of the reflection coefficient it was given that

$$Z_L = Z_0 \frac{1+\Gamma_L}{1-\Gamma_L} \tag{2.71}$$

which was derived for the case with a load connected to a transmission line. This expression is, however, not limited to this. By combining (2.43) and (2.62) the same expression is obtained for Z_d:

$$Z_d = Z_0 \left(\frac{1 + \dfrac{(Z_L - Z_0)}{(Z_L + Z_0)} e^{-2\gamma d}}{1 - \dfrac{(Z_L - Z_0)}{(Z_L + Z_0)} e^{-2\gamma d}} \right) = Z_0 \frac{1 + \Gamma_L e^{-2\gamma d}}{1 - \Gamma_L e^{-2\gamma d}} \tag{2.72}$$

and since $\Gamma_d = \Gamma_L e^{-2\gamma d}$,

$$Z_d = Z_0 \frac{1+\Gamma_d}{1-\Gamma_d} \tag{2.73}$$

which is, as expected, identical with (2.71). This expression reveals an important property of transmission lines, a repeated impedance pattern with increasing d. For a lossless line

$$\Gamma_d = \Gamma_L e^{-j2\beta d} \tag{2.74}$$

In other words, the reflection coefficient rotates clockwise with increasing distance from the load. This means that $\Gamma_d = \Gamma_L$ for $2\beta d = 2n\pi$ where n is an integer and consequently the impedance along a lossless line will be repeated for every interval of a half-wavelength distance:

$$Z(d) = Z\left(d + \frac{\pi}{\beta}\right) = Z\left(d + \frac{\lambda}{2}\right) \tag{2.75}$$

While the reflection coefficient has so far been defined for the case where waves can be identified on a transmission line equation (2.73) suggests that the reflection coefficient can be used to represent any impedance as long as Z_0 is given. In this context Z_0 may be referred to as a reference impedance rather than a characteristic impedance. Furthermore, the impedance itself, can by represented by a *normalised impedance*

$$z \equiv \frac{Z}{Z_0} = \frac{1+\Gamma}{1-\Gamma} = r \pm jx \tag{2.76}$$

and similarly the *normalised admittance* is given by

$$y \equiv \frac{Y}{Y_0} = \frac{Z_o}{Z} = \frac{1}{z} = \frac{1-\Gamma}{1+\Gamma} = g \pm jb \tag{2.77}$$

It should be noted that the lowercase letters are commonly designated for normalised quantities in describing distributed transmission-line circuits.

2.1.6 Waves and Power

As a wave may be represented by a voltage wave and an associated current wave the wave carries power. Considering the case when an incident wave propagates along line towards a load with $Z_L \neq Z_0$ a wave will be reflected. This wave will, of course also carry power.

Assuming that the incident voltage wave and the associated current wave at the load (see figure 2.14) are in phase (that is Z_0 is real-valued) the power carried by the wave is given by

$$P_L^+ = \frac{1}{2} \cdot V_L^+ \cdot (I_L^+)^* = \frac{\left|V_L^+\right|^2}{2Z_0} \tag{2.78}$$

where the voltage and the current are given as peak values. Similarly, the power of the reflected wave can be written as

$$P_L^- = \frac{1}{2} \cdot V_L^- \cdot (I_L^-)^* = \frac{\left|V_L^-\right|^2}{2Z_0} = P_L^+ \cdot \left|\Gamma_L\right|^2 \tag{2.79}$$

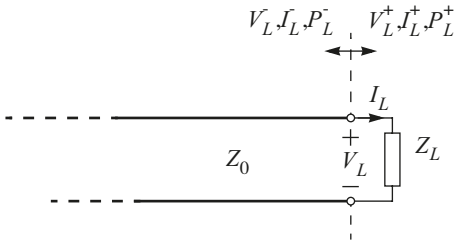


Figure 2.14 Transmission of power via transmission line to load.

According to the principle of conservation of energy, the incident power minus the reflected power must be equal to the power transmitted to the load:

$$P_L = P_L^+ - P_L^- = T_p P_L^+ \tag{2.80}$$

where T_p denotes the *transmission factor* which is defined as

$$T_p \equiv \frac{\text{transmitted power to the load}}{\text{incident power}} \tag{2.81}$$

The transmission factor can be expressed in terms of the reflection coefficient as

$$T_p = (1 - \left|\Gamma_L\right|^2) \tag{2.82}$$

The relation between the corresponding voltages or currents is called the *transmission coefficient* and is defined as

$$T \equiv \frac{\text{transmitted voltage or current}}{\text{incident voltage or current}} = \frac{V_L}{V_L^+} = \frac{I_L}{I_L^+} \tag{2.83}$$

Noting that the voltage across the load is given by

$$V_L = V_L^+ + V_L^- = V_L^+(1 + \Gamma_L) = V_L^+ \frac{2Z_L}{Z_L + Z_0} \tag{2.84}$$

the voltage transmission coefficient can be written as

$$T = \frac{V_L}{V_L^+} = \frac{2Z_L}{Z_L + Z_0} \tag{2.85}$$

2.1.6.1 Multiple Reflection

The discussion on transmission lines has so far mainly dealt with the case where a wave propagates along a transmission line and as the wave reaches the terminating load is more or less reflected. This scenario will now be extended to include the source that drives the circuit, see figure 2.15. Here the interface between the source and the transmission line is commonly termed the sending end and consequently the interface between the line and load is termed the receiving end.

As for the load, the impedance of the source Z_S may be represented by a reflection coefficient Γ_S and in terms of reflection of waves the source will behave identically with the load, that is an incident wave (travelling in the negative direction with respect to the z-axis) will be reflected, resulting in a wave travelling in the positive z direction.

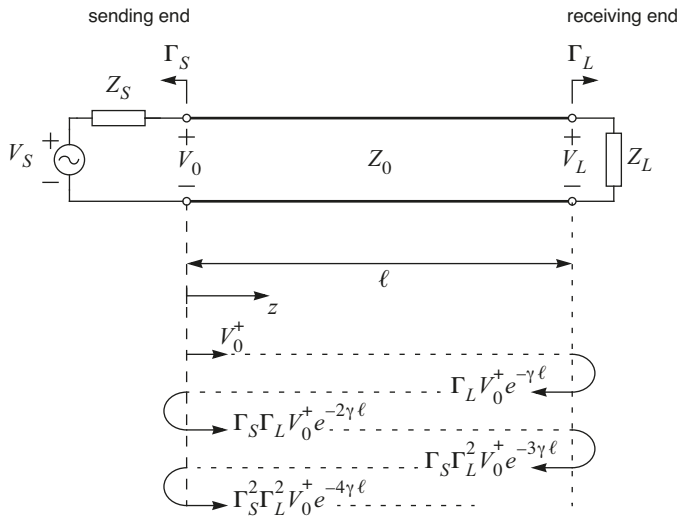


Figure 2.15 Transmission line with reflection both at the sending and receiving ends.

Thus, on a low-loss line with mismatch at both ends, there will be *multiple reflection* of waves that will 'bounce' back and forth along the line as illustrated in figure 2.15 until the amplitude of the wave has decreased to an insignificant level. This means that transfer of power from the source to the load has a complicated dependency on the interaction between the line length, source impedance and load impedance. However, multiple reflection can be avoided if either the sending end or the receiving end is properly terminated ($\Gamma = 0$).

A complete derivation of the expression for the power transmission from the source to the load is out of the scope for this text but the basic idea is to consider the primary wave (V_0^+ in figure 2.15) transmitted by the source before considering the effects of reflection. This wave is the wave that is obtained when the transmission line is assumed to be infinitely long, i.e. it presents an input impedance equal to its characteristic impedance Z_0 which gives

$$V_0^+ = V_S \frac{Z_0}{Z_S + Z_0} \tag{2.86}$$

Thus, the total wave, leaving the sending end is a sum of several waves such that

$$\begin{aligned} V_{0T}^+ &= V_0^+ \cdot (1 + \Gamma_S \Gamma_L V_0^+ e^{-2\gamma\ell} + \Gamma_S^2 \Gamma_L^2 V_0^+ e^{-4\gamma\ell} + \ldots) = \\ &= V_0^+ \cdot \frac{1}{1 - \Gamma_L \Gamma_S e^{-2\gamma\ell}} \end{aligned} \tag{2.87}$$

After a few more steps the following useful expression is found:

$$P_L = P_S e^{-2\alpha\ell} \frac{(1 - |\Gamma_S|^2)(1 - |\Gamma_L|^2)}{\left|1 - \Gamma_S \Gamma_L e^{-2\gamma\ell}\right|^2} \tag{2.88}$$

where P_S denotes the *available power from source* and equals

$$P_S = \frac{|V_S|^2}{8\mathrm{Re}[Z_S]} \tag{2.89}$$

which is the power delivered from the given source to a load that is (conjugately) matched to the source. Here, the source voltage V_S is given by its peak value.

Equation (2.88) can be explained by four sources of transmission losses: line loss, mismatch loss at the sending end, mismatch loss at the receiving end and a residual term arising from multiple reflections. The last term can be positive or negative and varies sharply with small variations in the length of the line, i.e. with small changes in frequency. This creates several drawbacks from a system point of view.

2.1.7 Standing Wave and Standing-Wave Ratio

The voltage or current at a point z on a transmission line is in the general case the sum of two waves travelling in opposite directions and with unequal amplitudes. This scenario is known to create a standing-wave pattern along the transmission line.

Consider the voltage at a given position z along a lossless line,

$$\begin{aligned} V(z) &= V_0^{+} e^{-j\beta z} + V_0^{-} e^{j\beta z} = \\ &= V_0^{+} e^{-j\beta z} \cdot (1 + \Gamma_0 e^{j2\beta z}) \end{aligned} \tag{2.90}$$

Noting that the magnitude of the factor $V_0^{+} e^{-j\beta z}$ is constant with z the magnitude of the voltage $V(z)$ will change in accordance with $\left|1 + \Gamma_0 e^{j2\beta z}\right|$ only which represents the voltage *standing-wave pattern*. Thus, maxima and minima of $V(z)$ will occur for

$$\begin{aligned} &\Gamma_0 e^{j2\beta z} = |\Gamma_0| \Rightarrow \arg(\Gamma_0) + 2\beta z = 2n\pi \Rightarrow \\ &z = \frac{n\pi}{\beta} - \frac{\arg(\Gamma_0)}{2\beta} = n\frac{\lambda}{2} - \frac{\arg(\Gamma_0)}{2\beta} \end{aligned} \tag{2.91}$$

where n is an integer. From (2.91) it can be concluded that the distance between any two successive maxima or minima is half a wavelength.

The ratio between the maximum and minimum value of $V(z)$ is termed the *standing-wave ratio* (denoted by SWR or ρ)

$$\begin{aligned} \text{SWR} &\equiv \frac{\text{maximum voltage or current}}{\text{minimum voltage or current}} = \\ &\equiv \frac{|V_{max}|}{|V_{min}|} = \frac{|I_{max}|}{|I_{min}|} = \frac{1 + |\Gamma_0|}{1 - |\Gamma_0|} \end{aligned} \tag{2.92}$$

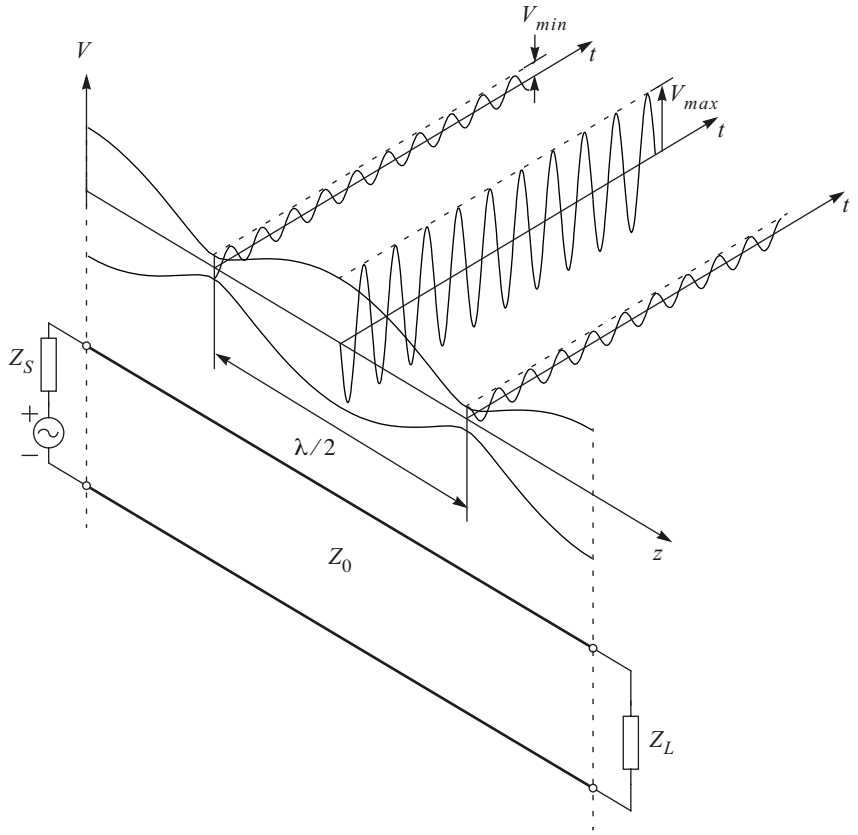


Figure 2.16 Standing-wave pattern in a lossless line.

In figure 2.16 a standing wave pattern is exemplified together with the time-domain signal at the minima and maxima of the pattern. Note that only the voltage standing-wave pattern is shown. The corresponding current standing-wave pattern is in anti-phase with the voltage pattern such that maxima of the current pattern coincide with the minima of the voltage pattern and vice versa.

The above analysis on standing waves applies to lines with no or negligible loss. The standing-wave ratio SWR was defined for this case as the ratio of the maximum to the minimum value of the voltage or current along the line. When the line has significant loss the SWR is of little use as the standing-wave pattern will change with distance.

2.2 Transmission Lines Resonators

To round of this chapter the transmission line used as a resonator will be investigated and the relationships between the line resonator and lumped resonant circuits presented in the previous chapter will be identified. Here a quarter-wavelength resonator terminated with a short-circuit will be investigated, see figure 2.17.

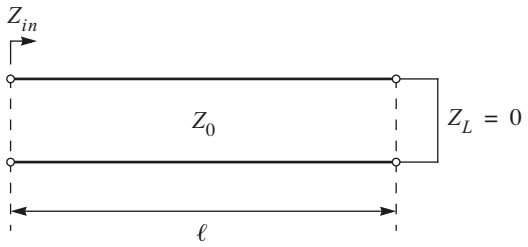


Figure 2.17 Short-circuit quarter-wavelength transmission line.

We will first investigate the case when the line is lossless. The input impedance for this circuit is obtained from equation (2.66) with $Z_L = 0$

$$Z_{in} = Z_0\left(\frac{Z_L + jZ_0\tan\beta\ell}{Z_0 + jZ_L\tan\beta\ell}\right) = jZ_0\tan\beta\ell \tag{2.93}$$

To identify the similarities between transmission line resonators and lumped resonators the input impedance for each circuit can be compared in the frequency domain. Thus, for the transmission line we have

$$Z_{in} = jZ_0\tan\beta\ell = jZ_0\tan\frac{2\pi\ell}{\lambda} = jZ_0\tan\frac{\omega\ell}{v_p} \tag{2.94}$$

This expression is best understood by plotting the reactance as a function of frequency where free-space parameters are assumed, that is $v_p = c$. The length is $\ell = 0.075\,\text{m}$ and $Z_0 = 50\Omega$.

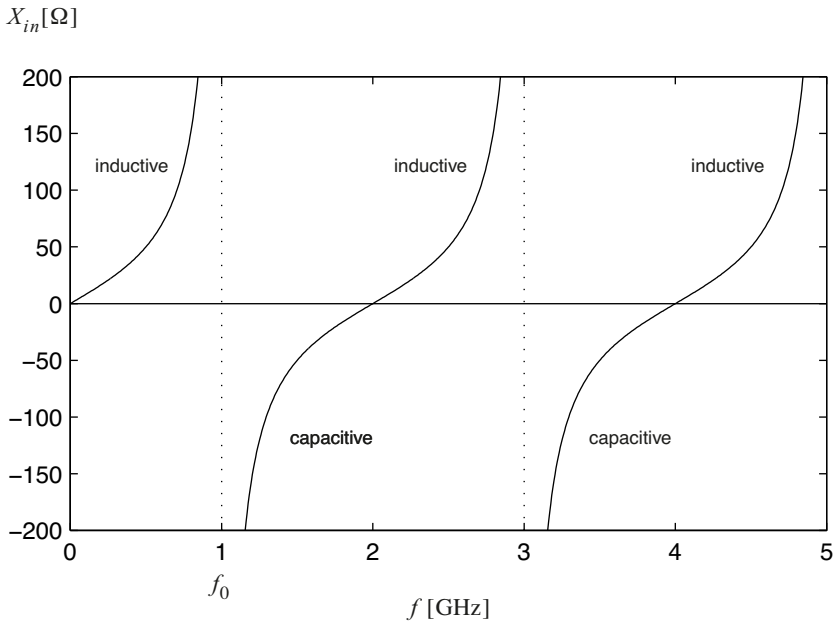


Figure 2.18 Reactance of 1GHz quarter-wavelength line resonator.

For frequencies below 1GHz the reactance of the resonator is inductive and for frequencies above (up to 2GHz) the reactance is capacitive and at 1GHz it is infinite. Thus, qualitatively this circuit has some resemblance with a parallel resonant circuit as shown in figure 2.19.

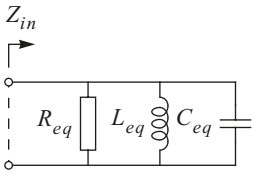


Figure 2.19 Parallel resonant circuit.

The purpose of the following analysis is find the equivalent parallel resonant circuit for the quarter-wavelength line resonator.

The frequency behaviour of the transmission line will be investigated in the proximity of the resonant frequency, that is at $\omega = \omega_0 + \Delta\omega$, thus

$$\begin{aligned} Z_{in} &= jZ_0 \tan\frac{\omega\ell}{v_p} = jZ_0 \tan\frac{(\omega_0+\Delta\omega)\ell}{v_p} = jZ_0 \tan\frac{\omega_0+\Delta\omega}{v_p}\ell \\ &= jZ_0 \tan\left(\frac{\pi}{2}+\frac{\Delta\omega\ell}{v_p}\right) \approx \frac{jZ_0}{-\sin\left(\frac{\Delta\omega\ell}{v_p}\right)} \approx -jZ_0\frac{v_p}{\Delta\omega\ell} \\ &= -jZ_0\frac{\lambda\cdot\frac{\omega_0}{2\pi}}{\Delta\omega\lambda/4} = -jZ_0\frac{2}{\pi}\frac{\omega_0}{\Delta\omega} \end{aligned} \tag{2.95}$$

For the parallel circuit (without a parallel resistor R_{eq}) we have

$$\begin{aligned} Z_{in} &= \frac{j\omega L_{eq} \cdot \frac{1}{j\omega C_{eq}}}{j\omega L_{eq} + \frac{1}{j\omega C_{eq}}} = \frac{j\omega L_{eq}}{1-\omega^2 L_{eq} C_{eq}} = \frac{j\omega L_{eq}}{1-\omega^2 L_{eq} C_{eq}} \\ &= \frac{j(\omega_0+\Delta\omega)L_{eq}}{1-(\omega_0+\Delta\omega)^2 L_{eq} C_{eq}} \approx \frac{j(\omega_0+\Delta\omega)L_{eq}}{1-\omega_0^2 L_{eq} C_{eq} - 2\Delta\omega\omega_0 L_{eq} C_{eq}} \\ &= \frac{j(\omega_0+\Delta\omega)}{-2\Delta\omega\omega_0 C_{eq}} \approx -j\frac{1}{2\Delta\omega C_{eq}} \end{aligned} \tag{2.96}$$

By equating (2.95) and (2.96) the equivalent capacitor is obtained:

$$-jZ_0 \frac{2}{\pi}\frac{\omega_0}{\Delta\omega} = -j\frac{1}{2\Delta\omega C_{eq}} \Rightarrow C_{eq} = \frac{\pi}{4\omega_0 Z_0} \tag{2.97}$$

Since $X_C = -X_L$ at resonance the equivalent inductance becomes

$$\omega_0 L_{eq} = \frac{1}{\omega_0 C_{eq}} \Rightarrow L_{eq} = \frac{4Z_0}{\omega_0 \pi} \tag{2.98}$$

Finally, a lossy line should be considered to identify the value of R_{eq}. The approach is similar to previous analysis but now with an attenuation constant α included. This complicates the derivation substantially and only the result is given here (a complete derivation is found in [1]):

$$R_{eq} = \frac{Z_0}{\alpha\ell} = \frac{Z_0}{\alpha\lambda/4} = \frac{2\beta Z_0}{\alpha\pi} \tag{2.99}$$

Thus, a Q-factor can be found for the line resonator:

$$Q = \frac{R_{eq}}{\omega_0 L_{eq}} = \frac{\beta}{2\alpha} \tag{2.100}$$

and, furthermore, relating to the parameters of the lumped circuit model for the line with low loss

$$\begin{aligned} \frac{1}{Q} &= \frac{2\alpha}{\beta} = \frac{1}{\omega_0\sqrt{LC}}\left(R\sqrt{\frac{C}{L}} + G\sqrt{\frac{L}{C}}\right) = \\ &= \frac{R}{\omega_0 L} + \frac{G}{\omega_0 C} = \frac{1}{Q_S} + \frac{1}{Q_P} \end{aligned} \tag{2.101}$$

where Q_S represents the quality factor of the series elements in the lumped circuit model for the transmission line whereas Q_P represents the shunt elements. Note that the Q-factor is independent of the length of the line and thus it should be regarded as a general quality factor for the transmission line.

The analysis presented above to find the equivalent parallel resonant circuit is based on the behaviour of the circuits in the vicinity of the resonant frequency only. Thus, the question that remains to be answered is how well the line resonator emulates the parallel resonant circuit over a wider frequency range. This is best illustrated by plotting the absolute value for the impedance of both circuits. Using the same parameters as above, that is $v_p = c$, $\ell = 0.075\,\text{m}$ and $Z_0 = 50\Omega$, with the addition of $Q = 10$, the result is shown in figure 2.20. The curves are very close from DC to about $1.5f_0$ but beyond this range the line resonator will have multiple resonant peaks with the first one appearing at $3f_0$.

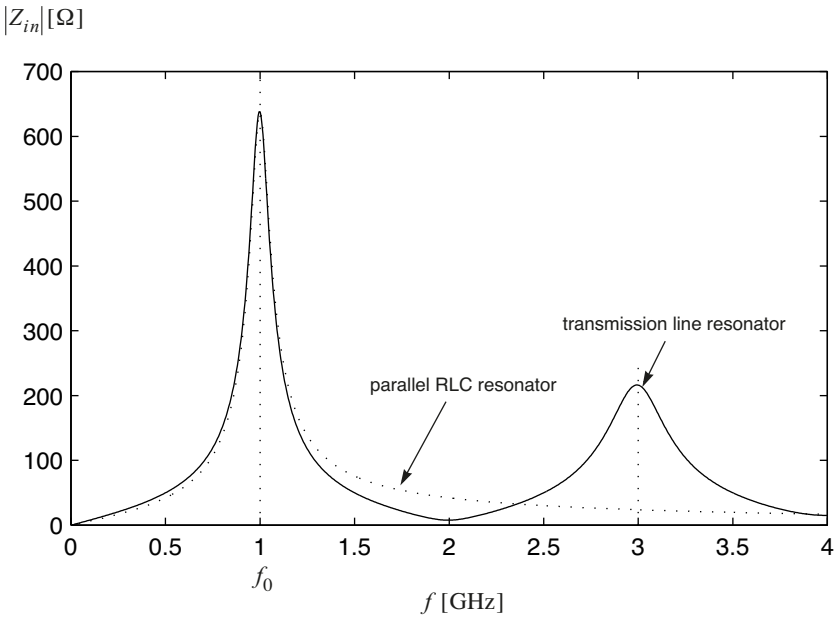


Figure 2.20 Reactance of 1GHz quarter-wavelength line resonator.

As a final remark it is important to note that it is not only the short-circuited quarter-wavelength line that is useful as a resonator. Repeating the analysis above will reveal that a quarter-wavelength line terminated with an open-circuit will behave as series resonant circuit, a half-wavelength line with an open-termination will behave as a parallel resonant circuit and so on.

2.3 References

[1] D. M. Pozar, Microwave Engineering, 2nd edition, Wiley, 1998.

Chapter 3

Passive Components

In the discussion of resonant circuits ideal circuit elements were assumed. This assumption of course keeps the basic analysis on a reasonable level. Unfortunately, the behaviour of real components deviates more and more from their ideal counterparts as the frequency is increased. For example, for a sufficiently high frequency a capacitor may very well behave as an inductor. The causes are found in limitations of the physical realisation and characteristics of the materials that are used.

Transmission lines were also discussed in a previous chapter without taking any consideration to implementation aspects. However, in contrast to lumped components the non-ideal behaviour of a transmission line does not so much lie in the line itself but rather in the interface, that is, when transmission lines are terminated (e.g. to ground or open-circuit) and when they are connected together in junctions or in corners. However, as the frequency becomes higher dispersive effects will show up which complicates the modelling of transmission lines significantly. These effects are, however, not discussed here and it is assumed that all travelling waves are transverse electromagnetic waves (TEM waves) without any longitudinal field component. With this assumption the physical design of transmission line circuits is quite straightforward although it may involve more or less complicated design equations and tables. Two commonly used geometries for transmission lines are depicted in figure 3.1.

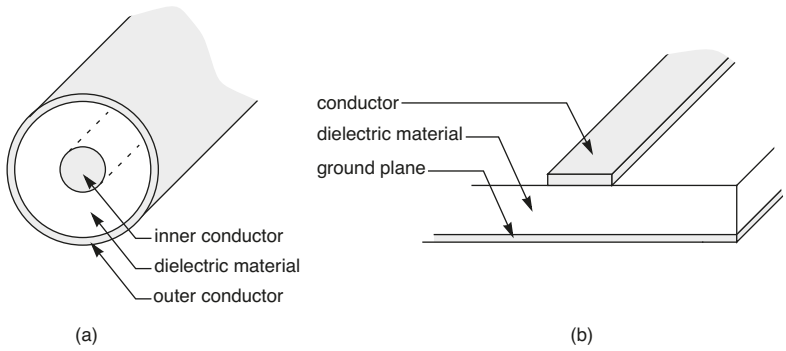


Figure 3.1 Two commonly used transmission line geometries (a) coaxial line (b) microstrip line.

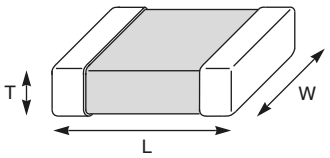
The purpose of this chapter is to discuss real lumped components, their behaviour and how they can be modelled. Furthermore, design formulas are presented for the most common transmission line geometries and parameters for various conductor and substrate materials are given. Finally, transmission line discontinuities as junctions, terminations and corners will be discussed briefly.

3.1 Lumped Components

In contrast to transmission lines, that are considered as distributed circuit elements, the lumped components are assumed to be infinitely small. This approximation holds as long as the wavelength is much longer than the physical dimensions of the components. However, even though a component can be considered as lumped and not distributed, the physical size of the component itself and the parts that resides therein will still introduce, what is referred to, parasitic effects that can be modelled with ideal circuit elements such as inductors and capacitors. The reactive parasitic effects will combined and together with the intrinsic component form a resonant circuit. Therefore, a resonance frequency can be found which is usually termed the *self-resonance frequency*, SRF of the component. The SRF is an important parameter that is used to specify the useful frequency range for a lumped component.

It is readily seen that to obtain small parasitic effects (and high self-resonance frequency) small components with very short leads should be used. For this purpose surface mount design (SMD) components exist in a number of standard sizes, given by size codes in mil (1/1000 inch) or mm units. Examples are given in figure 3.2 where dimensions are given for various sizes. However, the values given for the height T should just be considered as an example. Depending on the nature of the component and its value the height may very well be larger, of course still within the specification of the standard.

Another driver for small components is the continuous miniaturisation of electronic equipment. The penalty for using small components is, however, lower power and voltage handling capabilities which must be considered in the design work.



		Dimensions in mm		
size code (mm)	size code (mil)	L	W	T
1005	0402	1.0	0.50	0.4
1608	0603	1.6	0.8	0.5
2012	0805	2.0	1.2	0.6
3216	1206	3.2	1.6	0.6

Figure 3.2 Various surface mount design (SMD) sizes.

3.1.1 Resistors

A real resistor can be modelled with the equivalent circuit found in figure 3.3. The actual resistance is surrounded by inductive elements partly originating from the leads but also from the actual resistance, which physical realisation may very well resemble the structure of an inductor, e.g. a spiral. Resistors with long leads can have several nH in series inductance (L_{lead}) and the intrinsic inductor L_S may be in the same order of magnitude. The capacitive elements may arise from the material used but it also depends on the structure. Here, the parasitic capacitance C_P can be on the order of 1pF. From figure 3.3 one parallel and one series resonance circuit can be identified. Therefore, considering the values for the parasitic elements that were exemplified a resistor may exhibit a resonance frequency as low as 1GHz or even lower. Of course, in the vicinity of the resonance frequency and certainly above, the resistor is useless.

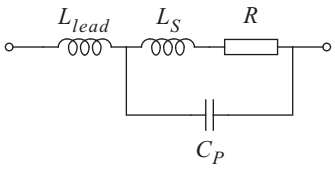


Figure 3.3 Equivalent circuit for a real resistor.

If we do not consider any particular application one of the most common techniques to manufacture resistors are based on a carbon-composition material shaped as a cylinder with leads at each end. The carbon-composition material is made of densely packed carbon granules and between each pair of granules there is a parasitic capacitor which will add up to 1pF or less in parasitic capacitance for the whole resistor. This disqualifies the carbon-composition material for use in high frequency applications.

The parasitic capacitance can be significantly reduced if a resistive material such as carbon is deposited as a thin film onto a ceramic cylinder. The carbon film is then spiralised to obtained the desired resistance by scribing a spiral pattern on the film, see figure 3.4. The spiral shape will inevitably result in a parasitic inductance but the parasitic capacitance being some tenth of pF will still dominate the reactive behaviour. Using a metal film instead of carbon can reduce the parasitic capacitance further and the resistor will be useful for high frequency applications.

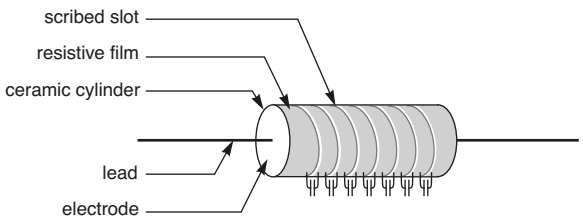


Figure 3.4 Principle of resistor with thin-film resistive material on ceramic cylinder. The distributed nature of the parallel capacitance is illustrated with capacitors between the spiral turns.

Thick-film techniques is used in SMD resistors. A resistive paste is applied to the surface of a high grade ceramic substrate using screen printing. The resistance value is determined by the mix of various components in the resistive paste as well as the physical size of the resistive layer. The value is trimmed to within tolerance, with laser cutting. SMD resistors are typically manufactured using this technique and the combination of a small intrinsic capacitance and the small SMD package makes it very useful for high frequency application. One example of a SMD resistor and its geometry is given in figure 3.5.

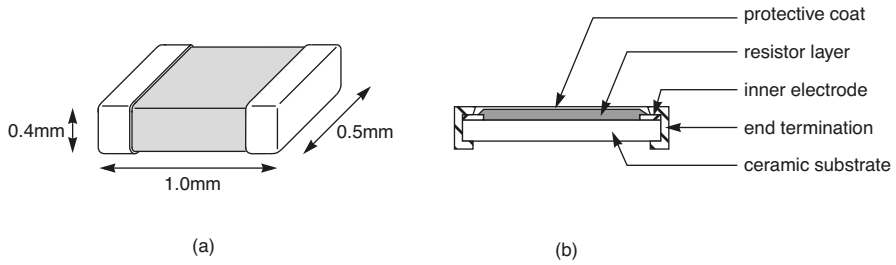


Figure 3.5 (a) Physical dimensions of SMD resistor with size code 0402 (b) cross section of resistor.

Example 3.1 High frequency behaviour of SMD resistor

A series of high frequency SMD resistors in 0603-size have the following parasitics:

$C_P = 50\text{fF}$ and $L_{lead} = 0\text{nH}$, $L_S = 0.4\text{nH}$

Plot the resistive and the reactive part of the impedance from 1MHz to 10GHz for such a resistor with 10Ω, 100Ω, 1kΩ, 10kΩ resistance. Normalise the resistance and the reactance with the DC value of the resistance.

To plot the impedance, first derive an expression for the impedance based on figure 3.3. By inspection of this circuit we obtain

$$Z_R = sL_{lead} + \frac{R + sL_S}{1 + sC_P(R + sL_S)}$$

and plots of resistance and reactance are shown below. From these plots it obvious that high resistor values cannot be used for high frequencies. Here, the parasitic capacitance will dominate as the frequency is increased and finally bypass the resistor completely. Also, there is a similar tendency for very low resistor values where the reactance of the series inductance will have a significant influence at high frequencies. To summarise, even though very small parasitics are found for a 0603-sized resistor they will still influence the characteristics of the component in the lower GHz-range.

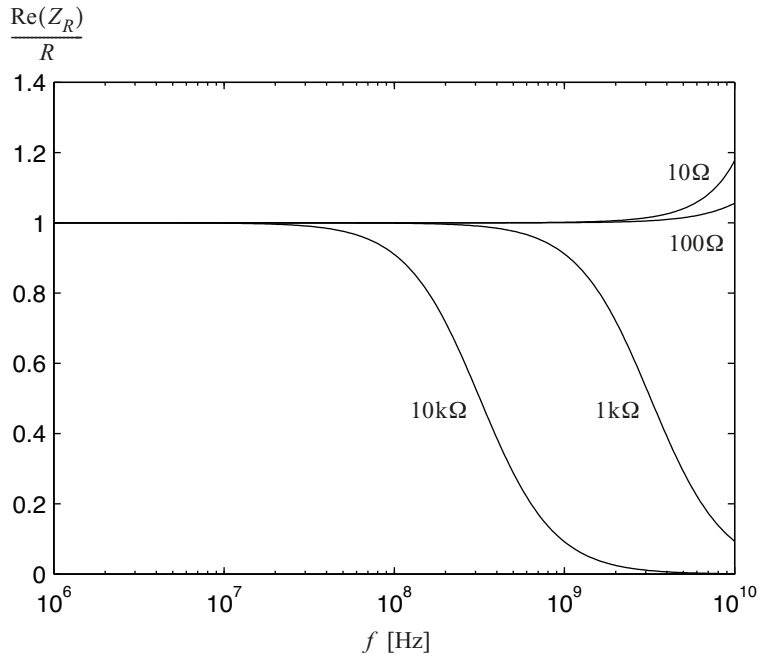


Figure 3.6 Normalised resistance of high frequency SMD resistors as a function of frequency.

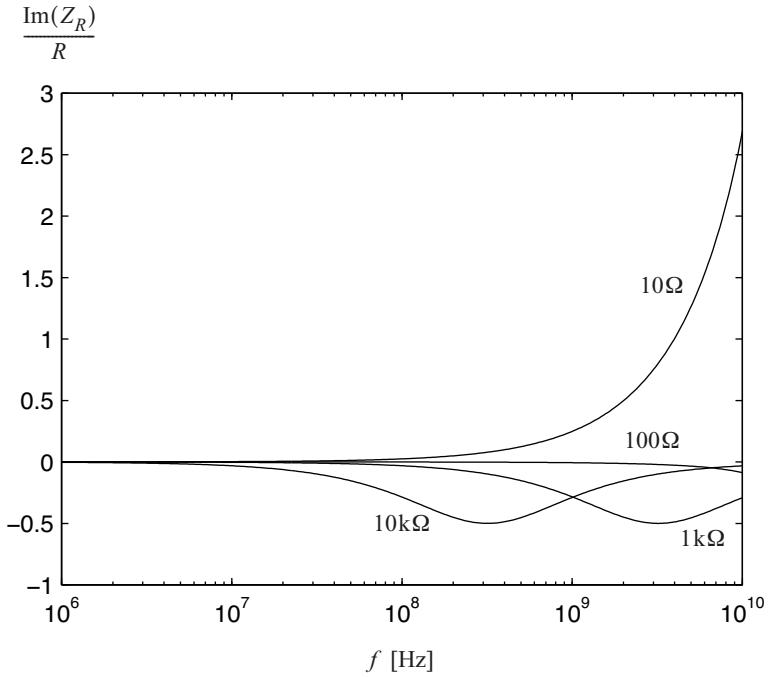


Figure 3.7 Normalised reactance of high frequency SMD resistors as a function of frequency.

3.1.2 Capacitors

The most simple capacitor consists of two plates separated by a thin dielectric material, see figure 3.8. This is the principle used for most capacitors, RF capacitors in particular.

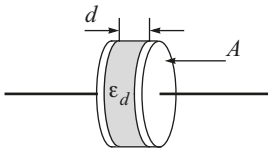


Figure 3.8 Simple capacitor with two plates separated by a dielectric disk.

The capacitance is given by

$$C = \varepsilon_d A/d = \varepsilon_0 \varepsilon_r A/d \tag{3.1}$$

where ε_0 is the permittivity of free-space, ε_r the relative permittivity for the dielectric material, A is the area of one plate and d the distance between the plates. For RF applications the size of the plates cannot be too large as it will give more parasitic effects. Also, if a small footprint is desired (e.g. SMD) the allowed size is certainly limited. Instead very thin dielectric sheets with a large permittivity can be used, which in turn makes it possible to stack several layers of plates to increase the area further. To obtain a very large capacitance another principle similar to that of a regular battery can be used with anode and cathode and an electrolyte in between. However, this principle is limited to low frequency applications.

To investigate high frequency performance of capacitors, an equivalent model including the parasitic effects should be used, see figure 3.9. As for the resistor, there are inductive elements due to the leads, but also the internal structure of the capacitor give rise to some inductance. The parallel parasitic capacitance, C_P, is typically quite small and can be ignored in most cases as it originates from the capacitive coupling between the electrodes, excluding the intrinsic capacitor. This assumes of course that we only consider frequencies below the first resonance frequency given by the series resonance circuit that is obtained when we disregard C_P.The series resistance, R_S, also termed the *equivalent series resistance* (ESR), is due to the internal loss in the capacitor and determines the quality factor of the component. In practice all the elements in this equivalent model are frequency dependent. However, for frequencies in the low GHz range the inductors and the capacitor can be assumed to be constant while The ESR gradually increases with frequency starting well below the GHz range. The lead inductance can be several nH whereas the intrinsic inductance L_S is typically somewhat smaller. For this reason the resonance frequency for capacitors are typically quite low and inevitably decreases with increasing capacitance values.

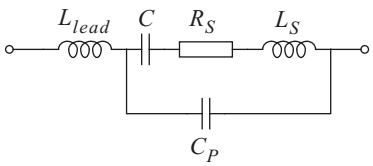


Figure 3.9 Equivalent circuit for a real capacitor.

A plethora of various capacitors exists. The film capacitor uses a plastic film as a dielectric material between the electrodes.The actual plastic material that is used (e.g. polyester, polypropylene etc.) determines the properties in terms of internal loss, temperature sensitivity, insulation resistance etc. Plastic film capacitors are not used in high frequency applications, they are more suitable in low frequency analog circuits and as decoupling capacitors in digital circuits.

The electrolytic capacitor have some similarities with a battery in that there is an anode and a cathode with a dry or wet electrolyte. They can be manufactured will very high capacitance values, up to several tenths of Farads but they suffer from high resistance. The use of capacitors with wet electrolyte is limited to energy reservoirs as found in power supplies whereas the dry counterparts are of more general purpose nature. However, as with batteries this capacitor has a polarity which must be considered in the design work. If applied with a reversed voltage the capacitor may explode.

For high frequency applications the ceramic capacitor is widely used. Here one or several layers of a ceramic material are printed with an electrode material which in turn are connected to the terminals, see figure 3.10. Ceramic capacitors are divided into three classes. Class 1 uses a material with a low dielectric constant and they have very low sensitivity in terms of temperature, frequency and voltage. They also exhibit a very low loss and they are suitable for high frequency applications. Class 2 uses materials with a high dielectric constant which enables more compact designs but they suffer form a nonlinear dependency on temperature, voltage and frequency. They are used in noncritical applications such as for decoupling. The class 3 ceramic capacitor is based on a semiconducting material where the capacitance between granules constitutes the capacitance. It is similar to class 2 in performance but they have lower breakdown voltage.

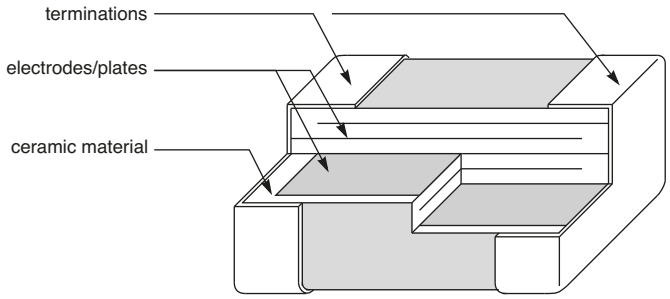


Figure 3.10 The structure of an SMD multi-layer ceramic capacitor.

Example 3.2 High frequency behaviour of SMD capacitor

A high frequency SMD capacitors in 0603-size with a nominal value of $C = 10\text{pF}$ has the following parasitics:

$C_P = 150\text{fF}$ and $L_{lead} = 0\text{nH}$, $L_S = 1.0\text{nH}$, $ESR = 0.5\Omega$

Plot the absolute value of the impedance from 10MHz to 50GHz. Note that in practice the parasitics will not be constant over such a wide frequency range.

The expression for the impedance is found by inspection of figure 3.9:

$$Z_C = sL_{lead} + \frac{R_S + sL_S + 1/sC}{1 + sC_P(R_S + sL_S + 1/sC)}$$

and the result is shown figure 3.11.

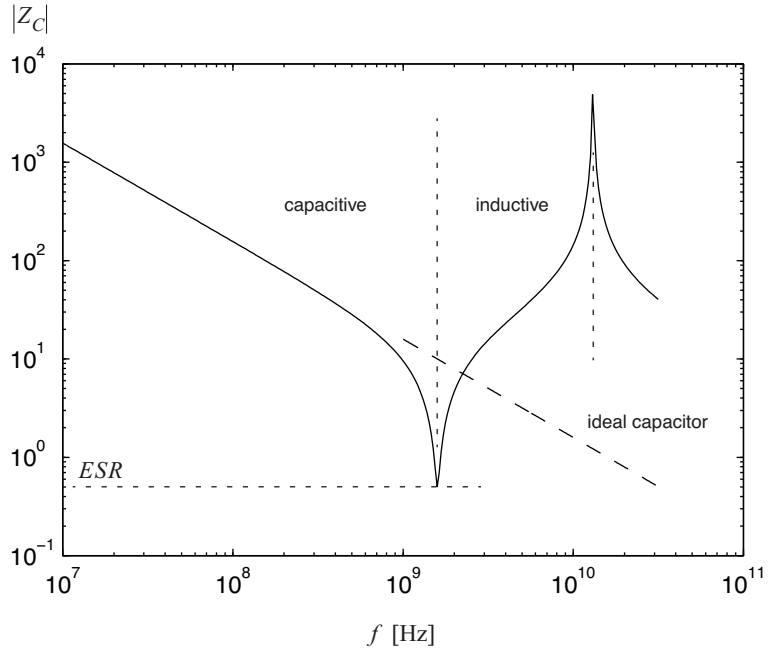


Figure 3.11 Absolute value of impedance for a high frequency SMD capacitor as a function of frequency.

From the plot it is evident that this capacitor will not be useful in the GHz range. A deviation from the slope of the ideal capacitor (dashed line) starts at around 700MHz and a series resonance dip (due to C and L_S) is found just below 2GHz where the impedance equals ESR. Yet another resonance effect is found above 10GHz now with C_P and L_S in resonance.

3.1.3 Inductors and Transformers

Even a straight piece of wire exhibit an inductive behaviour although its inductance is rather small per unit length of wire. Its inductance is given by

$$L = \ell \cdot 2{\times}10^{-7} \cdot \left(2.3 \cdot \log\left(\frac{4\ell}{d}\right) - 0.75\right) \tag{3.2}$$

where ℓ is the length of the wire and d the diameter (in metres).

With the straight wire as a starting point the inductance can be increased by increasing the magnetic flux linkage between different parts of the wire. This is best done by forming a loop with one or multiple turns as shown in figure 3.12.

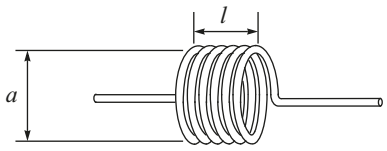


Figure 3.12 Simple air-core inductor.

The inductance of this coil with an air core is approximately given by [1]

$$L = \frac{10\pi\mu_0 a^2 n^2}{10l + 4.5a} \tag{3.3}$$

where l is the length of the wire, a the diameter of the coil, n the number of turns and μ_0 the permeability of free space. This formula is valid as long as the length is greater than the radius. For the special case of a single loop of wire the inductance is approximately

$$L = \mu_0 \frac{a}{2} \cdot \left(\ln\left(\frac{8a}{d}\right) - 2 \right) \tag{3.4}$$

where a the diameter of the coil and d the diameter of the wire.

The inductance of a coil can be increased further by inserting a magnetic high-permeability core which effectively increases the magnetic flux. This approach has some shortcomings though as the a magnetic core will introduce loss. More important, the permeability is frequency dependent and approaches the permeability for air as the frequency is increased. In practice, magnetic cores can be used up to 1GHz or less depending on the application.

An equivalent model for a real inductor is shown in figure 3.13. The resistor R_S models the losses in the inductor wire and if present the losses in the core. As for the wire the loss is small at low frequencies but as the frequency is increased the Skin effect (see section 3.2.1) will manifest itself by increasing the loss. The capacitor C_P models the capacitive coupling between the closely spaced turns of a multi-turn inductor and therefore can be quite large. As for the other components the inductor will also have a small parasitic inductance due to the leads or electrodes.

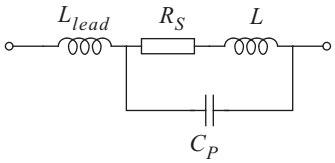


Figure 3.13 Equivalent circuit for a real inductor.

Example 3.3 High frequency behaviour of SMD inductor

A high frequency SMD inductor in 0603-size with a nominal value of $L = 10\text{nH}$ has the following parasitics:

$C_P = 150\text{fF}$ and $L_{lead} = 0\text{nH}$, $R_S = 3\Omega$

Plot the absolute value of the impedance from 10MHz to 50GHz. Note that in practice the parasitics will not be constant over such a wide frequency range, especially not the series resistance R_S.

The expression for the impedance is found by inspection of figure 3.13:

$$Z_L = sL_{lead} + \frac{R_S + sL}{1 + sC_p(R_S + sL)}$$

and the result is shown figure 3.14.

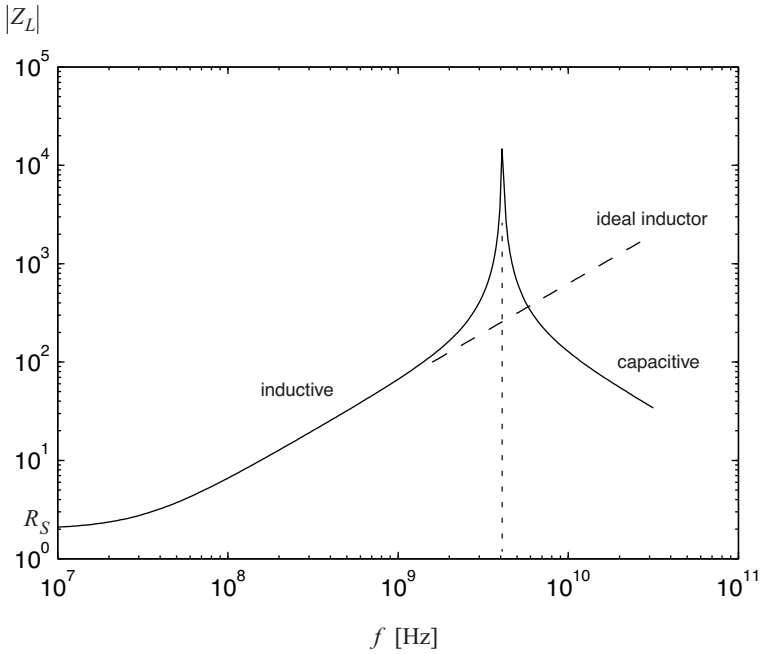


Figure 3.14 Absolute value of impedance for a high frequency SMD inductor as a function of frequency.

The self-resonance frequency is found to be 4GHz and beyond this point the inductor is not of much use. A deviation from the slope of the ideal inductor (dashed line) already starts at around 1GHz.

The Q-factor for inductors is usually quite limited and, moreover, it is frequency dependent. In figure 3.15 the qualitative frequency behaviour of the Q-factor is given with X_T denoting the total reactance of the component. At low frequencies the DC resistance will determine the losses whereas the reactance value will increase linearly with frequency. With increasing frequency the Skin effect will gradually increase the loss (see section 3.2.1) which slows down the Q-factor curve until it reaches a peak where the parasitic capacitor, C_P, causes the reactance to increase with the same rate as the loss. As the frequency is increased further the effective reactance will decrease due to the

parasitic capacitor which in turn will result in a decreasing Q-factor which finally becomes zero at SRF. Note that this way of reasoning only applies for an inductor that is supposed to be used as an inductor. If the inductor is part of a resonance circuit the parasitic capacitance may be included (if the tolerances are adequate) in the total capacitor of the resonance circuit and the Q-factor of the resonance circuit will therefore not degrade due to the parasitic capacitor.

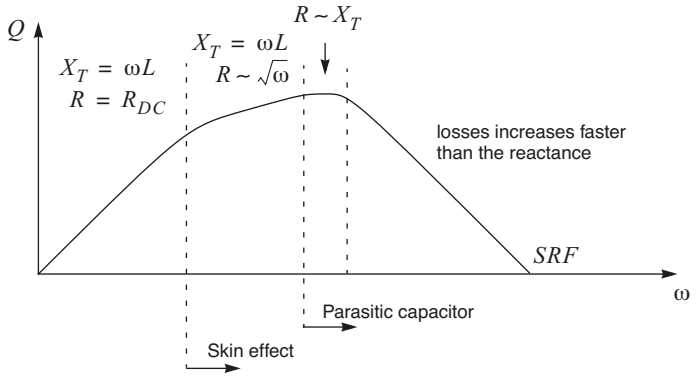


Figure 3.15 Qualitative frequency behaviour of Q-factor for an inductor.

The inductor is the only component that can be easily implemented in the lab. A piece of insulated copper wire wounded around a cylinder-shaped object will make a fairly good air-core coil, see figure 3.12. Although this is not a mass-production friendly approach air core inductors are actually manufactures in SMD sizes where the coil is fixed by an acrylic jacket as shown in figure 3.16a. The jacket also give the inductor a well-defined geometry with a flat top that makes them suitable for automatic placement. These inductors provide Q-factors well beyond 100 in the GHz range and self-resonance frequencies above 5GHz for 10nH and less.

Instead of using a jacket to fix the geometry of the inductor the wire can be wounded around a dielectric core with a well-defined size that also provides flat electrodes on the sides of the core, see figure 3.16b. These inductors have a somewhat lower Q-factor than for the air-core inductors but may still reach a value of 100.

Thin-film techniques can also be used in a fashion very similar to the construction of resistors where a low resistive planar and spiral-shaped inductor is grown on a ceramic substrate. The inductor geometry is accurately patterned by means of photolithography and therefore the inductance value becomes very accurate. These inductors does, however, not exhibit the high Q-factor of the wire-wounded inductors described previously and typically peaks at 50 or less.

Transformers are not as common as inductors at higher frequencies, say, above a few hundred MHz. To obtain a large coupling factor transformers are typically wounded around toroid-shaped magnetic cores which limits the maximum frequency of operation. Surface mount transformers with ferrite cores typically have a bandwidth well below 1GHz and at RF frequencies the insertion loss is significant, on the order of 0.5dB. However, transformers with a ceramic (non-magnetic) core are available with up to 2GHz bandwidth (3dB insertion loss).

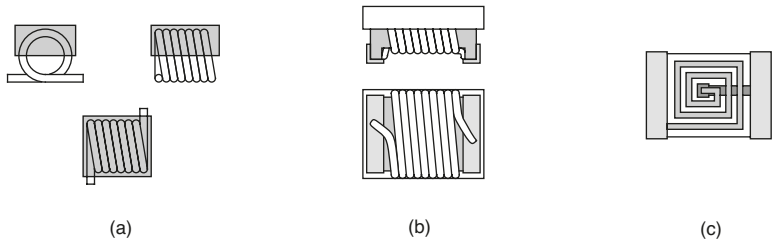


Figure 3.16 Various inductor implementations: (a) air core (b) dielectric core (c) thin-film inductor on ceramic substrate.

3.2 Transmission Lines

From the theoretical discussion on transmission lines in the previous chapter it is not easily seen how we may go by to implement these lines. However, from the schematic diagrams used to illustrate transmission lines one may conceive that two parallel wires could constitute a transmission line. Although not a practical solution this is indeed one possible geometry. Here, we refer to the geometry of the transmission line as the cross-section layout of the line.

In figure 3.17 various geometries are illustrated. The coaxial geometry is the most common one to realise transmission lines as cables. For printed circuit boards (PCBs) the microstrip geometry is widely used and in some cases also the stripline geometry. The characteristic impedance of these lines of course varies with how these geometries are scaled, e.g. the ratio between the radius of the conductor and the shield in a coaxial line determines the characteristic impedance together with the dielectric material between the conductor and the shield.

Considering transmission lines on PCBs, one reason for the microstrip geometry being widely used is the fact that it only requires two conductive layers separated with a dielectric material.This should be compared with the stripline technique which requires three conductive layers and two dielectric layers. Also, using the coaxial transmission line might seem to be excessive if it is compared with e.g. the simplicity of a twisted pair of wires.

However, the microstrip structure and twisted pair of wires are examples of what are usually referred to as open-boundary transmission lines. That is, the geometries do not limit the extension of the electric and magnetic fields to within a finite area. This means that signals carried by such transmission lines very well might couple to other lines in the vicinity and thereby cause undesired interference. The coaxial line on the other hand is a closed-boundary geometry which keeps the fields within the outer conductor. Finally, the stripline might be viewed as something in between, a semi-closed boundary line as the fields will not extend beyond the two ground planes but still may couple to other transmission also residing between the two ground planes.

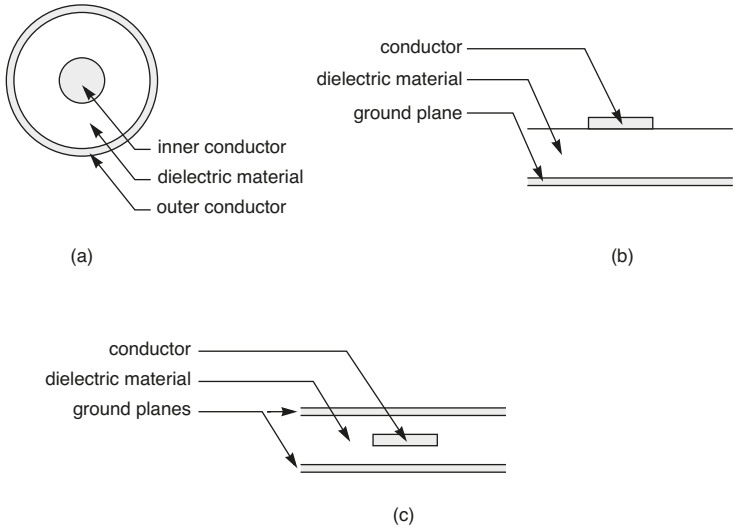


Figure 3.17 Various transmission line geometries: (a) coaxial (b) microstrip (c) stripline.

3.2.1 Conductor and Substrate Materials

The materials used to implement a transmission line controls the properties of the line. The permittivity of a dielectric material affects the phase velocity, v_p, and the characteristic impedance Z_0. The dielectric material also introduce loss. The conductor material, however, only contribute loss and does not influence v_p and Z_0.

For low frequencies the loss in the conductor is only given by the resistivity or is reciprocal parameter, the conductivity σ. Thus for a given line structure and dimensions it is straightforward to calculate e.g. the resistance per unit length of transmission line or the attenuation constant α. However, as the frequency is increased the *Skin effect* will manifest itself by increased loss in the line. The Skin effect is a result of the varying magnetic flux inside the conductor which effectively will push the current to a narrow region below the surface of the conductor. The depth of this region is termed the *Skin depth* and represents the effective depth of the conductive area. In more detail the current density decays exponentially from the surface towards the centre of the conductor as illustrated in figure 3.18. Here it is also shown that the Skin depth is equal to the distance from the surface for which the current density has decreased by a factor e. All this applies if the diameter or cross-sectional width of the conductor is, say, 5 times the Skin depth or more, otherwise a uniform current density can be assumed. The Skin depth is given by

$$\delta_i = \sqrt{\frac{2}{\omega \mu_r \mu_0 \sigma}} \tag{3.5}$$

where ω is the angular frequency, μ_r the relative permeability of the conductor material and μ_0 the permeability of free-space. For the materials used as conductors the relative permeability is equal to unity and thus only the conductivity and frequency is needed to calculate the Skin depth. The conductivity for some metals are given in table 3.1.

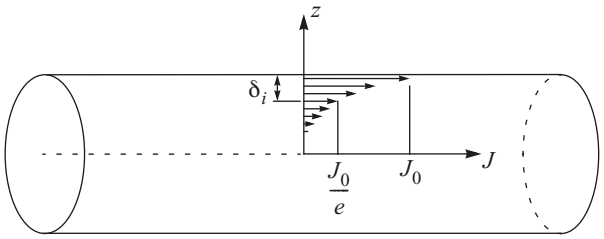


Figure 3.18 Skin effect: current density J as a function of distance from centre of circular conductor.

Table 3.1 Conductivity for some metals.

Metal	Conductivity σ [S/m] (20°C)
Aluminium	3.816×10^7
Copper	5.813×10^7
Gold	4.098×10^7
Silver	6.173×10^7

Example 3.4 Resistance of wire at various frequencies

Calculate the resistance of circular wire made of pure copper with a diameter of 0.1mm and 1 meter in length at 1MHz, 100MHz and 10GHz.

First, at DC there is no Skin effect and we assume that the current density is uniform over the cross-section area of the wire. That is, the conductive area is $\pi r^2 = 7.85\times10^{-9}\text{m}^2$.

At 1MHz the Skin depth becomes $\delta_i = 66\mu\text{m}$. This should be compared with the diameter of the wire which is 100μm. Since they are within the same order of magnitude we can assume a uniform current density over the cross-section area of the wire.That is, the conductive area is $\pi r^2 = 7.85\times10^{-9}\text{m}^2$.

For the other two frequencies the Skin depth becomes significantly less than the diameter and the conductive area will be a thin region below the surface:

100MHz: $\delta_i = 6.6\mu m \rightarrow A = 2\pi r\delta_i = 2.07\times10^{-9} m^2$

10GHz: $\delta_i = 0.66\mu m \rightarrow A = 0.21\times10^{-9} m^2$

Finally, the resistance becomes

$$R(1MHz) = \frac{l}{\sigma \cdot A} = 2.2\Omega$$

$R(100MHz) = 8.3\Omega$

$R(10GHz) = 83\Omega$

The relative permittivity (or dielectric constant) for a substrate material is usually preferred to be high as the phase velocity is decreased from the speed of light in free space by a factor of $\sqrt{\varepsilon_r}$. This in turn means a reduction of the wavelength by the same factor and therefore also the physical length of a transmission line to realise a certain electrical length. However, for very high frequencies (millimetre-wave) the small size may not be advantageous any more as the structures become to small to be manufactured with an adequate accuracy and in this case a low relative permittivity might be desired.

As mentioned above a dielectric material also adds loss and this is usually represented by the loss tangent $\tan\delta$ where δ is called the loss angle. For a dielectric material with low loss the loss tangent is given by

$$\tan\delta \approx \delta \approx \frac{\sigma}{\omega\varepsilon_r\varepsilon_0} \tag{3.6}$$

This is also related to the quality factor or *Q*-factor of the material such that

$$Q_d = 1/\tan\delta \tag{3.7}$$

The relative permittivity and loss tangent for some substrate materials are given in table 3.2 below. These figures should be considered as typical figures as they vary between different manufactures. Also, the parameters are both frequency and temperature dependent.

Table 3.2 Relative permittivity for some substrate materials

Material	ε_r	$\tan\delta$
Alumina (ceramic form - SiO_2)	10	0.0015
Epoxy fibre-glass (composite material)	4	0.02
Fused silica (amorphous form of quartz)	3.8	0.0001
Gallium arsenide (crystal - GaAs)	12.9	0.001
RT Duorid* 5880 (composite material)	2.20	0.0009
Silicon (crystal - Si)	11.9	0.004

* Rogers Corp., Chandler, Arizona

3.2.2 Transmission Line Geometries

3.2.2.1 Coaxial Geometry

The coaxial line, the most commonly used structure for cables, consists of an inner and an outer conductor usually supported by a homogeneous dielectric material such as Teflon or Polyethylene. The geometry is shown in figure 3.19 below together with important parameters that determines the characteristic impedance and the phase velocity.

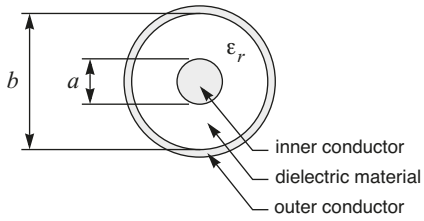


Figure 3.19 Coaxial line geometry.

Since the electrical field completely resides between the two conductors the phase velocity is only controlled by the relative permittivity of the dielectric material, that is,

$$v_p = \frac{c}{\sqrt{\varepsilon_r}} \tag{3.8}$$

Furthermore, it is quite straight forward to show that the characteristic impedance is given by [3]

$$Z_0 = \frac{1}{2\pi}\sqrt{\frac{\mu_0}{\varepsilon_r\varepsilon_0}} \cdot \ln\left(\frac{b}{a}\right) \approx \frac{138}{\sqrt{\varepsilon_r}} \cdot \log\left(\frac{b}{a}\right) \tag{3.9}$$

That is, as far as the dimensions are concerned only the ratio of the diameter of the inner conductor and the inner diameter of the outer conductor controls the characteristic impedance. However, the absolute dimensions are still important. When high power signals are to be transferred the distance between the conductors should be large enough to avoid dielectric breakdown. This occurs when the electrical field strength exceeds the maximum field strength for the dielectric material.

Because the coaxial line is one of the most commonly used structures for high power transmission and transmission over large distances it has served as the basis for determining the standard impedance of 50Ω which is the most widely used interface impedance found in many instruments. If no dielectric material is present (air between the conductors) it can be shown [4] that to obtain maximum power transfer capability the diameter ratio b/a should be 1.65 corresponding to a characteristic impedance of 30Ω whereas if the atten-

uation coefficient is to be minimised the ratio should instead be 3.6 which corresponds to $Z_0 = 77\Omega$. The mean value between these two falls in the neighbourhood of 50Ω.

3.2.2.2 Microstrip Geometry

The microstrip structure is easily adopted on PCB design since it only requires two layers of conducting materials, the ground plane and the plane with transmission lines, see figure 3.20.

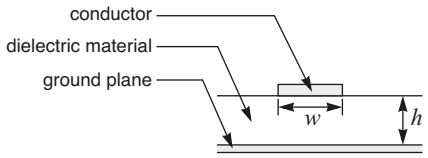


Figure 3.20 Microstrip geometry.

It is not as easy to calculate the properties of this geometry as was the case for the coaxial line. Not only do we have non-uniform distribution of the electric and magnetic fields due to its complex geometry but also because the fields penetrates both air and the substrate. Thus, to calculate e.g. the phase velocity we cannot simply use the relative permittivity of the substrate or air. Instead, a weighted mean value between the two should be used, here referred to as the effective permittivity ε_{eff}. Many formulas exists for microstrip calculations but they are all more or less empirical and are only accurate for a limited range of impedances. Then, on the other hand, the required accuracy does not need to be any better than the process that is used to manufacture the PCBs.

The characteristic impedance is easily found from figure 3.21 as a function of substrate height, h, and conductor width, w. Notice that these curves assume a zero thickness conductor which is an adequate approximation in most cases. These curves have been calculated using the design equations in reference 5. To determine the phase velocity the curves in figure 3.22 should be used which shows the effective permittivity as a function of the ratio w/h and the relative permittivity of the substrate. As w/h ratio increases ε_{eff} approaches ε_r. The reason is of course a larger portion of the electrical fields between the conductor and the ground plane will only penetrate the substrate.

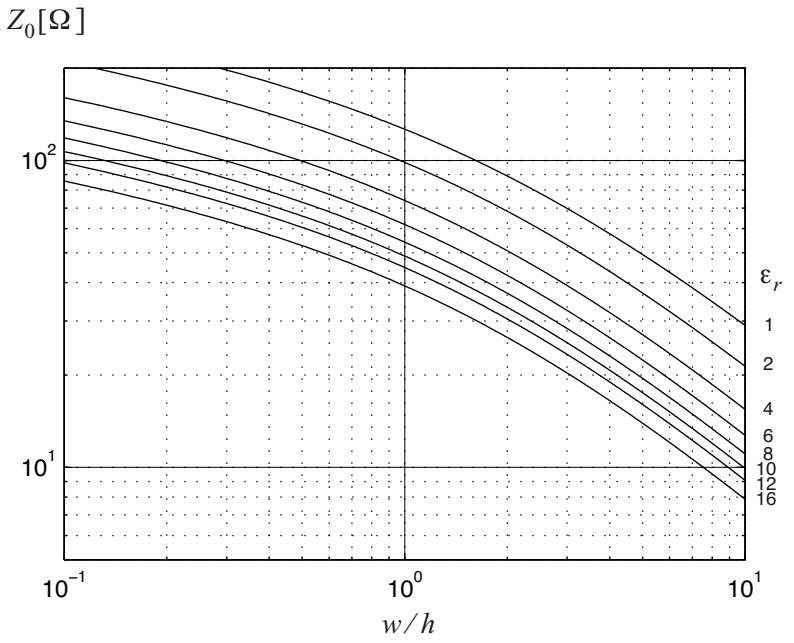


Figure 3.21 Characteristic impedance versus width/height ratio and relative permittivity of the substrate assuming zero thickness conductor.

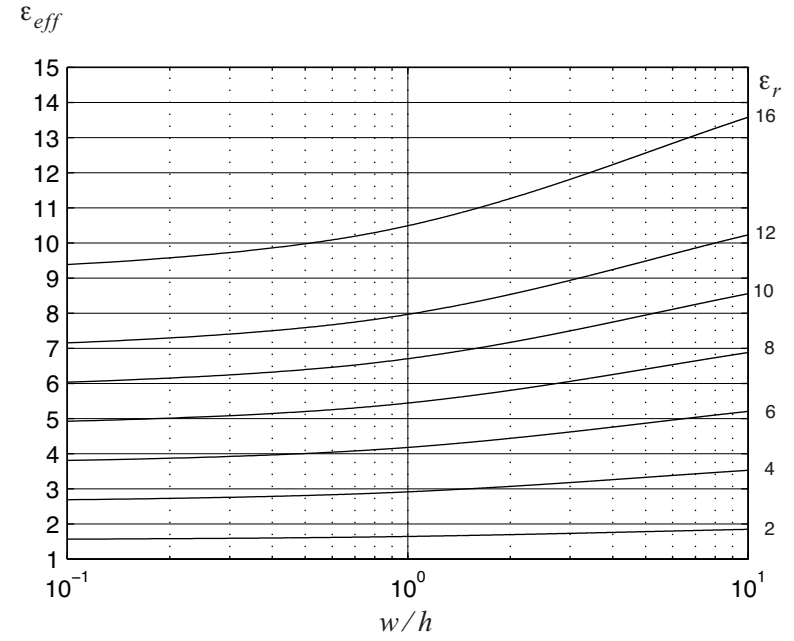


Figure 3.22 Effective permittivity versus width/height ratio and relative permittivity of the substrate assuming zero thickness conductor.

Example 3.5 Design of a microstrip line

Calculate the width of a microstrip line to obtain a characteristic impedance of 20, 50 and 100Ω. The line should have an electrical length of one wavelength at 1GHz. The substrate is 1mm thick and is made of epoxy fibre-glass, see table 3.2.

The substrate material has a relative permittivity of 4 and for the desired characteristic impedances figure 3.21 gives the following result:

20Ω: $w/h = 7.3 \rightarrow w = 7.3\text{mm}$

50Ω: $w/h = 2.1 \rightarrow w = 2.1\text{mm}$

100Ω: $w/h = 0.5 \rightarrow w = 0.5\text{mm}$

The wavelength in free-space is used as a starting point to calculate the physical length of the transmission line. At 1GHz we have that $\lambda_0 = 0.3\text{m}$ and therefore

20Ω: $w/h = 7.3 \rightarrow \varepsilon_{eff} = 3.4 \rightarrow \lambda = \lambda_0/\sqrt{\varepsilon_{eff}} = 0.163\text{m}$

50Ω: $w/h = 2.1 \rightarrow \varepsilon_{eff} = 3.1 \rightarrow \lambda = 0.170\text{m}$

100Ω: $w/h = 0.5 \rightarrow \varepsilon_{eff} = 2.8 \rightarrow \lambda = 0.179\text{m}$

3.2.3 Discontinuities

The previous treatment of transmission lines, where the lines had a continuous cross-section geometry with field patterns being the same along its complete extension, is a simplification which will cause more or less errors in practice. In a real circuit a transmission line will not be separate entity of a straight line. The fact that these lines will interface to other lines with different widths, have turns to change direction and maybe be terminated to ground or with an open end makes accurate modelling substantially more complicated. The effects of these discontinuities can be taken into account using models that are more or less empirical. The field patterns around these structures are to complicated to be handled purely analytical. Some models can be found in e.g. [6] but since these models still improve by continuous research the most advanced and up-to-date models are usually found implemented in CAD tools. Here, the most common effects of microstrip discontinuities are discussed briefly.

3.2.3.1 Open-Circuit Termination

The open-circuit termination shown in figure 3.23 is quite commonly used in matching networks and filters where the open-circuit impedance is transformed to a complex impedance by the transmission line. Unfortunately, the abrubt termination will not cause an equally abrupt end of the fields. Instead, the fields close to the termination will be distorted and fringing fields will extend beyond the end of the transmission line. This can be modelled as a capacitor at the termination as shown in figure 3.23b or more conveniently as an extension of the line, see figure 3.23c. This model is adequate for low fre-

quencies, that is when the line extension is substantially less then the wavelength. Typically, this line extension lies in a region from say 0.1 to 1 times the height of the substrate depending on the w/h .

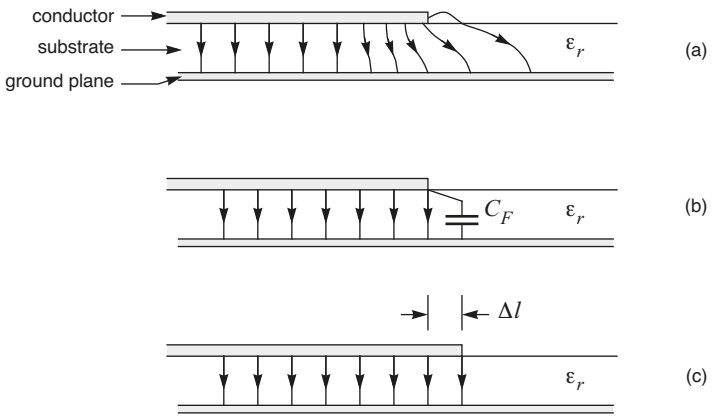


Figure 3.23 Microstrip line with open-circuit termination:
(a) electric field (b) equivalent capacitance (c) equivalent line extension.

3.2.3.2 Corners

Corners and bends are inevitable in practical transmission line circuits. Plain transmission lines might need to have several corners to be able to use the PCB area efficiently and in some transmission line structure, e.g. directional couplers and filters, corners simply are a must to obtain the desired function.

The ideal corner should of course be transparent. For an abrupt corner, this is not possible to achieve although it can be designed to have a minimal effect. Also, if there is enough space, a rounded corner can be used. If the radius of such a curvature is greater than 3 widths of the line the effect will be very small.

To understand the qualitative behaviour of the corner the structure in figure 3.24a is used as a starting point. The square corner, defined by the structure between the reference planes A_1 and A_2 , are connected to two perpendicular and uniform transmission lines. The corner can be modelled by a network as shown in figure 3.24b with inductors, L_c , representing the current and stored magnetic energy and the capacitor, C_c , representing the charge and the stored electric energy. The same network is also an equivalent circuit for a short (relative to the wavelength) transmission line with a characteristic impedance equal to

$$Z_{corner} = \sqrt{L_c / C_c} \tag{3.10}$$

Since the square corner presents an excess capacitance compared with the uniform transmission line the effective characteristic impedance of such an equivalent line will be lower than that of the connecting lines. Also, an equivalent line length is given by

$$\Delta l = \frac{c}{\sqrt{\varepsilon_{eff}}} \cdot \sqrt{L_c C_c} \tag{3.11}$$

Here, L_c and C_c are unknown parameters describing the square corner and can be obtained using appropriate models or CAD tools.

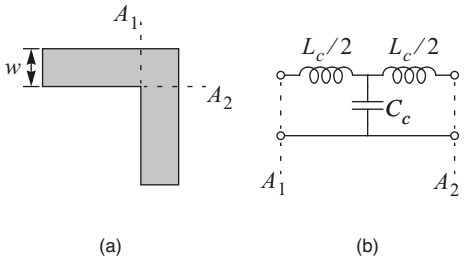


Figure 3.24 The geometry of a square corner and equivalent models.

There is not much that can be done to significantly reduce the excess line length that is introduced by the corner. This is usually not a problem since the corner is most often a part of the line that is very much longer than Δl and therefore the line can be compensated to account for it. The characteristic impedance, on the other hand, is of prime concern. If Z_{corner} can be increased to Z_0 of the connecting lines the corner can become almost transparent. This is best done by decreasing C_c as illustrated in figure 3.25. The depth of a suitable mitering cut (y) depends on the w/h ratio of the line and the permittivity of the substrate.

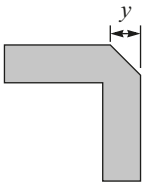


Figure 3.25 Corner geometry compensated with decreased capacitance by mitering the corner.

3.2.3.3 Symmetrical Step

The abrupt transition between two different line widths (and therefore Z_{01} and Z_{02}, as illustrated in figure 3.26a, is yet another very common structure. As for the open-circuit termination there is not abrupt change of the electric and magnetic fields around the transmission lines but rather a smooth transmission which, effectively, might be modelled by the T-network in figure 3.26b. This assumes that the transition region is small compared with the wavelength. The inductors represent the distorted current flow in the transition region whereas the capacitor represents the increased capacitance due to the fringing electric fields at the transition edges. A coarse approximation gives that the excess capacitance can be modelled as if the wider line ends at

reference plane A_2 rather than at the actual transition A_1, see figure 3.26c, and therefore the effect of this discontinuity can be compensated by altering the line widths accordingly.

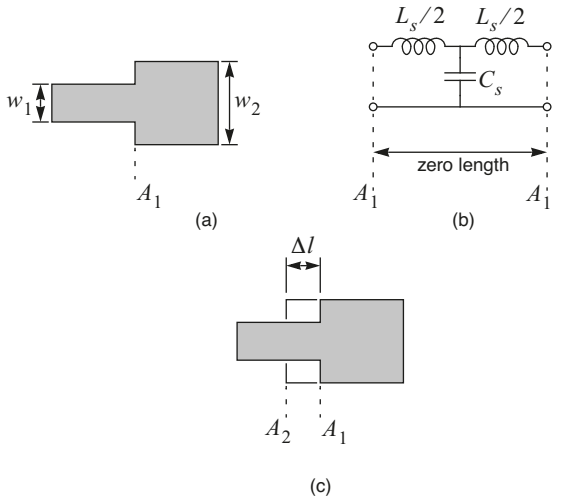


Figure 3.26 (a) The geometry of a symmetrical step, (b) equivalent model and (c) equivalent extension of wide line.

3.2.3.4 T-Junction

The last structure to be discussed in this chapter is the T-junction where three lines are joined as illustrated in figure 3.27a. In this particular case there is one main or 'through' line with the same width throughout the junction whereas the shunt line may have a different width. Note the location of the interface between the lines defined by the reference planes A_1, A_2 and A_3. Other reference planes could be devised to separate which part belongs to which line. This is of course only a matter of definition and the equivalent circuit model for the junction will change accordingly. However, the definition given in figure 3.27a is the one that is assumed for most models.

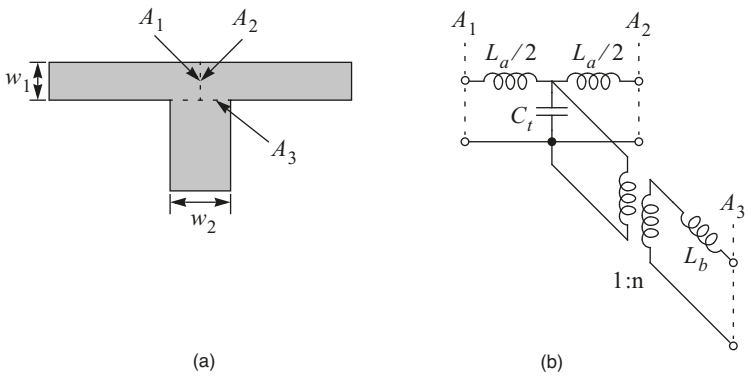


Figure 3.27 The geometry of a T-junction and its equivalent model.

Again, a T-network is used for modelling but now extended with another inductor and a transformer to account for the third line, see figure 3.27b. This model is only valid when the dimensions of the junction is much smaller than the wavelength. The transformer models dispersion effects. If there is no dispersion, n equals unity. When the reference planes are taken as in figure 3.27a the inductors and the capacitor in the equivalent model have negative values.

3.3 References

[1] H. A. Wheeler, "Simple Inductance Formulas for Radio Coils", In *Proceedings of the IRE*, vol. 16, no. 10, October 1928, pp. 1398-1400.

[2] S. Ramo, J. R. Whinnery and T. van Duzer, *Fields and Waves in Modern Radio*, Wiley, 1965, p. 311.

[3] S. A. Maas, *The RF and Microwave Circuit Design Cookbook*, Artech House, 1998.

[4] D. Roddy, *Microwave Technology*, Prentice-Hall, 1986.

[5] E. Hammerstad and O. Jensen, "Accurate models for microstrip computer-aided design", In *IEEE MTT-S International Microwave Symposium Digest*, 1980, pp. 407-409.

[6] G. Matthaei, L. Young and E. M. T. Jones, *Microwave filters, impedance-matching networks, and coupling structures*, Artech House, 1980.

Chapter 4

The Smith Chart

As seen in the previous chapters the mathematical treatment of transmission-lines and resonant circuits shows a tendency to derive intricate expressions. The important relationships between the reflection coefficient, impedance and standing-wave ratio are complex and difficult in general. That is why there is a need for a graphical tool to show the results and even make it possible to graphically solve very lengthy and complex equations.

The most useful graphical tool available to the rf designer is the *Smith Chart,* shown in figure 4.1, that transforms the z-plane into the Γ-plane . The first draft was conceived in the early 1930s by Phillip Smith, at that time an engineer at Bell Laboratories.

This chapter will explain what the chart represents and how it can be used at the rf workbench. In several of the succeeding chapters the Smith Chart will be used in the design of impedance matching networks, amplifiers and other tasks.

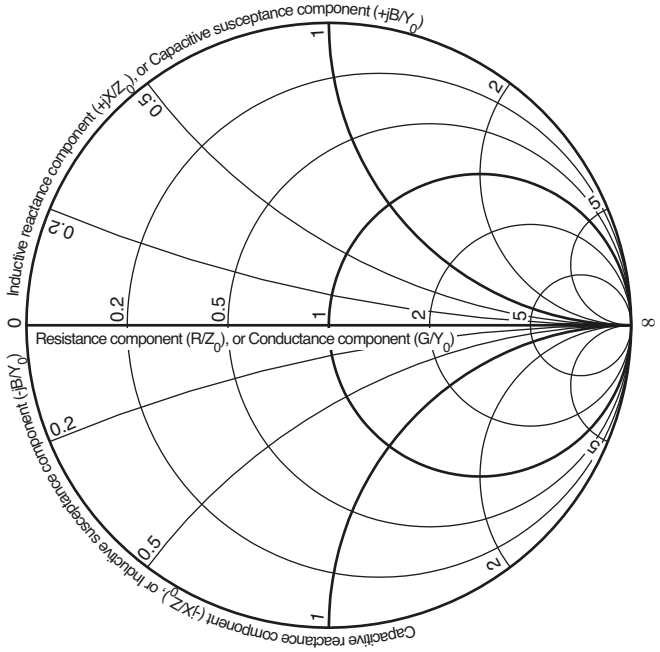


Figure 4.1 The basic Smith Chart

4.1 Construction of the Smith Chart

A complex impedance of the form $Z = R + jX$ can equally be defined by the corresponding reflection coefficient $\Gamma = \rho\angle\phi = \rho e^{j\phi}$. Especially when transmission-lines are involved, the latter is the most convenient parameter to use.

The fundamental relationship between reflection coefficient and normalised impedance at a pair of terminals may, as stated in Chapter 2, be written as

$$\Gamma = \frac{z-1}{z+1} \tag{4.1}$$

The Smith Chart is a mapping between the Γ-plane

$$\Gamma = \rho\angle\phi = p + jq \tag{4.2}$$

and the normalised z-plane

$$z = r + jx \tag{4.3}$$

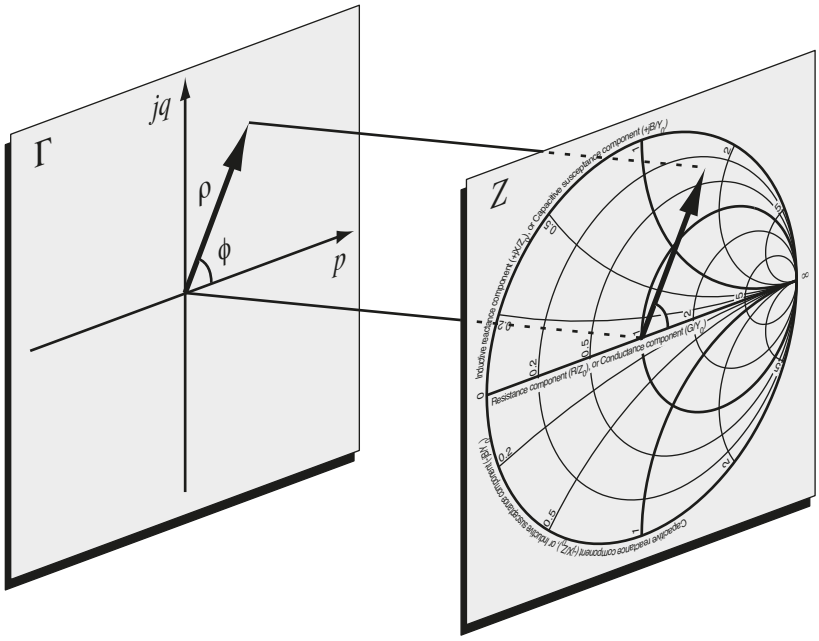


Figure 4.2 The Γ-plane in rectangular coordinates is mapped into the Smith Chart calibrated in impedance units.

To show the relationship it is first necessary to sort out the real and imaginary parts, therefore equation (4.1) is written as

$$p + jq = \frac{r - 1 + jx}{r + 1 + jx} \tag{4.4}$$

When this is separated into its real and imaginary parts, the two resulting equations are

$$\left(p - \frac{r}{r+1}\right)^2 + q^2 = \left(\frac{1}{r+1}\right)^2 \tag{4.5}$$

$$\left(q - \frac{1}{x}\right)^2 + (1-p)^2 = \left(\frac{1}{x}\right)^2 \tag{4.6}$$

Equation (4.5) represents a family of circles whose centres are located at $p = \frac{r}{r+1}$ and $q = 0$, and with radius equal to $\frac{1}{1+r}$.
Similarly, equation (4.6) describes another set of circles whose centres are at $p = 1$ and $q = \frac{1}{x}$, and with radius equal to $\frac{1}{x}$.
The conclusion is that impedances with fixed resistance are mapped into constant resistance circles in the Γ-plane, and impedances with fixed reactance are mapped into constant reactance circles as shown in figure 4.3.

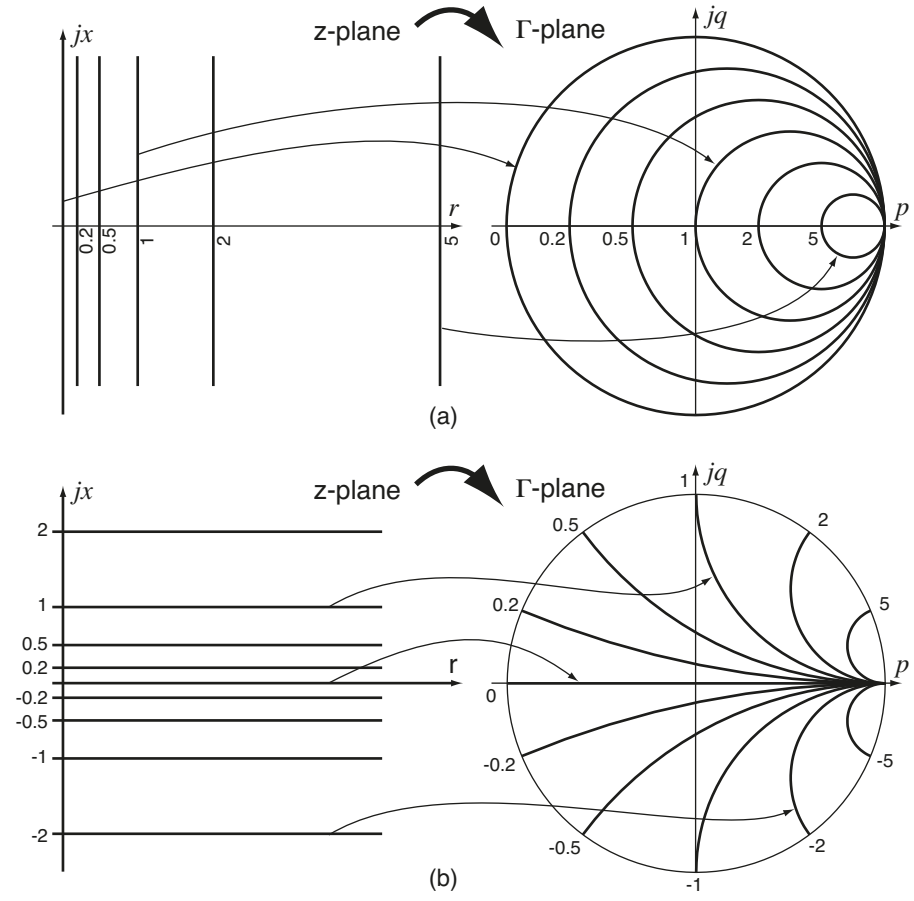


Figure 4.3 a) The mapping of impedances into the Γ-plane. (a) Constant resistance circles and (b) constant reactance circles.

On the Smith Chart the p-axis is calibrated in r-values, instead of being calibrated linearly. The calibration is obtained from equation (4.4) setting x equal to zero:

$$p = \frac{r-1}{r+1} \tag{4.7}$$

The r-scale is seen to extend from zero to infinity for p-values ranging from -1 to $+1$. Generally the chart is used only for load impedances with a positive resistive component. That means that all impedances with positive real part are mapped into the unit circle in the Γ-plane. Thus, the outer limit of the Smith Chart is the circle $|\Gamma| = 1$, which corresponds to maximum mismatching of a lossfree line with an open alternatively short-circuited load or a purely imaginary load.

Also on the Smith Chart, there is no need to show the q-axis as the segments of x-circles are marked directly in the x-values. Additional scales are added around the perimeter of the chart, showing the reflection coefficient angle, transmission coefficient angle and wavelengths along a transmission-line.

Below the chart is a linear scale found for reading the reflection coefficient magnitude. Further scales are added for determining other radially scaled parameters. A complete Smith Chart is shown in figure 4.5.

As it is seen that all impedances with a positive real part are mapped within the unit circle, the consequence is that impedances with a negative real part are mapped into the are outside the unit circle. Negative resistance does never exist in passive components, however in the discussion of stability in amplifiers and in oscillator design the phenomena will show up.

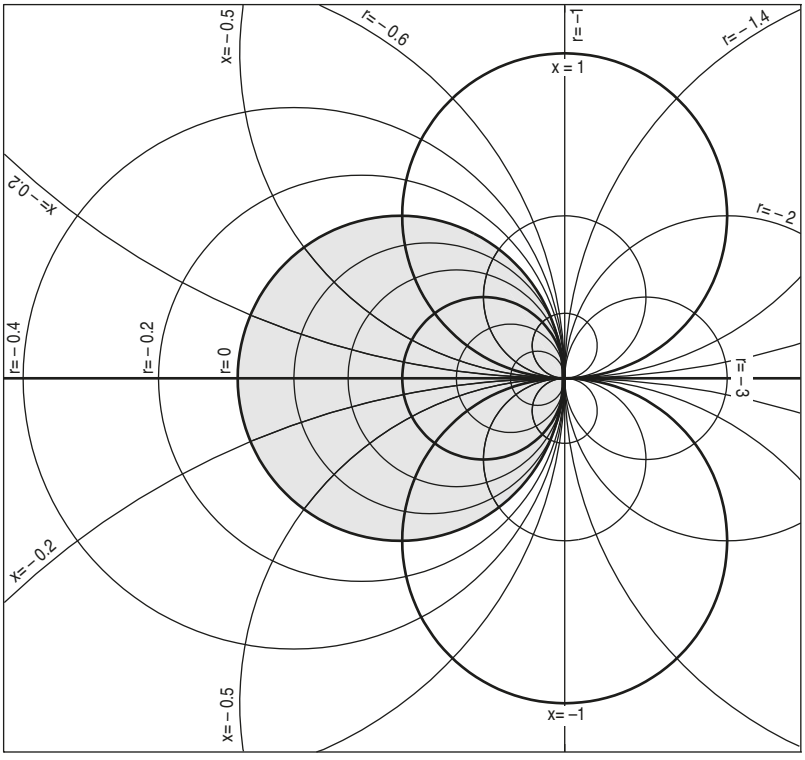


Figure 4.4 Extended Smith Chart showing negative resistance coordinates outside the unit circle.

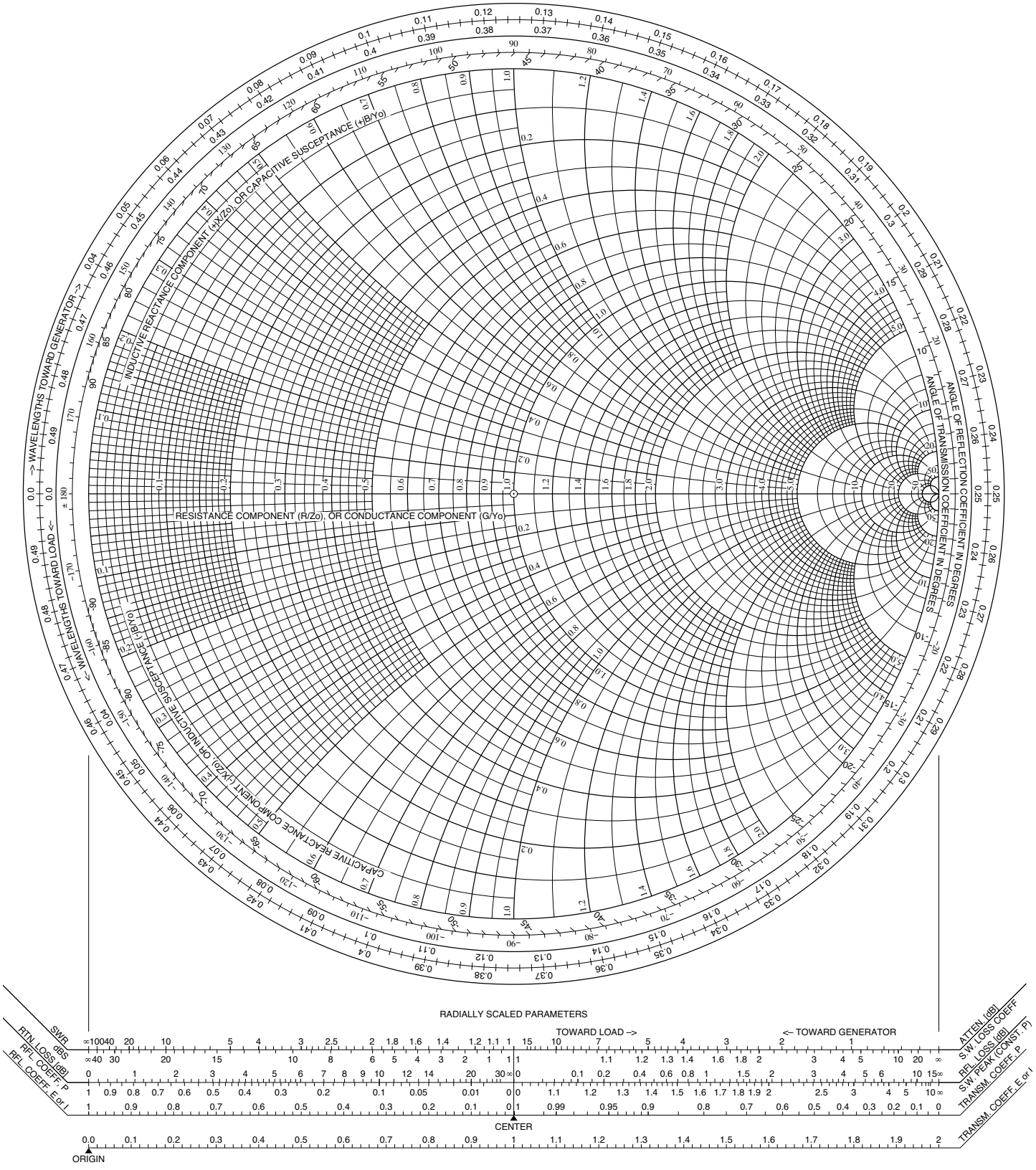


Figure 4.5 The Smith Chart.

4.2 Practical Applications of the Smith Chart

The Smith Chart is basically of use for the transformation $\Gamma = \Gamma(z)$ or $z = z(\Gamma)$. The network analyser is a good example. The instrument measures the reflection coefficient, but when the load impedance is requested, the result can instantly be presented on the screen in a Smith Chart. However, many other parameters related to these quantities can be graphically determined in the chart. This section will describe the most common applications. Indeed, there are so many uses for the chart that several books have been written on the subject. A detailed description of the chart and its applications is given by Phillip H. Smith. [1]

4.2.1 Reflection Coefficient and Electrical Length

If we move the observation point along a transmission-line by a distance d, from the load and in direction to the source, the reflection coefficient will be transformed as stated in section 2.1.4:

$$\Gamma_d = \Gamma_L e^{-2\gamma d} = \Gamma_L e^{-2\alpha d} e^{-j2\beta d} \tag{4.8}$$

This relationship corresponds to a clockwise rotation of the Γ-vector. The angle is found on the inner concentric scale marked "ANGLE OF COEFFICIENT IN DEGREES". An example is shown in figure 4.6 where $\Gamma_L = 0.8\angle 60°$ and $d = 0.35\lambda$. Curve a) shows a lossfree line, whereas curve b) shows a lossy line with the attenuation constant $\alpha = 0.8$.

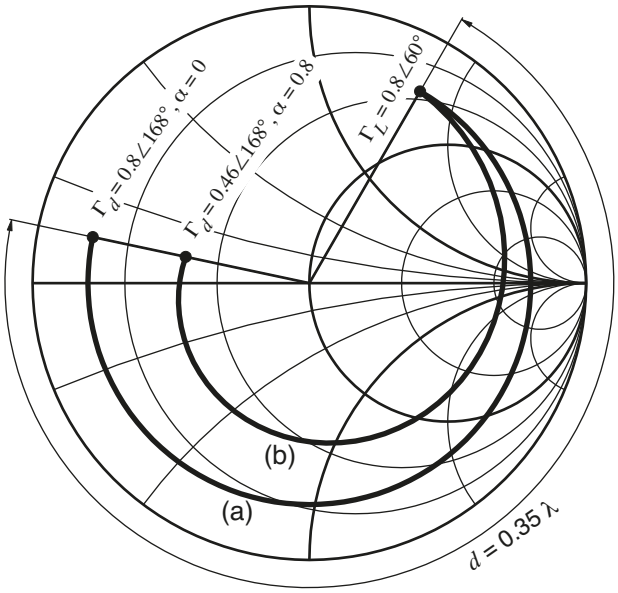


Figure 4.6 Transformation of the reflection coefficient along (a) a lossfree line and (b) a lossy line.

A complete turn, 360°, in the chart corresponds to a 2π change in

$$2\beta d = \frac{4\pi}{\lambda} d$$

that conforms with a line with an electrical length of $\lambda/2$. Analogous will the result of a $\lambda/4$ line be a half turn. This is shown on the outermost circular scale, which is calibrated in the electrical length of the transmission-line. The direction is indicated on the chart by the arrows marked "WAVELENGTHS TOWARD GENERATOR" and "WAVELENGTHS TOWARD LOAD".

The magnitude of Γ determines the distance from the centre of the chart. This is a linear relationship from 0 to 1 that is denoted on the linear scare beneath the chart and marked "RFL COEFF, E OR I". An alternative scale gives the corresponding dB relationship, "ATTEN [dB]".

It is clearly seen that the Smith Chart is a very convenient tool to calculate the input impedance of a transmission-line when a known load is connected to the other side:

1. Plot the load impedance or Γ_L on the chart.
2. Determine the electrical length of the line.
3. Rotate the locus of Γ_L around the centre according to the appropriate scales calibrated in wavelengths.
4. If the actual line is loss free, the locus traces an arc with constant radius. On the other hand, if the losses in the line are significant the radius has to be decreased according to the attenuation constant α:

$$|\Gamma_d| = |\Gamma_L| e^{-2\alpha d}$$

4.2.2 Impedance and Admittance

Any plot on the Smith Chart represents a *series* combination of resistance and reactance of the form $z = r + jx$. Note that the scales on the chart are always denoted in normalised quantities. If the impedance variables in the basic expression

$$\Gamma = \frac{z-1}{z+1} \tag{4.9}$$

are replaced by admittance quantities, $z = 1/y$, the locus for the corresponding admittance is found to be

$$\Gamma(y) = \frac{1/y - 1}{1/y + 1} = -\frac{y-1}{y+1} = -\Gamma(z) = e^{j\pi}\Gamma(z) \tag{4.10}$$

This corresponds to a rotation of 180° around the constant $|\Gamma|$-circle. Thus the Smith Chart can be used for admittance quantities as well. That makes a convenient way to do serial-to-parallel-conversion ab shown in figure 4.7. It

is to be remembered that an inductance corresponds to a positive reactance but to a negative susceptance. An inductive impedance is located in the upper half of the chart, while an inductive admittance is found in the lower half.

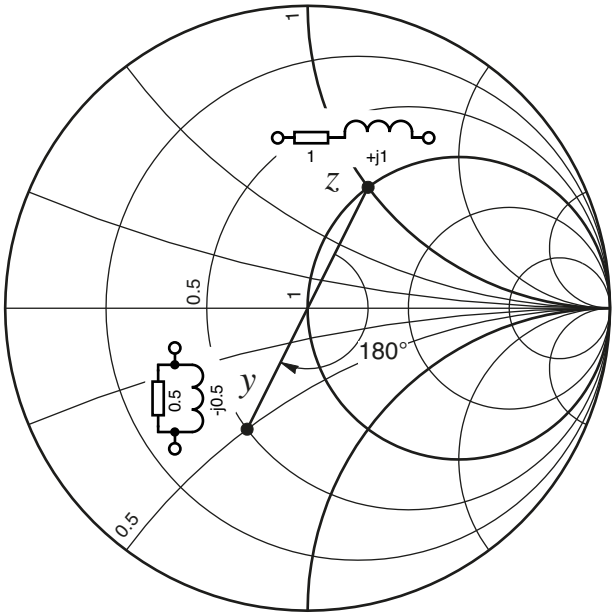


Figure 4.7 Impedance-admittance conversion on the Smith Chart.

The chart can also be used for impedance calculation of series connected circuit elements. When a resistive component is added, the locus of the impedance moves along a constant reactance arc. Addition of a reactive component moves the locus along a constant resistance circle. An example is shown in figure 4.8. Connection in parallel is carried out in the same way, except that admittances instead of impedances are to be added which is illustrated in figure 4.9.

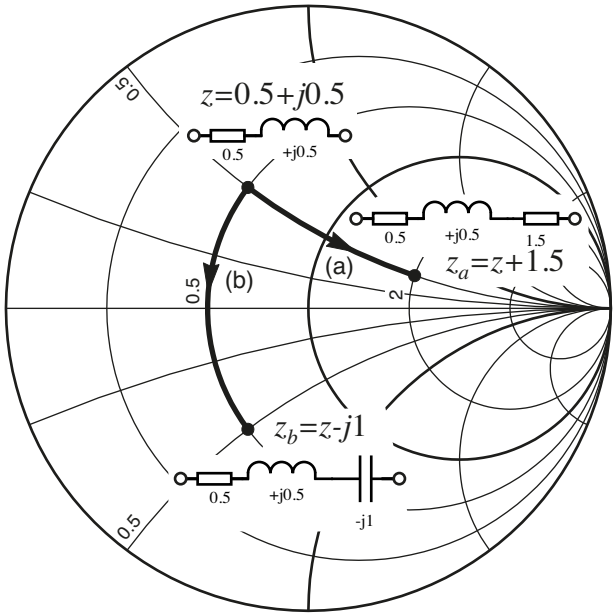


Figure 4.8 Series addition of circuit elements. (a) a resistance $r = 0.5$ and (b) a capacitive reactance $x = -1$.

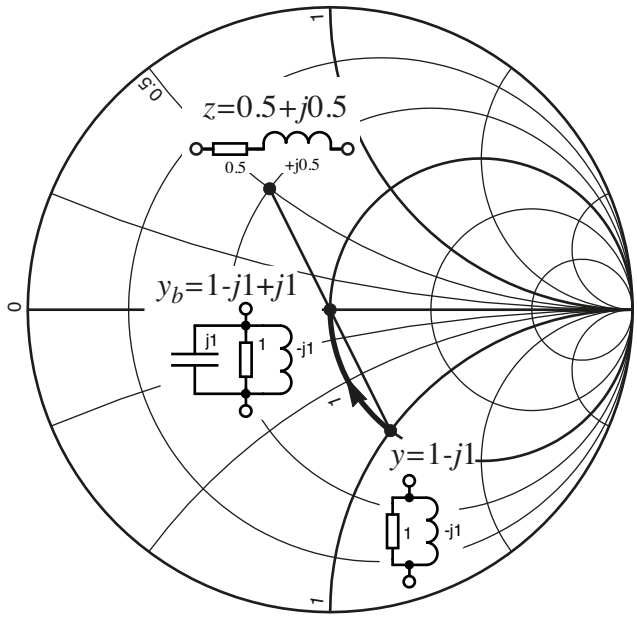


Figure 4.9 Shunt addition of a capacitance with the reactance x = -1 which corresponds to the susceptance b = 1.

When consecutive connections with mixed series and shunt elements are to be calculated, a Smith Chart with superimposed admittance coordinates is very useful. This eliminates the over and over again need for conversion between impedance and admittance. However, it is advisable to keep one's mind on the scales, as there are twice as many it is very easy to get lost.

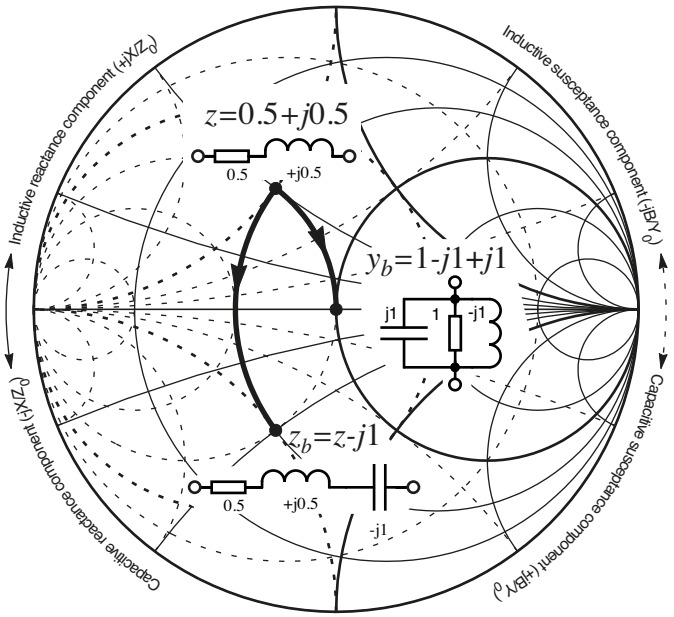


Figure 4.10 The Smith Cart with superimposed admittance coordinates.

4.2.3 The Circuit Q Factor

The Q factor, which is measure of the losses in a circuit, is as stated in Chapter 1 defined as

$$Q = \frac{x}{r} \tag{4.11}$$

for a series circuit. Therefore it is seen that impedances located near the resistance axis corresponds to low Q factors. On the contrary are impedances near the unit circle implying high Q factors. This is illustrated in where several constant Q lines are traced.

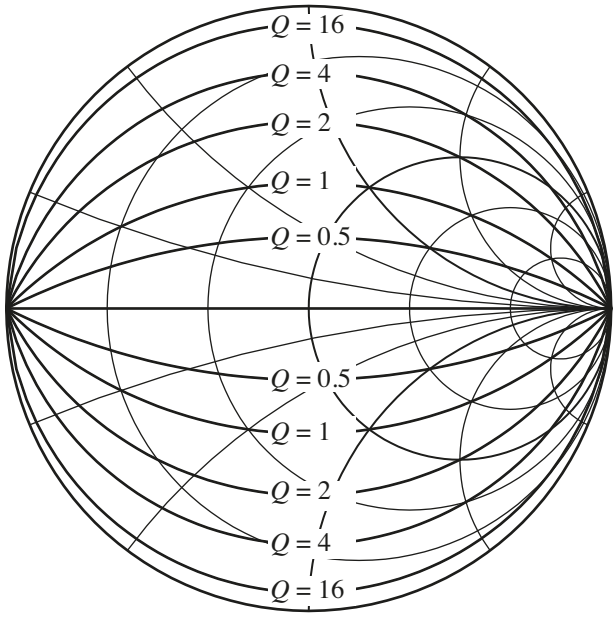


Figure 4.11 Constant Q lines in the Smith Chart.

4.2.4 Power Relations

Beneath the complete Smith Chart, shown in figure 4.5, are also some scales printed for determining some of the power relations that are discussed in section 2.1.6.

The relationship

$$\frac{P^-}{P^+} = |\Gamma|^2 \tag{4.12}$$

is the ratio between the power of the reflected wave and the incident wave and can be read from the scale marked "RFL COEFF P". The corresponding logarithmic expression is given by the scale marked "RTN LOSS [dB]".

The power transmission factor or the mismatch loss is defined as the ratio between the power transmitted to the load and the power in the incident wave. The relation is depicted as

$$T_p = 1 - |\Gamma|^2 \tag{4.13}$$

and is given by the scale marked "TRANSM. COEFF. P".

4.2.5 Standing-Wave Ratio

Among the additional scales below the Smith Chart there are also scales marked "SWR" and "dBS" intended for reading the standing-wave ratio, earlier defined in section 2.1.7:

$$\rho \equiv \text{SWR} \equiv \frac{|V_{max}|}{|V_{min}|} = \frac{|I_{max}|}{|I_{min}|} \tag{4.14}$$

However it is usually easier to determine the SWR directly in the chart by reading the value of r where the constant Γ-circle intersects the horizontal axis at the right side $(r \geq 1)$. This follows from

$$\rho = \frac{1 + |\Gamma|}{1 - |\Gamma|} \tag{4.15}$$

As Γ is real at the horizontal axis and assumed to be positive, then

$$\rho = \frac{1 + \Gamma}{1 - \Gamma} = r \tag{4.16}$$

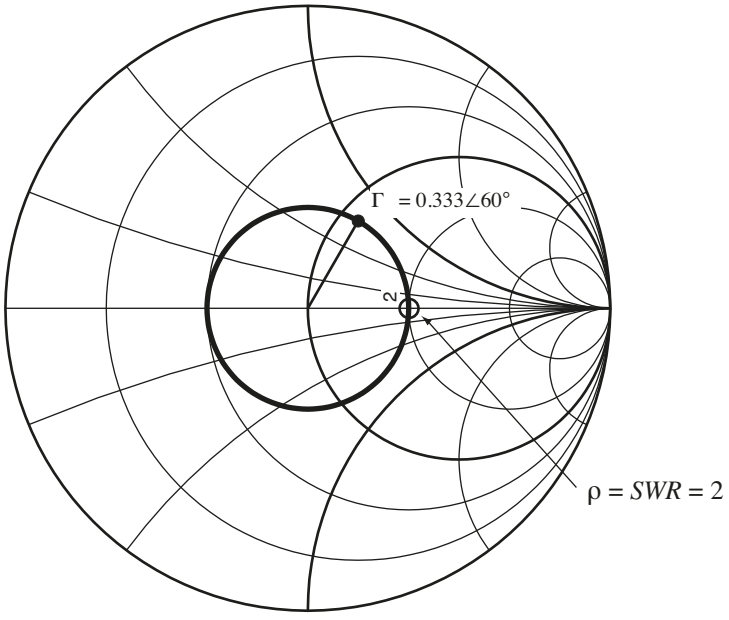


Figure 4.12 Example of reading the SWR in the Smith Chart.

4.3 References

[1] Phillip H. Smith “Electronic Applications of the Smith Chart”, McGraw-Hill, 1969

Chapter 5

Network Design

Impedance matching is one of the most important tasks for the RF (Radio Frequency) designer to deal with. As the power gain as a rule is poor at high frequencies and the circuits often have to handle weak signals, special care has to be taken in the design of RF networks. A true illustration is the front-end of any sensitive receiver and transmitter where different kinds of impedance matching are necessary to meet a hard specification.

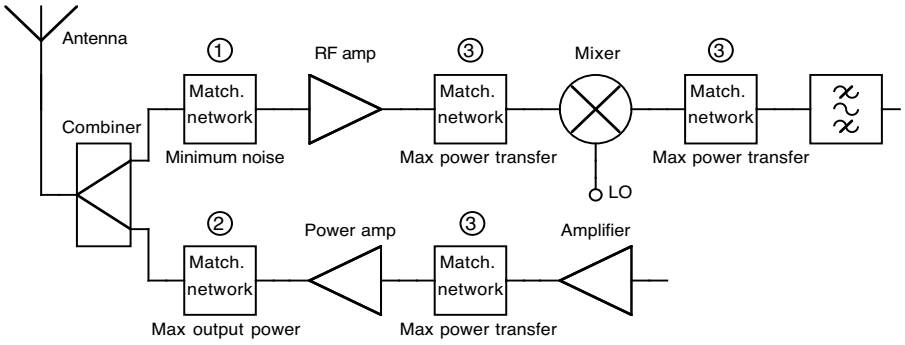


Figure 5.1 Three different kinds of matching can be found in the front-end of a low noise receiver.

Three different approaches can be identified:

1. The input signal (antenna) is most likely connected to the amplifier by a matching network that transforms the antenna impedance to the required source impedance for optimal noise performance. At this point *noise matching* is essential as the signal level can be just slightly exceeding the noise level. The topic is discussed in detail in Chapter 8.
2. In order to deliver a high level of output power and adequate efficiency, the matching network between the power output amplifier and the antenna is designed to provide a specified amount or *maximum output power*. The network transforms the antenna impedance to the optimum load impedance that is required for the amplifier to deliver maximum output power. This topic is discussed in detail in Chapter 14.
3. The remaining interstage matching networks are designed to provide *maximum possible transfer of power*. The method, called complex conjugate matching, will be further explained in this chapter.

Apparently, any needless loss in a circuit, that is already carrying extremely low signal levels, simply cannot be tolerated. Therefore, in most instances, extreme care is taken during the initial design of such a front-end, to make sure that each device in the chain is properly matched to its load.

Different approaches to the design of matching networks will be discussed in more detail in the following sections. No matter if the purpose of the network is to match for minimum noise factor, maximum output power or maximum transfer of power, the same design methods can be applied. The only difference is how the source and load impedances are selected. However, in this chapter we will focus on networks designed for maximum transfer of power (complex conjugate matching).

5.1 Matching

When a load is to be connected to a signal source there are three options that might be referred to as current matching, voltage matching and finally power matching. If a purely resistive source and load are considered the following applies:

- Current matching - the load resistance should be low with respect to the source resistance.
- Voltage matching - the load resistance should be high with respect to the source resistance.
- Power matching - the load should be power matched, i.e. $R_L = R_S$. As will be demonstrated in this chapter, a matching network can be designed such that any load resistance can be transformed to be power matched with the source resistance. This results in a higher current and voltage for the load than what is possible to achieve with current and voltage matching.

Figure 5.2 shows the results of various choices of load resistance. The source voltage is assumed to be 1 V_{rms} and $R_S = 100\Omega$.

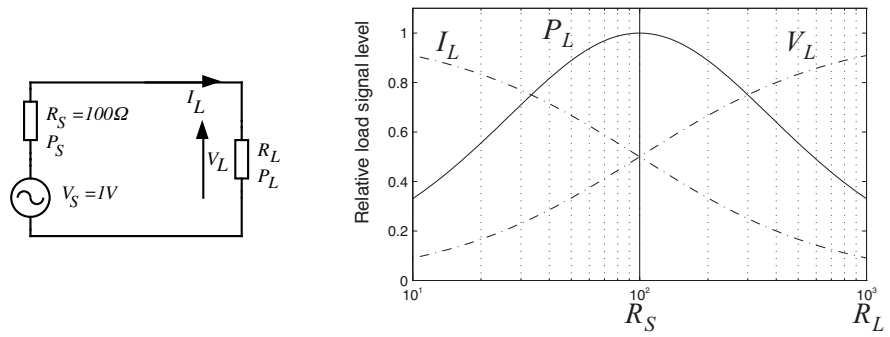


Figure 5.2 The choice of load resistance determines the kind of signal transfer. Maximum power transfer occurs when $R_L = R_S$.

Example 5.1 Signal transfer when R_L is varied.

Explore the circuit in figure 5.2.
The current and the voltage at the load is

$$I_L = \frac{V_S}{R_S + R_L} = \frac{1}{100 + R_L} \tag{5.1}$$

$$V_L = V_S \frac{R_L}{R_S + R_L} = \frac{R_L}{100 + R_L} \tag{5.2}$$

Obviously maximum current is delivered simultaneously as the voltage is zero if the load is a short circuit. On the contrary, the result will be maximum voltage and zero current if the load resistance is infinite.

However, at both these extreme conditions no power will be generated in the load. Then it remains to find out how to choose R_L for maximum power transfer. The power produced at R_L is

$$P_L = V_L I_L = V_S^2 \frac{R_L}{(R_S + R_L)^2} \tag{5.3}$$

where the voltages and the current are given as RMS values. By deriving P_L with respect to R_L it is easily found that maximum power is transferred to the load when $R_L = R_S$.

As seen in example 5.1, maximum power will be transferred to the load if *the load resistance is equal to the source resistance* as stated in the power match theorem:

$$R_L = R_S \tag{5.4}$$

In this case equal amounts of power will be produced in the load as in the source itself. All other ratios between the source and the load resistance will result in a worse transfer of power, i.e. more power will be dissipated in the source than transferred to the load.

The maximum amount of power that can be produced at the load is defined as *available power from source*, P_{AVS}

$$P_{AVS} = V_S^2 \frac{R_L}{(R_S + R_L)^2}\bigg|_{R_L = R_S} = \frac{V_S^2}{4R_S} \tag{5.5}$$

In a high frequency environment the impedances are normally not purely resistive. That is, we have to find an optimum load impedance to obtain the best transfer of power. Real power is purely produced in the real part of the

load impedance and the highest amount is delivered only if the current and the voltage across the load resistance are in phase. As previously discussed in Chapter 1, this condition will be fulfilled when the circuit is in resonance.

$$Z_L = R_L + jX_L = Z_S^* = R_S - jX_S \tag{5.6}$$

This technique is referred to as *complex conjugate matching*.

Figure 5.3 shows that the matching problem is brought back to the power match theorem stated in equation (5.4).

For example, the source, with a series reactive component of $-jX_S$ (a capacitor), is driving its complex conjugate load impedance consisting of a $+jX_L$ reactance (an inductor) in series with R_L. Then the reactive parts will cancel each other leaving only R_S together with R_L and since they are equal, maximum transfer of power will occur.

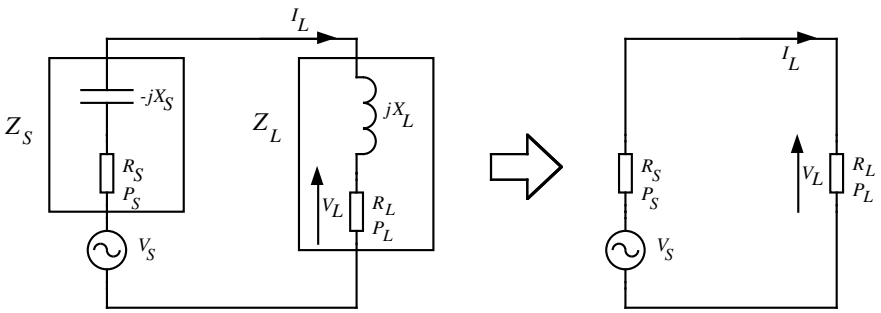


Figure 5.3 Maximum transfer of power occurs when the source impedance is driving its complex conjugate load impedance.

So when we are dealing with complex conjugate matching, we are simply referring to a condition in which any source reactance is resonated with a reactance of equal magnitude and opposite sign, leaving equal resistor values for the source and load terminations.

The elimination of the reactive parts implies that the available power from source can be calculated as previously defined in equation (5.5):

$$P_{AVS} = \frac{V_L^2}{R_L} = V_S^2\left(\frac{R_L}{Z_S + Z_L}\right)^2 \frac{1}{R_L}\bigg|_{Z_L = Z_S^*} = \frac{V_S^2}{4R_S} \tag{5.7}$$

In most cases the source and the load impedances are fixed and that makes it impossible to provide proper matching without additional components. The solution is to insert a matching network between the signal source and the load as shown in figure 5.4. The primary objective is to force the load impedance to "look like" the desired impedance from the source point of view. If

the network matching on one side, matching is automatically obtained on the other side. The design and the qualities of frequently used networks will be illustrated in the next sections.

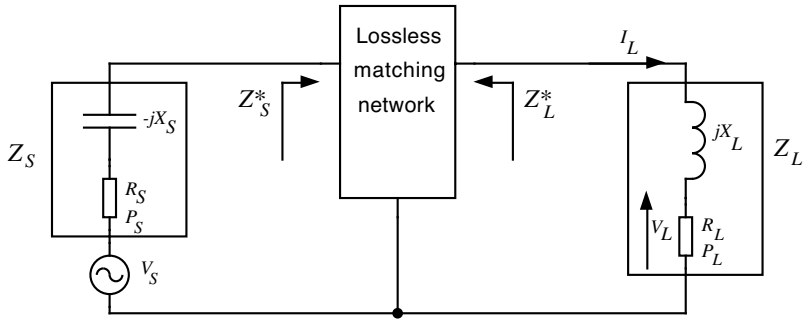


Figure 5.4 A network is needed to carry out proper matching when the load and source impedances are fixed.

5.2 Network Design by Lumped Circuits

There are an infinite number of possible structures which can be used to fulfil the impedance matching function of figure 5.4. However, because we are dealing with reactances, which are frequency dependent, there is perfect match only at one frequency. At all other frequencies diverging from the matching centre frequency, the impedance match becomes gradually worse and eventually nonexistent.

Simple networks designed with only two reactances, such as the L network, leaves no freedom to control the bandwidth. This can be a problem in broadband circuits where a perfect match is wanted everywhere within a broad passband. One solution is to increase the number of reactances in the matching network, in that the designer gains the freedom to choose both the impedances to match and the bandwidth at the expense of a more complex network. This will be discussed in the succeeding sections.

5.2.1 The L Network

The most simple and therefore widely used matching circuit is the L network. If the load and source impedances are resistive the two reactances that forms the network must be of opposite kind, i.e. if the series element is an inductor the shunt element must be a capacitor. Depending on the desired transfer function there are four possible matching circuits as shown in figure 5.5.

The choice of circuit is based on the ratio between R_L and R_S as well as the required filter shape. For example, an amplifier output matching network is often combined with the low-pass filtering circuitry needed to suppress unwanted harmonics. In other situations it is essential to maintain a dc path through the network and then a low-pass configuration will be the typical choice.

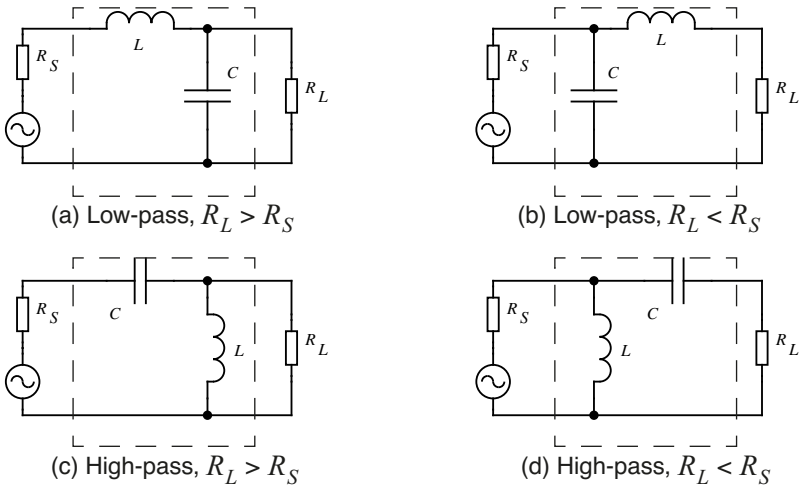


Figure 5.5 The L network has four possible configurations depending on the magnitude of the source and load impedances and the desired filter shape.

However, the primary idea is that the source and load impedances should see their complex conjugate when they look into the network.

Example 5.2 Analysis of a simple L network.

Analyse the matching network in figure 5.6 and verify that the 50Ω source is matched to the 250Ω load.

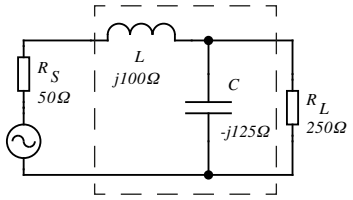


Figure 5.6 A simple impedance-match network.

Without the impedance-matching network, if the load were connected to the source directly, the relative output power would decrease by approximately 2.5 dB according to equations (5.3) and (5.5)

$$\frac{P_L}{P_{AVS}} = \frac{4R_S R_L}{(R_S + R_L)^2} = 0.56 = -2.5 \text{ dB}$$

This loss will be eliminated by the matching network assuming that the reactive components do not introduce any additional losses.

The mechanism of the impedance matching is best understood by transforming the circuit to the serial equivalent as described in section 1.2.3.

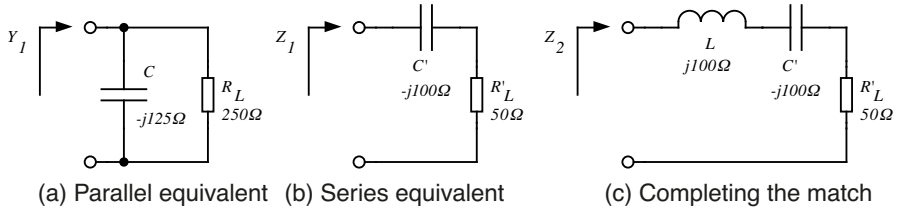


Figure 5.7 Analysis of the network by parallel-to-series conversion.

The parallel combination of R_L and C is transformed into the series equivalent circuit as stated in the equations (1.29) and (1.30)

$$Q = \frac{R_L}{X_C} = \frac{250}{125} = 2$$

$$R'_L = \frac{R_L}{1+Q^2} = \frac{250}{1+4} = 50\Omega$$

$$X_{C'} = X_C \frac{Q^2}{1+Q^2} = -j125\frac{4}{1+4} = -j100\Omega$$

After the parallel-to-series conversion it is clearly seen that the reactive component of -j100Ω will be cancelled by the series inductance of +j100Ω and Z_2 in figure 5.7c will be purely resistive and equal to the source resistance 50Ω.

The conclusions are:

- The function of the shunt component of the impedance-matching network is to transform a larger impedance to a smaller value, with a real part equal to the real part of the other terminating impedance.
- The series impedance-matching element then resonates or cancels any reactive component present, thus leaving the source driving an apparently equal load for optimum transfer of power.

The design of the L network is fairly simple by using the quality factor as illustrated in figure 5.8:

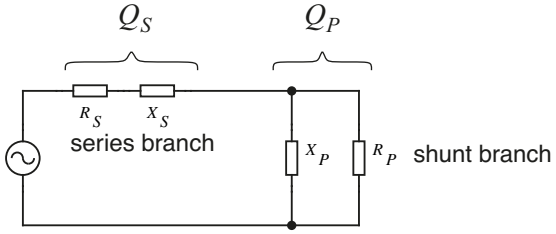


Figure 5.8 Summary of the L-network design.

If power match is obtained it is obvious that the Q factors related to the series and the shunt leg of the circuit has to be equal. By the use of the definitions stated in section 1.2.1 and 1.2.2 the network can be easily designed using the following equations:

$$Q = Q_S = Q_P = \sqrt{\frac{R_P}{R_S} - 1} \tag{5.8}$$

$$|X_S| = QR_S \tag{5.9}$$

$$|X_P| = \frac{R_P}{Q} \tag{5.10}$$

The quantities X_P and X_S may be either capacitive or inductive reactance but, if both the source and load impedances are pure resistances, each must be of the opposite type.

It is very rare that the terminating impedances both are pure resistances in the real world. Transistor input and output impedances are almost always complex valued. Antennas, mixers, filters and most other sources and loads are no different in that respect. That is, they contain both resistive and reactive components. It is therefore essential to handle these reactances in an intelligent manner to obtain adequate impedance matching.

The design procedure is quite similar to the above described method. The idea is to let the source and load reactances be included into the matching network. Almost every set of complex source and load impedances can be matched by the following approach:

1. Choose a network configuration from figure 5.5 to meet the initial conditions and desired filter shape.
2. The matching elements are first calculated as stated in equations (5.8) to (5.10), ignoring the stray reactances.
3. At last, the stray component values are combined with the calculated element values leaving final values for the network elements.

The method is illustrated in the following example:

Example 5.3 Design of a matching network with complex terminations.

Design a network that will match the impedances Z_S and Z_L as shown in figure 5.9. The dc path between the source to the load should be blocked and the operating frequency is $f_0 = 400\text{MHz}$.

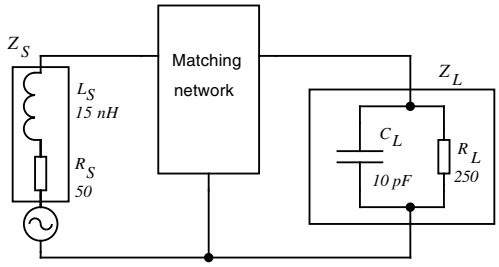


Figure 5.9 Complex source and load impedances that are to be matched.

Since $R_L > R_S$ and dc blocking is required only one network configuration is possible, namely figure 5.5c.

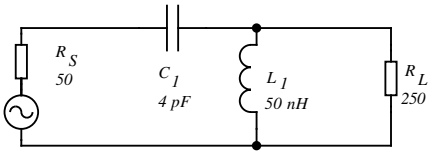


Figure 5.10 The choice of configuration and initial calculations of the network elements.

Disregarding the reactive source and load components, initial values for the two network elements C_1 and L_1 are calculated as defined in equations (5.8) to (5.10):

$$Q = \sqrt{\frac{R_L}{R_S} - 1} = \sqrt{\frac{250}{50} - 1} = 2$$

$$\left|X_{C_1}\right| = QR_S = 2 \cdot 50 = 100\Omega \quad \Rightarrow \quad C_1 = \frac{1}{\omega X_{C_1}} = 4 \text{ pF}$$

$$X_{L_1} = \frac{R_L}{Q} = \frac{250}{2} = 125\Omega \Rightarrow L_1 = X_{L_1}/\omega = 50 \text{ nH}$$

Now we have to take care of the reactive source and load components. From figure 5.11 it is understood that L_S connected in series to C_2 should equal C_1:

$$X_{C_1} = X_{L_S} + X_{C_2} \Rightarrow X_{C_2} = X_{C_1} - \omega L_S = -138\Omega$$

$$C_2 = \frac{1}{\omega X_{C_2}} = 2.9 \text{ pF}$$

The same approach is used at the shunt leg, L_2 connected in parallel to C_L equals L_1:

$$B_{L_1} = \frac{1}{X_{L_1}} = B_{L_2} + B_{C_L} \quad \Rightarrow \quad Y_{L_2} = \frac{1}{X_{L_1}} - \omega C_L = 33 \text{ mS}$$

$$L_2 = \frac{1}{\omega Y_{L_2}} = 12 \text{ nH}$$

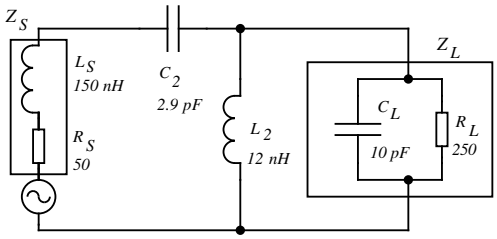


Figure 5.11 The final network design.

Note that in some occasions it will not be possible to compensate the reactive source and load components within the suggested network configuration. There are two different approaches to handle this problem:

1. If there is no disadvantages due to the filter or dc requirements one or both of the network elements can be changed to the opposite type, i.e. a capacitor is replaced with an inductor or vice versa.
2. In other cases additional reactances may be inserted in the network to resonate the source or load reactances at the actual frequency.

The last suggestion, when the network has to be extended with one or more components, leads to the next section where three-element matching will be treated.

5.2.2 The Pi and T Network

As seen in the previous section, the Q of the 2-element network is determined by R_S and R_L. This is a disadvantage as it prevents the designer to choose an arbitrary circuit bandwidth, since the source and load impedances usually are given by other requirements. If the network is extended by one element this disadvantage will be overruled. The 3-element network, shown i figure 5.12, can be designed for just any bandwidth in high-Q applications. The designer has almost full freedom to select any practical circuit Q as long as it is greater that Q which is the result of a 2-element network.

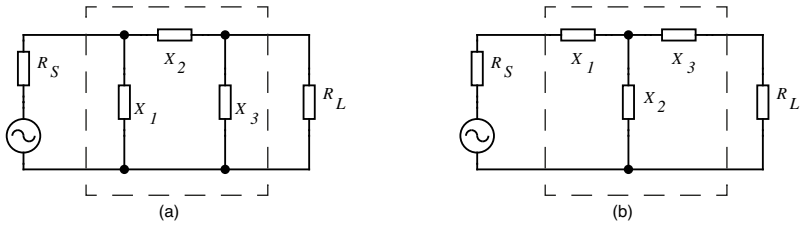


Figure 5.12 The three-element network, (a) Pi configuration and (b) T configuration.

The Pi and T networks can be described as two "back-to-back" connected L networks. Both sides are designed to match the terminating impedances to a virtual resistance, supposed to be an invisible load between the two networks.

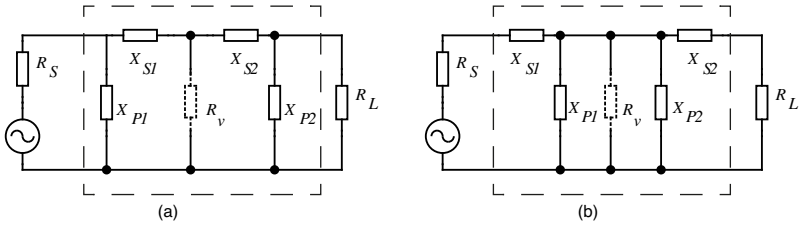


Figure 5.13 Equivalent circuits showing the Pi and T networks as double back-to-back L networks. R_v is a virtual resistance.

The loaded Q of one of the L networks is used to calculate an approximate value of the virtual resistance R_v. The magnitude of Q can be derived from the bandwidth requirements or simply be settled by the designer.

If a Pi network is to be designed, the highest value R_{max} of either R_S or R_L should be used. R_v must be smaller than the terminating resistances as it is connected to the series branch each L section.

$$R_v = \frac{R_{max}}{1+Q^2} \qquad R_{max} = \max(R_S, R_L) \tag{5.11}$$

The same approach is valid for the T network with one exception. As R_v is connected in the shunt arm of the L sections, the smallest of the terminating resistances is used. Thus R_v will always be larger than either R_S or R_L.

$$R_v = R_{min}(Q^2+1) \qquad R_{min} = \min(R_S, R_L) \tag{5.12}$$

When R_v is defined, each L network is calculated in exactly the same manner as was described in the previous section. The design procedure is displayed in the following example.

Example 5.4 Design of a Pi network

Design a Pi network as shown in figure 5.12a to match a 50Ω source to a 250Ω load. The loaded Q_L of the network should be 20. What are the possible configurations of the network?

From equation (5.11) the virtual resistance is calculated as

$$R_v = \frac{R_L}{1 + Q_L^2} = \frac{250}{1 + 20^2} = 0.62\Omega$$

In the design of the L network on the load side the source resistance is replaced by the virtual resistance R_v. The reactances are given from equations (5.9) and (5.10):

$$X_{S2} = Q_L R_v = 20 \cdot 0.62 = 12.5\Omega \tag{5.13}$$

$$X_{P2} = \frac{R_L}{Q_L} = \frac{250}{0.62} = 12.5\Omega \tag{5.14}$$

Before we proceed with the calculations of the reactances on the source side, the Q value for the L network on that side has to be defined. This is done by using equation (5.8) where R_L is replaced by $R_S = 50\Omega$ because the source is connected to the shunt branch. Equally is R_S in the formula substituted by R_v as the load is connected to the series leg of the network.

$$Q_S = \sqrt{\frac{R_S}{R_v} - 1} = \sqrt{\frac{50}{0.62} - 1} = 8.9$$

Now the remaining reactances can be calculated:

$$X_{P1} = \frac{R_S}{Q_S} = \frac{50}{8.9} = 5.6\Omega$$

$$X_{S1} = Q_S R_v = 8.9 \cdot 0.62 = 5.5\Omega$$

The entire design is shown in figure 5.14. Note that the virtual resistance R_v is not really in the circuit, and is shown just for clarity.

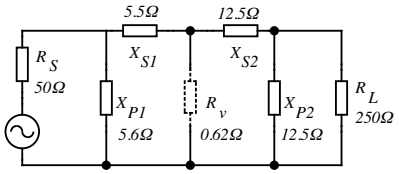


Figure 5.14 The equivalent circuit for the Pi network with calculated reactances.

So far in the design only the reactances are calculated. The last step will be to figure out possible configurations to implement the network.

All the reactances can all be either capacitive or inductive. The only limitation is that X_{P1} and X_{S1} as well as X_{P2} and X_{S2} must be of opposite types. Just as all the four elements are determined, the final circuits are formed by combining the series branches X_{S1} and X_{S2}. Note that if an inductance is selected, the reactance should be added and if a capacitor is selected, the reactance should be subtracted. The four possible configurations are shown in figure 5.15.

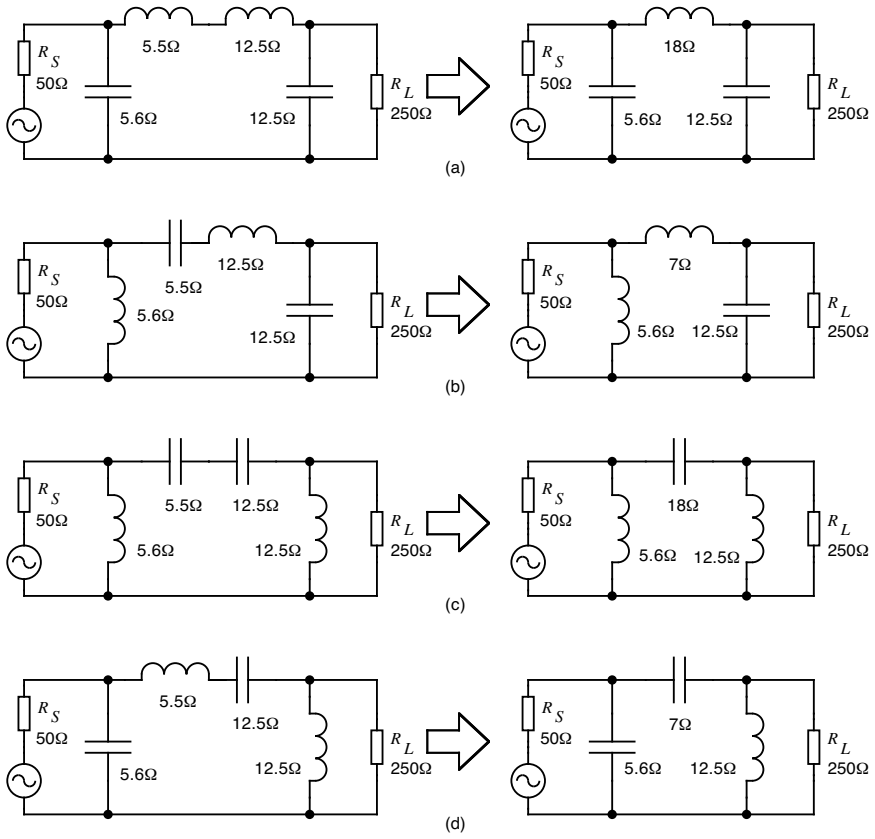


Figure 5.15 The four possible configurations for the Pi matching network.

The example above illustrates how a 3-element matching network can be formed in several shapes. The choice has to be made for each application depending on factors such as the elimination of stray reactances or the need to pass or block dc.

5.2.3 Network Design Using the Smith Chart

The Smith Chart is, as shown in section 4.2.2, an excellent tool to calculate the impedance when series or shunt elements are connected in ladder-type arrangements. An important advantage of the method is that each node of the network can be verified by measurement of the impedance before the calculation is resumed.

A common simplification in the design work is to assume lossless matching. The procedure then is simple:

1. Normalise the required load and source impedances and plot on the chart

$$z_L = \frac{Z_L}{Z_0} \quad \text{and} \quad z_S = \frac{Z_S}{Z_0}$$

2. Choose the desired network configuration.
3. Add shunt and series elements on the chart until the requested impedance is achieved.
   Remember, when a series element is added its reactance is added in z-coordinates, i.e. the locus of the impedance is moving along a constant resistance circle. When a shunt element is added the susceptance is added in y-coordinates, i.e. the locus is moved along a constant conductance circle.

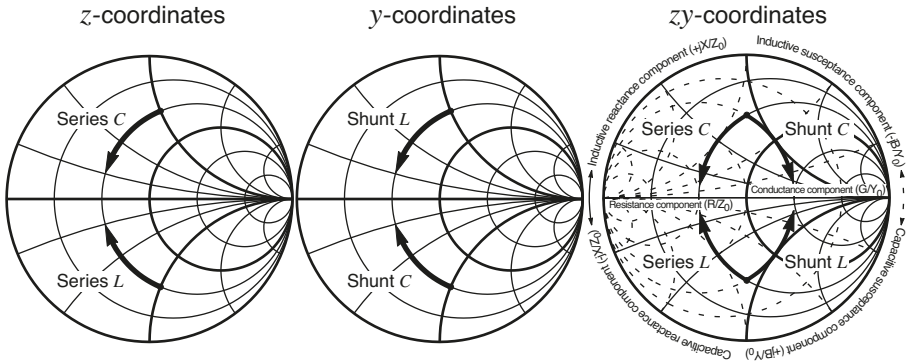


Figure 5.16 Summary of addition of reactive components on a Smith Chart.

4. Read from the chart and denormalise the selected reactances or susceptances and calculate the component values at the actual frequency.

$$\text{For series elements: } C = \frac{1}{\omega x} \cdot \frac{1}{Z_0} \quad \text{or} \quad L = \frac{x}{\omega} \cdot Z_0 \tag{5.15}$$

$$\text{For shunt elements: } C = \frac{b}{\omega} \cdot \frac{1}{Z_0} \quad \text{or} \quad L = \frac{1}{\omega b} \cdot Z_0 \tag{5.16}$$

The procedure is illustrated in the following example.

Example 5.5 Design of a matching network on the Smith Chart.

Design a network, operating at 400 MHz, that will match the load impedance $Z_L = 10 + j10\Omega$ to a 50Ω for optimum transfer of power. Calculate the components and estimate the network bandwidth. Use the configuration shown in figure 5.17 and assume lossless reactances.

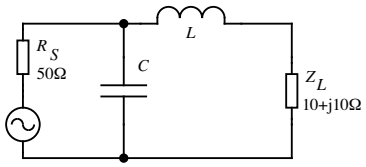


Figure 5.17 A two-element network.

First of all the impedances are normalised to $Z_0 = 50\Omega$ and plotted in the chart

$$z_L = \frac{Z_L}{Z_0} = 0.2 + j0.2 \quad \text{and} \quad z_S = \frac{Z_S}{Z_0} = 1$$

Optimum transfer of power is achieved when the source "sees" the complex conjugate of its impedance, i.e. $z_S^* = z_S = 1$, when it is looking into the network. Due to reciprocity the same condition is true as $z_L^* = 0.2 - j0.2$ should be seen when looking into the network from the load's point of view. It is therefore obvious that there are two different paths to follow in the chart. However, any of them produces equal solutions. For clarity, both paths are illustrated in figure 5.18.

Let's start at the load.
As the inductor is connected in series, its reactance is added to z_L.
In order to determine the amount of reactance, an additional constant conductance circle showing $g = 1$ is drawn.
Then follow the constant resistance circle from z_L until it intersects the help circle. The normalised value of the inductive reactance, x_i, is equal to the difference between the two points read from the scale at the circumference.

$$jx_i = z_1 - z_L = (0.2 + j0.4) - (0.2 + j0.2) = j0.2$$

The next step is to calculate the value of the capacitive reactance, x_c, that is connected in parallel. As calculation of shunt connections are best done by addition of admittances, the value corresponding to z_1 is plotted:

$$y_1 = \frac{1}{z_1} = 1 - j2$$

As capacitive susceptance is added to y_1, the locus is shifted at the constant conductance circle towards $y_S^* = 1/z_S = 1$. The normalised value of the susceptance, b_c, is read from the scale as the difference between y_1 and y_S^*:

$$jb_c = y_S^* - y_1 = 1 - (1 - j2) = +j2$$

This completes the calculation on the chart and the result is shown in figure 5.18a.

If the start point is chosen to the source, the procedure will be almost the same except that the path will end in y_L^* as illustrated in figure 5.18b. The values of x_i and b_c are exactly the same as the values derived by the former path.

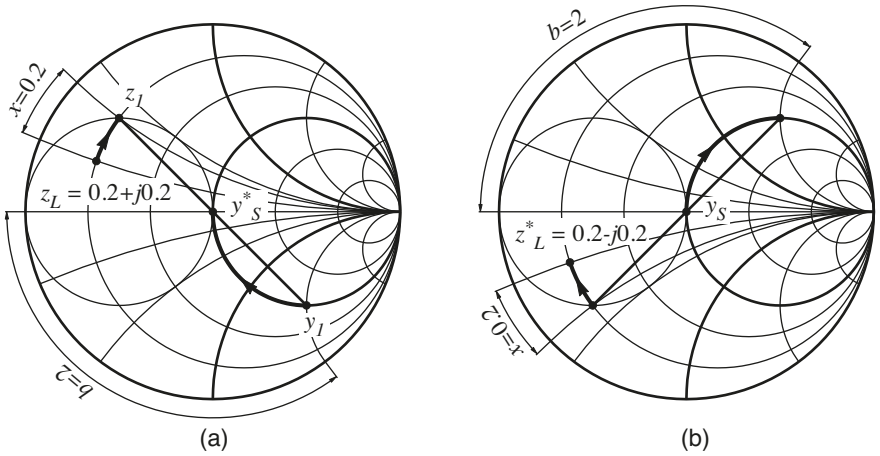


Figure 5.18 Calculation of the reactances on the Smith Chart. a) shows the path from the load to the source and b) shows the reverse direction.

An approximate value of the network bandwidth is derived from the node that corresponds to maximum Q. Referring to figure 4.11, that point seems to be $z_1 = r_1 + jx_1$ that gives

$$B = \frac{f_0}{Q} = \frac{f_0 r_1}{x_1} = \frac{400 \cdot 10^6 \cdot 0.2}{0.4} = 200\text{MHz}$$

Finally, the component values are calculated at $f_0 = 400\text{MHz}$ by using equations (5.15) and (5.16):

$$L = \frac{x_i}{\omega} \cdot Z_0 = \frac{0.2}{2\pi \cdot 400 \cdot 10^6} \cdot 50 = 4\text{nH}$$

$$C = \frac{b}{\omega} \cdot \frac{1}{Z_0} = \frac{2}{2\pi \cdot 400 \cdot 10^6} \cdot \frac{1}{50} = 16\text{pF}$$

In the initial phase of the design, the designer has to decide what kind of network configuration to use. The selection is normally based on the load and source impedance ratio and other practical aspects, such as filter characteristics or the need of a dc path.

The qualities of four kinds of L networks was discussed in section 5.2.1 and figure 5.5. The circuits are reprinted in figures 5.19a-d and completed with outlines of the Smith Chart that defines the matching capabilities of the possible L-type combinations. The remaining figures 5.19e-h shows L networks based on reactances of equal type, known as tapped-capacitor circuit or tapped-inductor circuit in Chapter 1.

Figure 5.19 serve as a guide to the selection of a suitable L-type matching circuit for any particular impedance transformation. If the locus of the load impedance, z_L , falls anywhere inside of the unshaded area, it is transformable to a pure resistance z_0 .

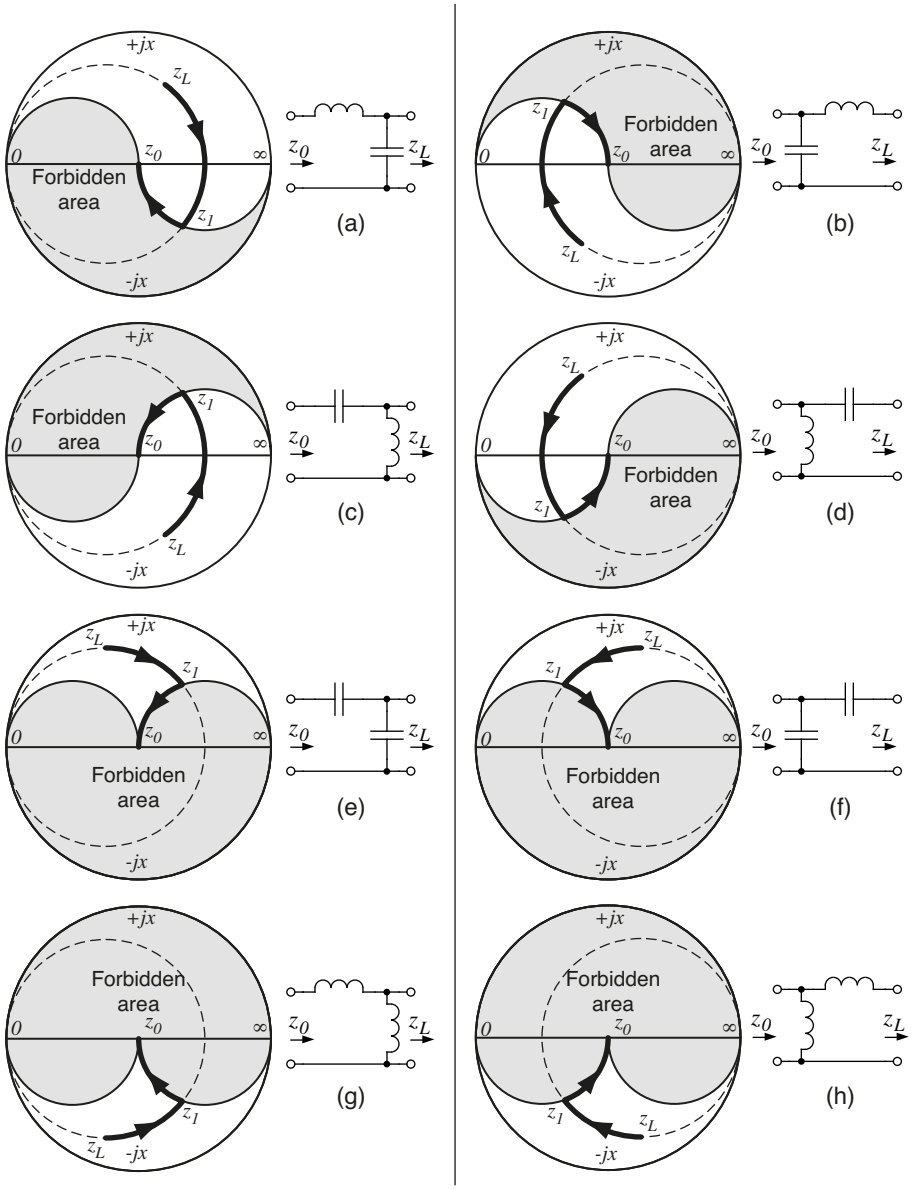


Figure 5.19 Eight possible L-type networks for transforming a complex load impedance z_L to a pure resistance z_0.

It is clear that the L-match is very common due to its simplicity. However, there are only two degrees of freedom as only the values of L and C are possible to select. Hence, once the impedance transformation ratio and the resonant frequency have been specified, the network Q is automatically determined. If a different value of Q is required, the network has to be expanded at least by one more element that provides additional degrees of freedom. This is illustrated by analysing example 5.4 in the Smith Chart:

Example 5.6 Analysis of a Pi network on the Smith Chart

Analyse the network in figure 5.20 that was designed in example 5.4 to match a 50Ω source to 250Ω load.

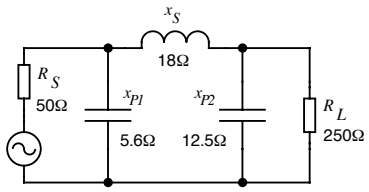


Figure 5.20 The Pi network with reactances calculated in example 5.4

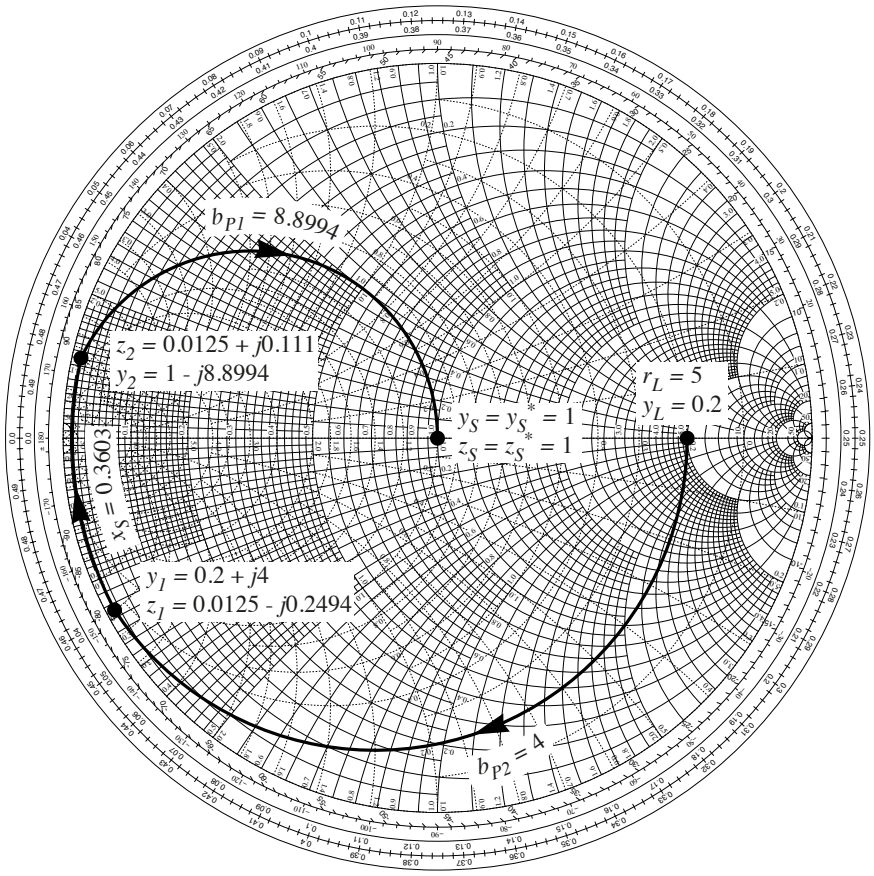


Figure 5.21 The impedance path from load to source plotted on the Smith Chart. All values are normalised with $Z_0 = 50\Omega$.

The loaded Q_L of the network can be calculated from z_1:

$$Q_L = \frac{x_1}{r_1} = \frac{g_1}{b_1} = \frac{4}{0.2} = 20$$

It can be found out from example 5.6 that the shape of the impedance path gives a sense of the network Q. Figure 4.11, showing constant Q lines in the Smith Chart, is a valuable help on how to choose a proper path for a required Q. This is illustrated in example 5.22 where two different paths are outlined, both matching z_L to z_S. The conclusion is that the Q will decrease when the path is hold close to the resistance axis.

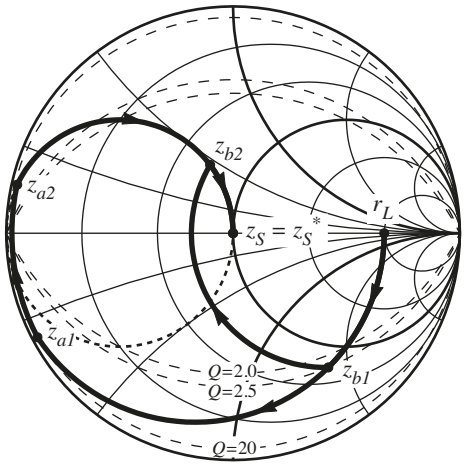


Figure 5.22 An alternative solution for the problem in example 5.6.

There is however a lower limit of Q that can be implemented by a three-element network. In the actual example the minimum possible Q is 2, and then the circuit is reduced to a two-element L network.

Hence, under some conditions it may be necessary to increase the number of network elements to fulfil the specification. Figure 5.23 illustrates the concept where two networks of higher order are performing equal impedance transformation, but are showing different values of Q. Under these circumstances it is not possible to determine the Q value numerically on the chart. However, it is a good indicator to find out the order of Q. Compare with the systematic methods for filter design in Chapter 12.

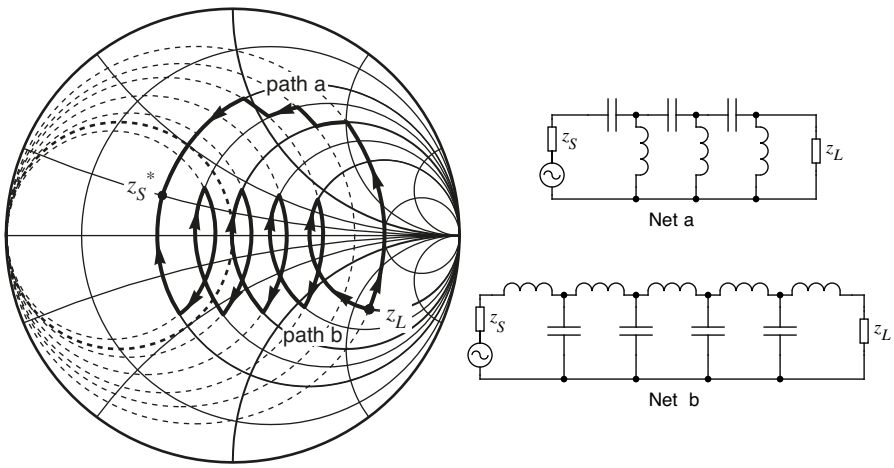


Figure 5.23 Multielement matching. Path a shows a high Q and path b shows a low Q network.

5.3 Network Design by Line Structures

When the operating frequency advances into the micro-wave region the practical problems to handle lumped reactive components will be evident. Parasitic reactances as well as rising losses are no longer negligible. Unfortunately this will result in complex models and tedious calculations to meet, if it is ever possible, any stipulated specification.

Luckily there are alternative techniques that are independent of lumped elements to achieve impedance transformation. As seen in section 2.1.5 the input impedance of a transmission line with an electrical length of ℓ and the opposite end is connected to a load impedance is given by

$$Z_{in} = Z_0 \frac{1 + \Gamma_{in}}{1 - \Gamma_{in}} \text{ where} \tag{5.17}$$

$$\Gamma_{in} = \Gamma_L e^{-2\gamma\ell} = \Gamma_L e^{-2\alpha\ell} e^{-j2\beta\ell} \tag{5.18}$$

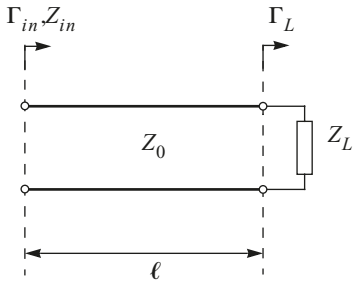


Figure 5.24 Transformation of impedance by a transmission line.

Substituting equation (5.18) into (5.17) yields the final expression for Z_{in}:

$$\begin{aligned} Z_{in} &= Z_0 \left(\frac{(Z_L + Z_0)e^{\gamma\ell} + (Z_L - Z_0)e^{-\gamma\ell}}{(Z_L + Z_0)e^{\gamma\ell} - (Z_L - Z_0)e^{-\gamma\ell}} \right) \\ &= Z_0 \left(\frac{Z_L + Z_0 \tanh\gamma\ell}{Z_0 + Z_L \tanh\gamma\ell} \right) \end{aligned} \tag{5.19}$$

For a lossless line when $\gamma = j\beta$, and by using
$\sinh j\beta\ell = j\sin\beta\ell$ and $\cosh j\beta\ell = \cos\beta\ell$, equation (5.19) is reduced to

$$Z_{in} = Z_0 \left(\frac{Z_L \cos\beta\ell + jZ_0 \sin\beta\ell}{Z_0 \cos\beta\ell + jZ_L \sin\beta\ell} \right) = Z_0 \left(\frac{Z_L + jZ_0 \tan\beta\ell}{Z_0 + jZ_L \tan\beta\ell} \right) \tag{5.20}$$

A look into the Smith Chart exposes a new kind of path where the impedance locus is navigating along a Γ arc. If the transmission line can be considered as lossless, which is the normal approximation, the arc turns into a constant Γ circle, as shown in figure 5.25.

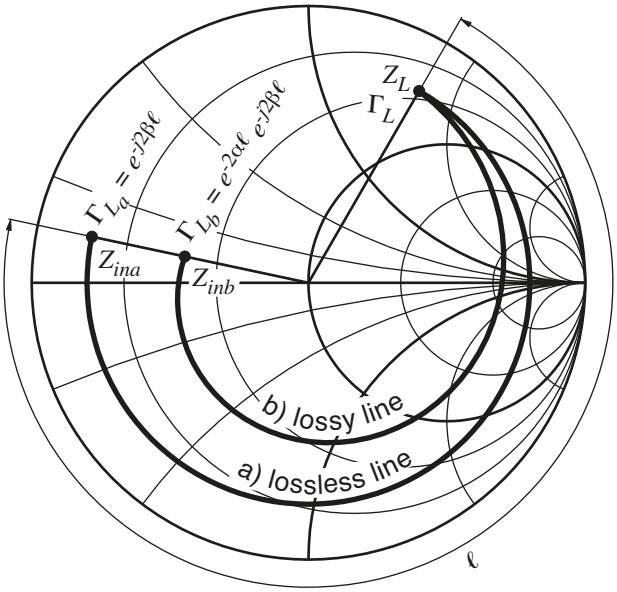


Figure 5.25 transmission line as an impedance transformer.

Almost any impedance matching problem can be solved theoretically by a single line element by selecting proper electrical length and characteristic impedance. However, in many cases the solution for the single element line will assume unreasonable dimensions besides the lack of freedom to control the network bandwidth. These problems can be mastered by connecting several line elements with changed properties in cascade or by connecting reactive shunt elements, so called stubs, to the network. These methods will successively be discussed in the following sections.

Since the electrical length of a transmission line is inversely proportional to the frequency, these kind of matching networks are only utilised at frequencies in the GHz range. At lower frequencies the dimensions of the network tends to be unrealistically bulky. Nevertheless, line elements can reliably be combined with lumped circuit elements in order to perform impedance matching at intermediate frequencies.

5.3.1 Quarter-Wave Transformer

An important application is a line with fixed electrical length of $\lambda/4$, often named as "*quarter-wave transformer*". Assuming a lossless line the required Z_0 is derived from equation (5.20):

$$\ell = \lambda/4 \quad \Rightarrow \quad \beta\ell = \pi/2$$

$$Z_{in} = \frac{Z_0^2}{Z_L} \quad \Rightarrow \quad Z_0 = \sqrt{Z_{in} Z_L} \tag{5.21}$$

The procedure is illustrated in the following example where the problem in example 5.2 is solved by a transmission line.

Example 5.7 Design of quarter-wave transformer.

Calculate the characteristic impedance of a $\lambda/4$ line to properly match a 50Ω source to a 250Ω load.

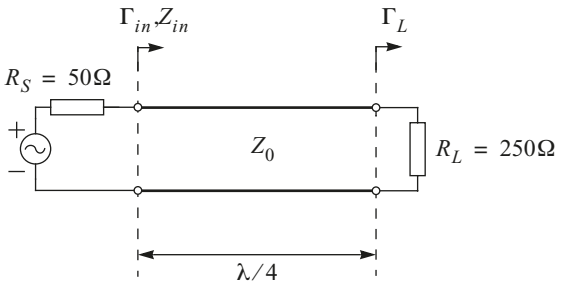


Figure 5.26 Quarter-wave transformer.

$$Z_0 = \sqrt{R_s R_L} = \sqrt{50 \cdot 250} = 111.8\Omega$$

The result is depicted in the Smith Chart as a half turn via the constant Γ circle, that is intersecting the normalised load and source resistances.

$$r_S = \frac{R_S}{Z_0} = \frac{50}{111.8} = 0.45\,, \quad r_L = \frac{R_L}{Z_0} = \frac{250}{111.8} = 2.24$$

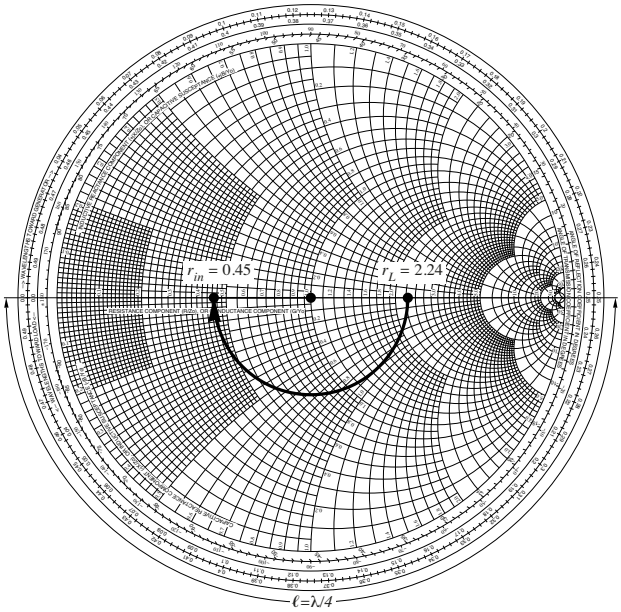


Figure 5.27 The result of the quarter-wave transformer plotted in the Smith Chart

When the ratio between the load and source impedances is large, the bandwidth associated with the quarter-wave transformer will be fairly narrow. Unfortunately, as there are only two degrees of freedom, the characteristic impedance and the length, there is no possibility to control the bandwidth unless the network is extended with one or more elements. Note the similarity to the discussion concerning the L network in the previous section.

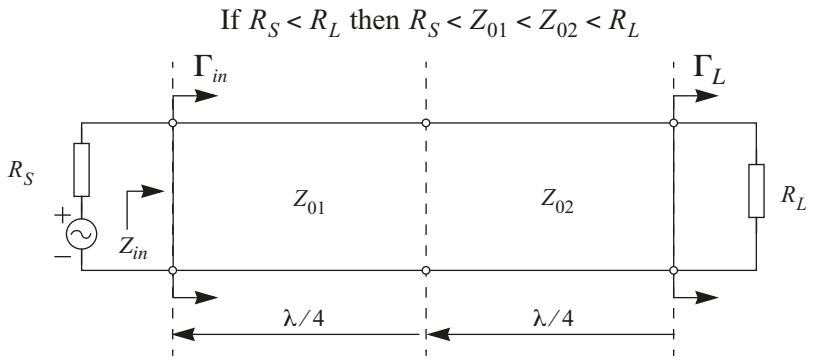


Figure 5.28 Impedance transformation by cascaded quarter-wave lines.

Thus, a low standing-wave can be obtained over a wider frequency band if the impedance is changed gradually by two or more quarter-wave lines. Maximum-flatness is achieved when the subsequent impedance shifts are exhibiting geometrical relationships determined by the binomial coefficients. Each change in impedance is then determined by

$$\frac{Z_{0(k)}}{Z_{0(k-1)}} = \left(\frac{R_L}{R_S}\right)^{\binom{n}{k-1}/2^n} \tag{5.22}$$

where n is the total number of line sections and k is the position of each section starting with $k = 1$ at the source, i.e. $Z_{0(0)} = R_S$ and $Z_{0(n+1)} = R_L$. This procedure can, of course, also be applied to lumped ladder-type networks.

Table 5.1 Binomial Coefficients $\binom{n}{k}$

$n \downarrow$	$k \rightarrow$ 0	1	2	3	4	5
0	1					
1	1	1				
2	1	2	1			
3	1	3	3	1		
4	1	4	6	4	1	
5	1	5	10	10	5	1

For example, if three quarter-wave sections are used to match R_S to R_L, the four steps are determined from the number series 1-3-3-1. Thus the characteristic impedance for each section is given by

$$Z_{01} = R_S\left(\frac{R_L}{R_S}\right)^{\frac{1}{8}}$$

$$Z_{02} = Z_{01}\left(\frac{R_L}{R_S}\right)^{\frac{3}{8}} = R_S\left(\frac{R_L}{R_S}\right)^{\frac{4}{8}}$$

$$Z_{03} = Z_{02}\left(\frac{R_L}{R_S}\right)^{\frac{3}{8}} = R_S\left(\frac{R_L}{R_S}\right)^{\frac{7}{8}}$$

Figure 5.29 shows the broadening of bandwidth when the matching network in example 5.7 is designed with one, two or three quarter-wave sections for maximum-flatness. The frequency scale is normalised to the mid-frequency f_0 corresponding to the electrical wavelength $\lambda/4$. It can also be observed that perfect matching, SWR = 1:1, in all cases is obtained at f_0.

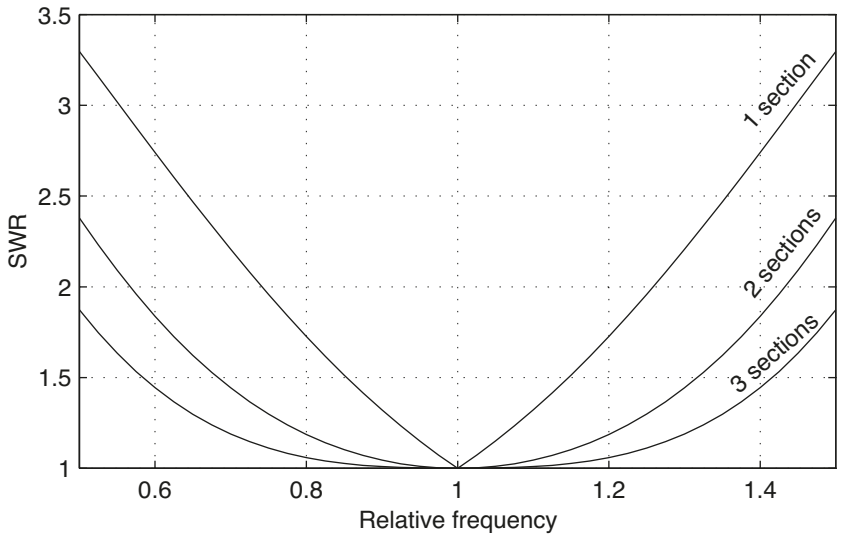


Figure 5.29 Comparison of the SWR at the source when one, two or three quarter-wave sections are used to match $R_S = 50\Omega$ to $R_L = 250\Omega$.

5.3.2 Line Section with Optimised Length and Z_0

Obviously, quarter-wave transformers are only capable of matching purely resistive impedances as each line section always moves the impedance locus one half turn in the Smith Chart. This was illustrated in figure 5.27.

Nevertheless, matching of a complex-valued load impedance to a complex-valued source can, with a few exceptions, be obtained through optimising both Z_0 and the length of the line. If the source impedance is equal to $Z_S = R_S + jX_S$, complex conjugate matching is obtained when the load impedance $Z_L = R_L + jX_L$ is transformed into $Z_{in} = Z^*_s = R_S - jX_S$.

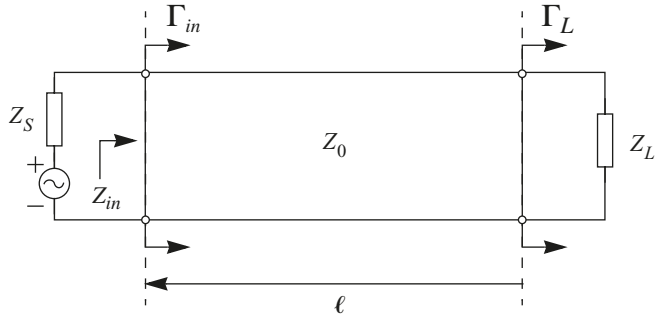


Figure 5.30 Impedance transformation by a single line with optimised characteristic impedance and length.

The optimum characteristic impedance of the line, assuming a lossless line, conforms to

$$|\Gamma_L| = |\Gamma_{in}| \tag{5.23}$$

and can be obtained by solving equation (5.23) with respect to Z_0:

$$Z_0 = \sqrt{R_L R_{in}}\sqrt{1 + \frac{X_L^2}{R_L(R_L - R_{in})} - \frac{X_{in}^2}{R_{in}(R_L - R_{in})}} \tag{5.24}$$

As a positive and real Z_0 is required, it is obvious that not all combinations of impedances are possible to match with one single line. Some limitations when Z_S is pure resistive are depicted in figure 5.31.

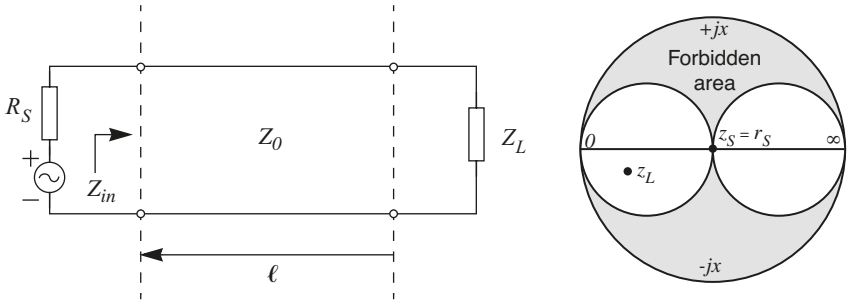


Figure 5.31 If the locus of the load impedance, z_L, falls anywhere inside of the unshaded area, it is transformable to a pure resistance $z_S = r_S$ by a single line.

When Z_L and Z_S are normalised by the value of Z_0 from equation (5.24) and plotted on the Smith Chart, they will both be found on the same constant $|\Gamma|$ circle. The electrical line length is then finally read on chart or derived from

$$\Gamma_{in} = \Gamma_L e^{-j2\beta d} \quad \Rightarrow \quad \frac{d}{\lambda} = \frac{1}{4\pi}|\theta_L - \theta_{in}| \tag{5.25}$$

The procedure is illustrated in the following example.

Example 5.8 Calculation of optimum Z_0 and line length.

Design a single transmission line as shown in figure 5.30 to perform matching between
$Z_S = 65 - j50\Omega \Rightarrow Z_{in} = Z_S^* = 65 + j50\Omega$, and
$Z_L = 10 - j10\Omega$.

Equation (5.24) is used to calculate the characteristic impedance:

$$Z_0 = \sqrt{65 \cdot 10}\sqrt{1 + \frac{(-10)^2}{10(10-65)} - \frac{50^2}{65(10-65)}} = 30\Omega$$

The source and load impedances are normalised by Z_0 and then plotted on the chart:

$$z_{in} = \frac{Z_{in}}{Z_0} = \frac{65 + j50}{30} = 2.24 + j1.73$$

$$z_L = \frac{Z_L}{Z_0} = \frac{10 - j15}{30} = 0.35 - j0.52$$

It is clear that r_S and z_L both are located on the same constant $|\Gamma|$ circle.

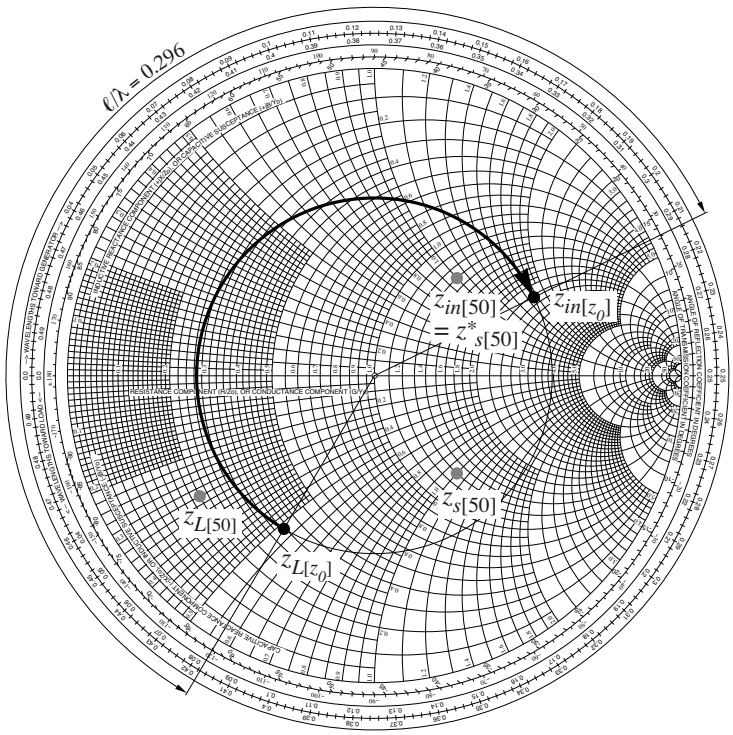


Figure 5.32 Both the characteristic impedance and the electrical length are optimised to obtain matching between a complex load and a complex source impedance by a single line segment.

The electrical length of the line, expressed in wavelengths, is at last read from the outermost circular scale:

$$\frac{\ell}{\lambda} = 0.214 + 0.082 = 0.296$$

It is of course possible to design cascaded line sections, each with optimised Z_0 and length. However, the necessary calculations for this kind of application is to extensive to be carried out by hand.

At higher frequencies, corresponding to centimetre or millimetre waves, the impedance transformation often is achieved by line structures where the characteristic impedance continuously is altered from the load to the source impedance, so called tapered lines. However, the theory needed to calculate the shape of such structures is fairly comprehensive and is therefore not covered by this text.

5.3.3 Stubs, Short-Circuited and Open Line Sections

As seen in the previous section, there are sets of impedances that can not be matched entirely by one line section. A more reliable method is to combine a transmission line with a shunt connected susceptance as shown in figure 5.33. As a matter of principle, matching can be obtained between every pair of impedances by this approach:

1. The load impedance is transformed by a line to an impedance where the real part of the corresponding admittance is equal to the real part of the desired input admittance.
2. The remaining imaginary part is adjusted by a shunt susceptance to attain the input impedance.

The shunt element may be a lumped component, but as stated earlier the performance is limited with increasing frequency. A more common application is the use of open or short-circuited line elements, so called *stubs*.

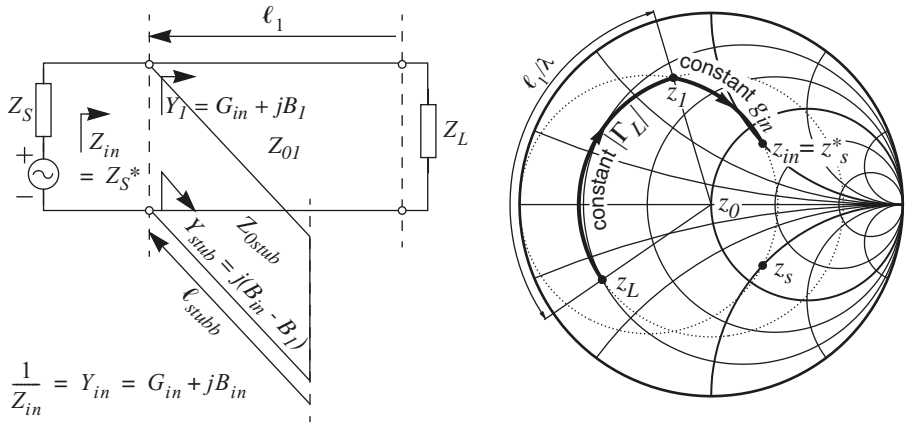


Figure 5.33 An example of a matching network that uses one line segment and a short-circuited stub. The corresponding impedance path is shown in the Smith Chart.

Depending on the ratio between R_L and R_{in}, the stub often needs to be connected to the proper end of the line. In the example shown in figure 5.33 $R_L < R_{in}$ and therefore the stub is connected to the source side of the line. Compare with the discussion concerning the L network in section 5.2.1.

In contrast to the design of lumped networks, where the calculations are done in sequence on a single Smith Chart, it is preferable to handle the design of the stub separately as soon as the required shunt susceptance is determined.

The stub can consist either of an open or a short-circuited line. The choice depends on practical matters such as the need for a dc path and obtaining a convenient electrical length. As a rule the short-circuited stub is preferable, as the end effects are easier to handle.

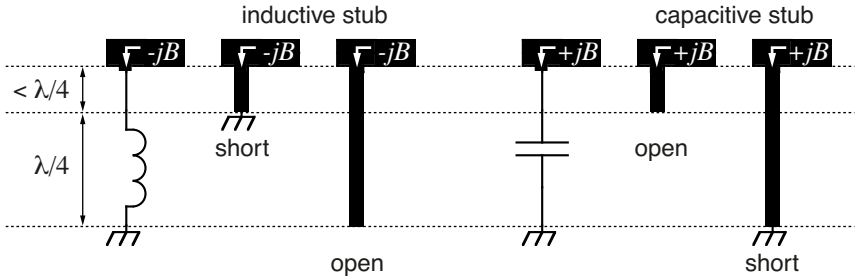


Figure 5.34 The susceptance is controlled by the length of the microstrip and the termination.

The input impedance of the stub is pure reactive and determined from equation (5.20) for a short-circuited stub

$$Z_{in} = Z_0\left(\frac{Z_L + jZ_0\tan\beta\ell}{Z_0 + jZ_L\tan\beta\ell}\right)\Bigg|_{Z_L = 0} = jZ_0\tan\beta\ell = jX_{in} \tag{5.26}$$

or for an open stub

$$Z_{in} = Z_0\left(\frac{Z_L + jZ_0\tan\beta\ell}{Z_0 + jZ_L\tan\beta\ell}\right)\Bigg|_{Z_L = \infty} = -jZ_0\tan\beta\ell = jX_{in} \tag{5.27}$$

The characteristic impedance of the stub can be chosen arbitrary. However, higher values of Z_0 reduces the consequences of stray effects at the end of the stub. By substituting $\beta = 2\pi/\lambda$ the electrical length ℓ/λ of the short-circuited stub is calculated as

$$\begin{cases} \dfrac{\ell}{\lambda} = \dfrac{1}{2\pi}\operatorname{atan}\left(\dfrac{X_{in}}{Z_0}\right)\Bigg|_{X>0} = \dfrac{-1}{2\pi}\operatorname{acot}(B_{in}Z_0)\big|_{B<0} & \text{(inductive)} \\ \dfrac{\ell}{\lambda} = \dfrac{1}{2} + \dfrac{1}{2\pi}\operatorname{atan}\left(\dfrac{X_{in}}{Z_0}\right)\Bigg|_{X<0} = \dfrac{1}{2} - \dfrac{1}{2\pi}\operatorname{acot}(B_{in}Z_0)\big|_{B>0} & \text{(capacitive)} \end{cases} \tag{5.28}$$

and the length of the open stub is derived in the same manner:

$$\begin{cases} \dfrac{\ell}{\lambda} = \dfrac{-1}{2\pi}\operatorname{acot}\left(\dfrac{X_{in}}{Z_0}\right)\Big|_{X<0} = \dfrac{1}{2\pi}\operatorname{atan}(B_{in}Z_0)\big|_{B>0} & \text{(capacitive)} \\ \dfrac{\ell}{\lambda} = \dfrac{1}{2} - \dfrac{1}{2\pi}\operatorname{acot}\left(\dfrac{X_{in}}{Z_0}\right)\Big|_{X>0} = \dfrac{1}{2} + \dfrac{1}{2\pi}\operatorname{atan}(B_{in}Z_0)\big|_{B<0} & \text{(inductive)} \end{cases} \tag{5.29}$$

It is very convenient to calculate the stubs on the Smith Chart as shown in figure 5.35. Starting at $y_{short} = \infty$ in the case of short-circuited line, or $y_{open} = 0$ by an open line, the admittance locus is moved along the perimeter of the chart as the stub length is increased. The line is assumed to be lossless. The electrical length is found on the outermost circular scale. If the calculated stub length is inconveniently short, it can be extended by one or more $\lambda/2$ as this corresponds to complete turns in the chart.

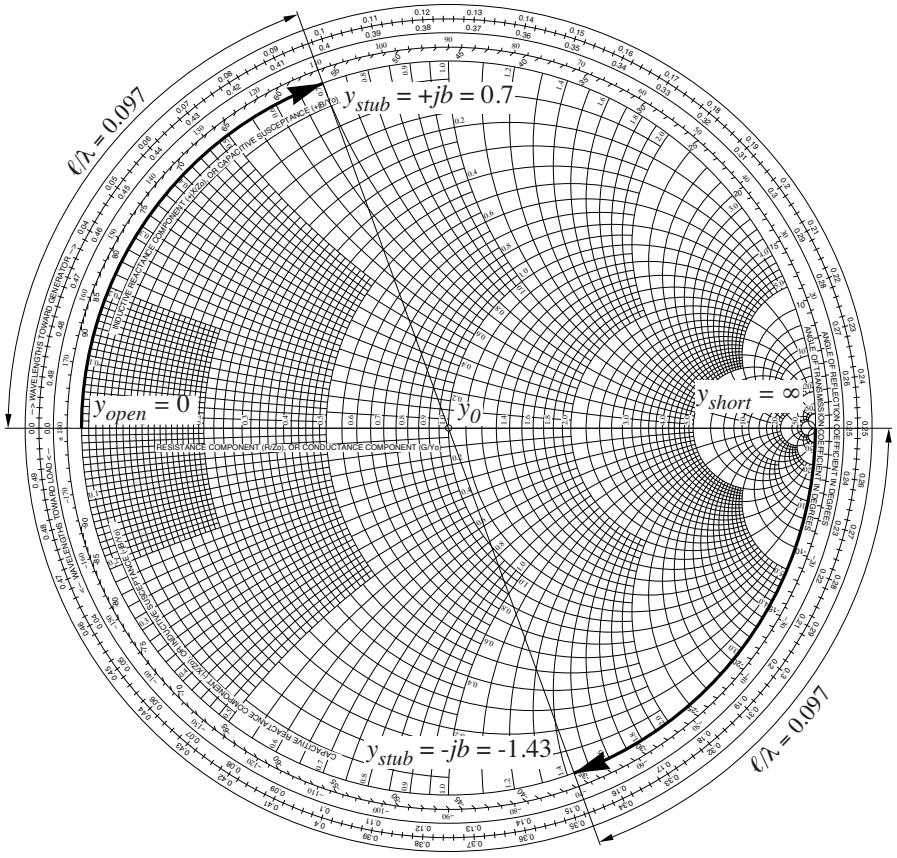


Figure 5.35 Determining the electrical length of a stub on the Smith Chart.

The shunt stub can be split into two parallel stubs in order to give a more symmetrical structure as shown in figure 5.36. By this action each stub must produce half the value of the requested susceptance as they are connected in parallel. Depending on the choice of open or short-circuited stubs and the requested susceptance, the length of the symmetrical stubs may be either shorter or longer than the single stub.

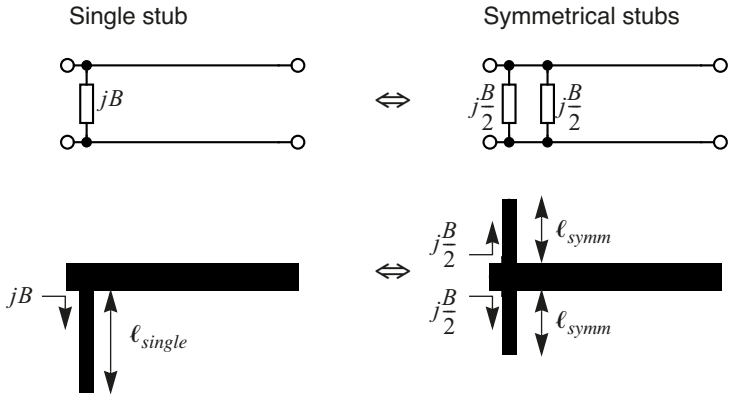


Figure 5.36 A single stub can be split into two parallel stubs.

Example 5.9 Design of a matching network using one stub.

Design on the Smith Chart a network that consists of a line and a short-circuited stub, as shown in figure 5.33, to obtain complex conjugate match between the load impedance $Z_L = 10 - j15\Omega$ and the source impedance $Z_S = 65 - j50\Omega$. The characteristic impedance for the line and the stub is $Z_0 = 50\Omega$.

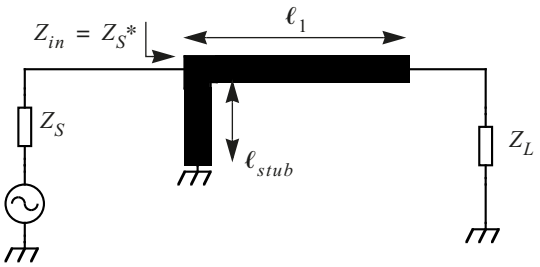


Figure 5.37 Single stub matching network.

The procedure will be to find a suitable electrical length ℓ_1 of the line that transforms z_L to a value z_1 where the real part of the corresponding admittance $y_1 = 1/z_1$ is equal to the input conductance. Then the stub is shunt connected to the end of the line and y_1 is adjusted by addition of the susceptance b_{stub} until the imaginary part of y_1 is equal to the input susceptance. Finally the stub length ℓ_{stub} is determined to transform the short-circuited end 0Ω to the required value b_{stub}.

Start by calculating the normalised input and load impedances.

$$z_{in} = z_S^* = \frac{65 + j50}{50} = 1.3 + j$$

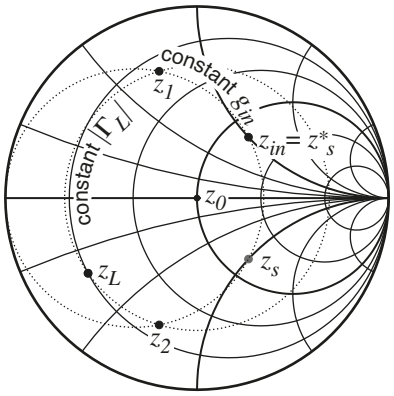
$$z_L = \frac{10 - j15}{50} = 0.2 - j0.3$$

When the impedances are plotted in the Smith Chart two help circles are drawn:

1. A constant $|\Gamma|$ circle through z_L

2. A constant g circle through z_{in}.

There will be two intersections, z_1 and z_2. between the circles. Each of them is a possible solution and z_1 that gives the shortest transmission line is selected.



The electrical length ℓ_1 of the line is read on the outermost circular scale:

$$\ell_1 = 0.15\lambda$$

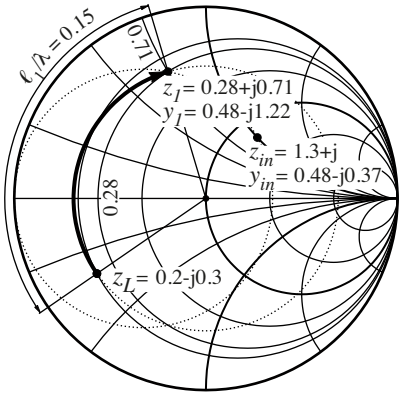
and the impedance z_1 is found to be

$$z_1 = 0.28 + j0.71$$

When the admittances corresponding to z_1 and z_{in} are calculated or read on the chart it is seen that their real part are equal

$$y_1 = g_1 + jb_1 = 0.48 - j1.22$$

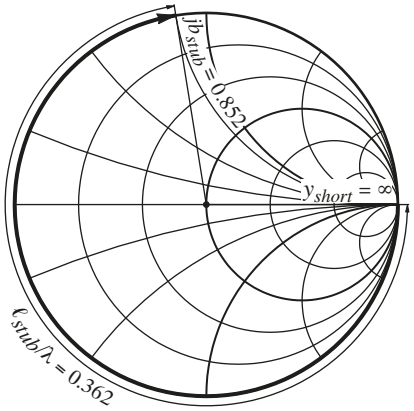
$$y_{in} = g_{in} - jb_{in} = 0.48 - j0.37$$



The difference in the imaginary parts has now to be compensated by addition of a susceptance from the shunt connected stub. The proper value of the sub susceptance is

$$b_{stub} = b_{in} - b_1 = 0.85$$

Then the electrical length is calculated on a separate chart.



As a short-circuited stub was specified the length ℓ_{stub} will be a turn from $y_{short} = 0$ to b_{stub}:

$$\ell_{stub} = 0.362\lambda$$

Addition of the shunt susceptance to y_1 will result in

$$y_1 + jb_{stub} = 0.48 - j1.22 + j0.85 = 0.48 - j0.37 = y_{in}$$

The complete solution is shown in figure 5.38.

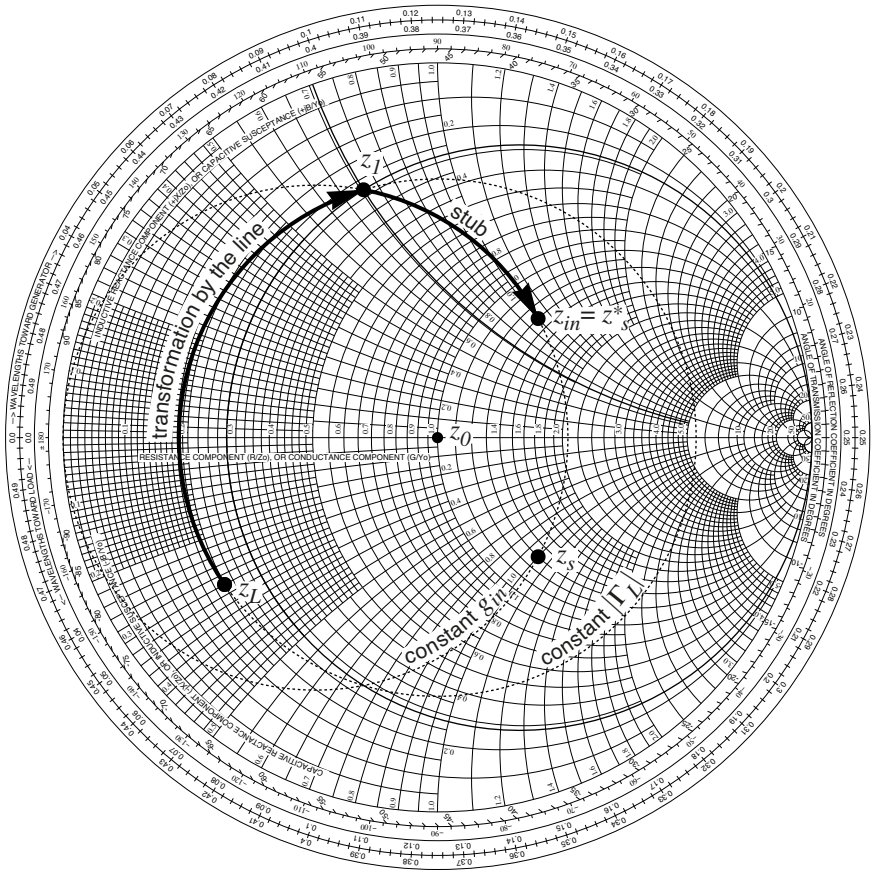


Figure 5.38 The complete solution to the matching network in example 5.9.

If there are bandwidth requirements that can not be fulfilled by single stub matching, the network can be extended by several sections. Refer the discussion concerning the lumped Pi network, figure 5.22 and 5.23

For laboratory work matching units are available based on two adjustable stubs having fixed connection points at a transmission line.

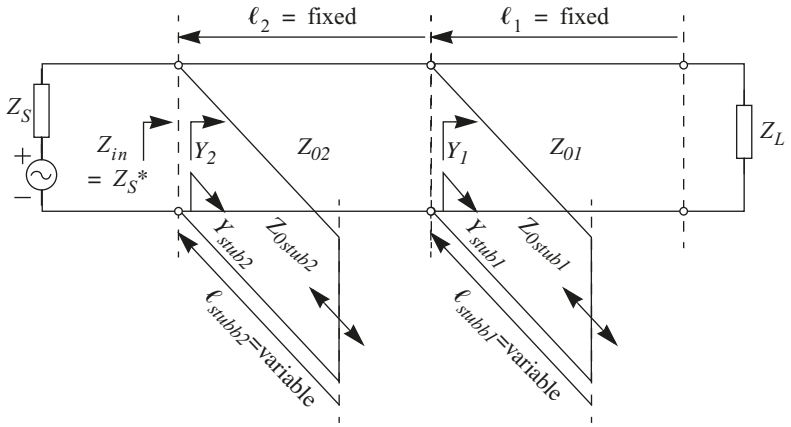


Figure 5.39 A matching unit using fixed lines but variable sub lengths.

Chapter 6

Introduction to HF Amplifier Analysis and Design

In the audio frequency range designing is simplified by the fact that there are amplifiers that can be regarded as ideal for many applications. For example, a general purpose operational amplifier operating at a couple of kHz exhibits many desirable features like voltage gain over 60dB, an input impedance of several MΩ:s and output impedance is in the range of a few Ω:s. This along with the possibility of adding feedback without risking instability enables us to more easily design voltage, current, transconductance and transresistance amplifiers with an optimal behaviour for each purpose.

However, when increasing the frequency up towards the MHz area, one soon realises that the former amplifier is now anything but ideal. An evident example of this is when designing active low pass filters where one sometimes incorrectly assumes that if one has made a 1kHz low pass filter, then signals at frequencies in the MHz range should be heavily attenuated. When evaluating the design it may be a painful experience to realise that this was not the case since our formerly ideal operational amplifier now has a gain close to unity.

This chapter will deal with the problems of analysing and designing amplifiers at higher frequencies. Focus is especially placed on what special considerations that have to be taken to get the wanted performance. Analysis on what trade-offs are needed between different amplifier parameters to fulfil the specifications. It would be nice to be able to design a low noise-high power-wideband amplifier with low current consumption, but generally one has to let go on some of the wishes. High frequency amplifier design is in this way an excellent example of the old saying that you always have to pay in one end to gain something in the other.

6.1 Transistor Model and Frequency Dependency

So why do amplifiers start to behave differently as the operating frequency is increased? To answer this question one has to look at the hybrid-π model of in this case a bipolar transistor, a technology still used in many high frequency applications

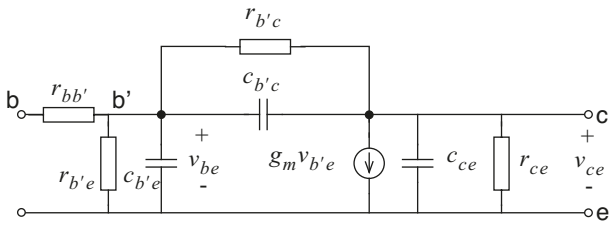


Figure 6.1 Hybrid-π model of a bipolar transistor.

In Figure 6.1 the different impedances connected between the nodes of the transistor are depicted. Not only are there resistive components along with a generator but also reactive and thus frequency dependent impedances. At DC, these capacitors could be disregarded having infinite impedance but as soon as frequency increases, so does also the influence from these capacitors

The notation of transistor parameters in the hybrid-π model varies significantly in the literature. In Table 6.1 some common parameters are listed along with definitions and alternate names.

Table 6.1 Hybrid-π transistor parameters for a bipolar IF transistor.

Parameter	Defined as	Typical value at $I_C = 1\text{mA}$	alternative notation
g_m	I_C / v_T	40mA/V	
$r_{b'e}$	β_0 / g_m	5kΩ	r_π
r_{ce}	V_A / I_c	50kΩ	r_o
$r_{bb'}$	device dependent	10Ω	
$r_{b'c}$	$\beta_0 \cdot \frac{V_a}{I_c}$	5MΩ	r_μ
$c_{b'e}$	device dependent $\tau_F \cdot g_m$	10pF	c_π
$c_{b'c}$	device dependent	0.5pF	c_μ, c_{fb}, c_r
c_{ce}	device dependent	1pF	c_o

Example 6.1 Transistor model at different frequencies

A:
What are the impedances in the hybrid-π model for a small signal RF npn transistor biased at 1mA at DC, 10 and 1000 MHz?

The transistor has the following parameters:

$c_{b'c} = 0.5\text{pF}$, $c_{b'e} = 9\text{pF}$, $c_{ce} = 1\text{pF}$

$r_{bb'} = 10\Omega$, $\beta \approx 300$, $V_A = 50\text{V}$

B:
How is the common emitter circuit affected by a load of $1\text{k}\Omega$ and a source impedance of 50Ω ?

Solution:

The missing transistor parameters can be calculated as:

$g_m = I_C / v_T = 1\text{mA}/26\text{mV} = 0.038\Omega^{-1}$

$r_{b'e} = \beta / g_m = 7.8\text{k}\Omega$

$r_{b'c} = \beta V_A / I_c = 15\text{M}\Omega$

$r_{ce} = V_A / I_c = 50\text{M}\Omega$

At DC all capacitors can be omitted from the model

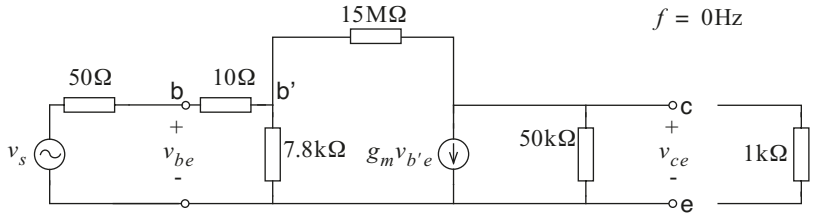


Figure 6.2 Hybrid-π transistor model at DC.

With a source impedance of 50Ω the voltage division on the input will be almost 1 making $v_s = v_{b'e}$. The output load is primarily the $1\text{k}\Omega$ load making the amplifier an almost perfect transconductance amplifier with the transconductance equal to g_m . The feedback from the output to input through the $15\text{M}\Omega$ resistor is marginal and the amplifier can be regarded as unilateral, i.e. isolated between input and output in the reverse direction.

If the same circuit is observed at 1MHz the reactances are included since they may now be significant.

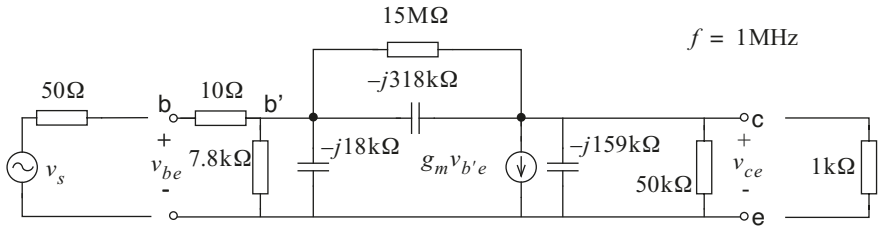


Figure 6.3 Hybrid-π transistor model at 1MHz.

The feedback path, although still with a high impedance, is now completely determined by the capacitor. The input impedance is now shunted by $-j18\text{k}\Omega$ which is of the same magnitude as $r_{b'e}$ and thus causes a low-pass behaviour on the input. At the output the total load is still determined by the load impedance.

Finally at 1000MHz (which is above the normal operating range of a transistor with the parameters as stated above) the capacitive components are quite dominating.

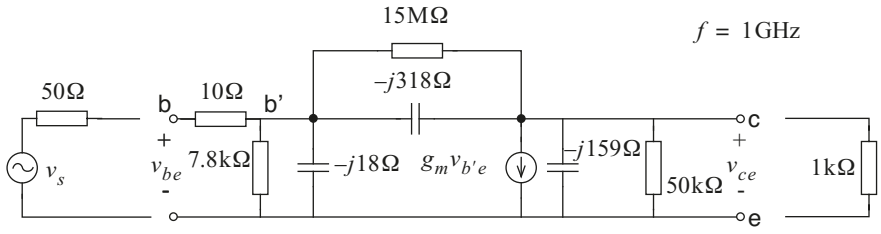


Figure 6.4 Hybrid-π transistor model at 1GHz.

We can now state that all resistances in the transistor are fully shunted by their respectively parallel capacitance. The input capacitance forms a clear low pass filter not only with the source impedance but also with $r_{bb'}$. Even if we had the opportunity of lowering the source impedance still a significant part of the signal would be attenuated at the input. Shunt feedback is provided by $c_{b'c}$ making the input impedance even lower. On the output, X_{ce} is now almost ten times lower than the resistive load causing low-pass behaviour as well a phase shift. At such a high frequency as 1GHz, the amplifier will have a quite different behaviour compared to the cases at lower frequency in that gain will be lower along with input and output impedance. There is also a clear reverse path making the input sensitive on the output load and vice versa. Plotting the amplifier gain versus frequency yields the following result.

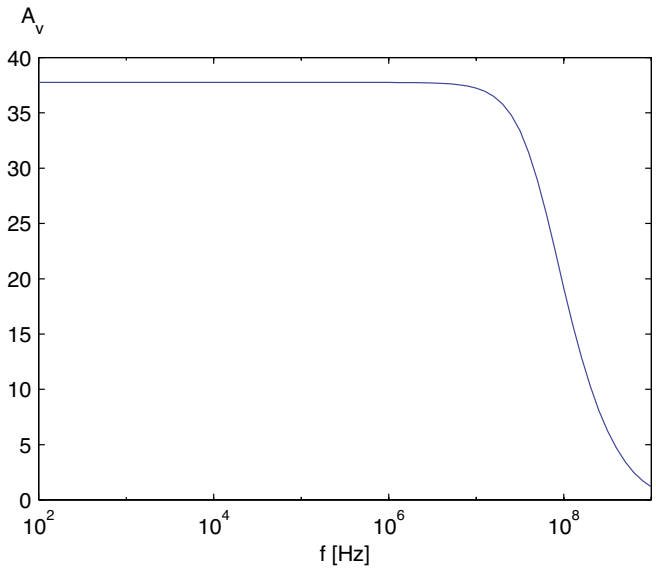
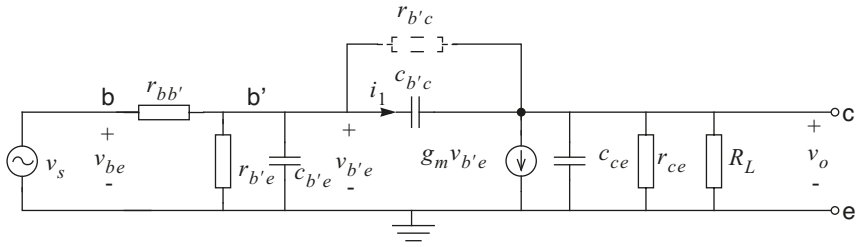


Figure 6.5 Voltage gain versus frequency for the loaded transistor.

The amplifier shows a clear low-pass behaviour due to the capacitive loading on both input and output as well as the feedback capacitor from collector to base. Determining exactly which capacitor that is the main cause behind the reduction in gain requires a more thorough analysis of the poles and zeroes of the circuit. The 3dB-breakpoint in the gain curve is also a function of both source and load impedances so it is not possible to talk about transistor bandwidth without specifying the circumstances.

A useful way of analysing the circuit is by the Miller theorem. The purpose if this is to transform the feedback capacitor $c_{b'c}$ to an equivalent capacitor at the input (and output).



The voltage source v_s causes a voltage over the node b' and a current i_1 through the feedback path (based on the results above the current through $r_{b'c}$ is neglected).

$$i_1 = (v_{b'e} - v_o)sc_{b'c}$$

the currents in the collector node will then be

$$-(v_{b'e} - v_o)sc_{b'c} + g_m v_{b'e} + \frac{v_o}{R_L \| r_{ce}} = 0 \Rightarrow v_o = \frac{sc_{b'c} - g_m}{sc_{b'c} + \frac{1}{R_L \| r_{ce}}} v_{b'e} \tag{6.1}$$

and thus

$$\frac{i_1}{v_{b'e}} = \left(1 + \frac{g_m - sc_{b'c}}{sc_{b'c} + \frac{1}{R_L \| r_{ce}}} \right) sc_{b'c} \tag{6.2}$$

which can be approximated by

$$\frac{i_1}{v_{b'e}} \approx (1 + g_m(R_L \| r_{ce}))sc_{b'c} = (1 + A_v)sc_{b'c} \tag{6.3}$$

This approximation is valid as long as $\omega c_{b'c} \ll 1/(R_L \| r_{ce})$ and $\omega c_{b'c} \ll g_m$.

The result of the Miller theorem is that the feedback capacitor can at the input be modelled as a capacitor between base and emitter (ground) and with a magnitude of $(1 + g_m(R_L \parallel r_{ce}))sc_{b'c}$. The behaviour can also be regarded as a capacitor between two nodes as in figure 6.6. When the voltage on the left side is raised 1V, the current should be $j\omega c_{b'c}$ if the right side was connected to ground (a). But since the node on the capacitors right side decreases $1 \cdot A_v$ volts (b), the total voltage over the capacitor is now $1 + A_v$ volts and thus the current is $(1 + A_v)j\omega c_{b'c}$ making the capacitor look like a $1 + A_v$ times bigger than in reality, viewed from the left side (c).

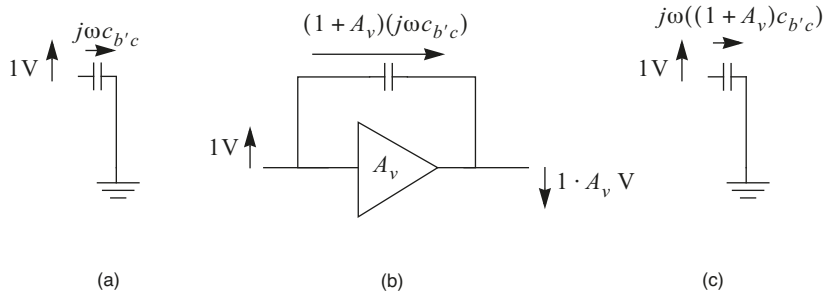


Figure 6.6 Miller equivalent of feedback capacitor.

By using the same method it can be shown that $c_{b'c}$ as seen from the output will look as it has a magnitude of $c_{b'c} \cdot (1 + 1/A_v) \approx c_{b'c}$. The Miller equivalent of the transistor is thereby reduced to the unilateral model in figure 6.7

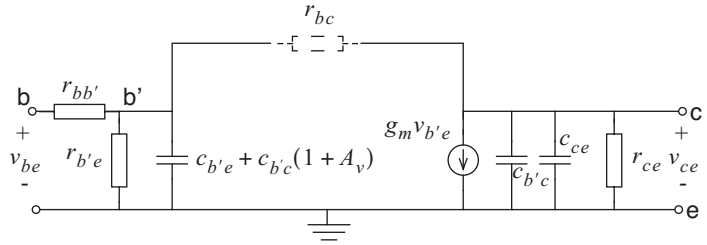


Figure 6.7 Miller equivalent of a bipolar transistor in CE coupling.

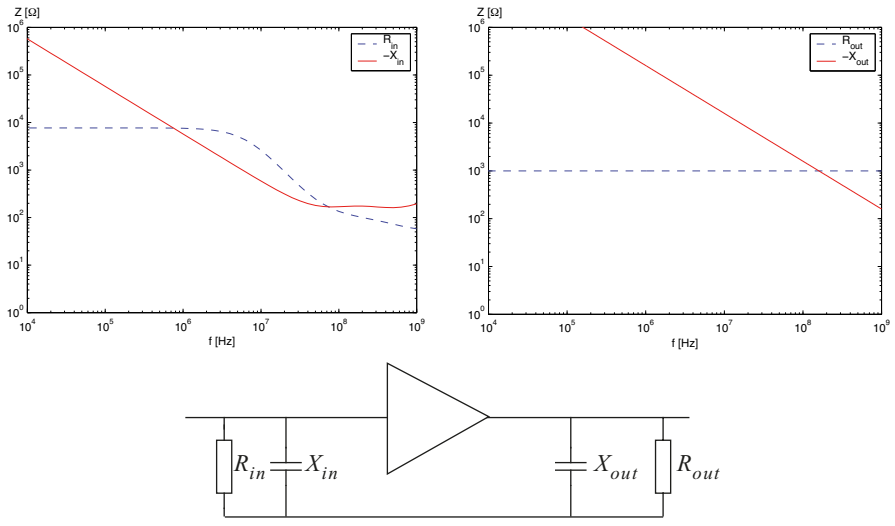


Figure 6.8 Input and output impedance as parallel equivalents. Results derived from a SPICE simulation.

The input and output impedance will vary with frequency as can be seen in Figure 6.8. At lower frequencies the input impedance is well approximated by the Miller theorem with $R_{in} = r_{b'e}$ and $X_{in} = 1/(j\omega(c_{b'e} + g_m R_L c_{b'c}))$.

At the same frequency range the output circuitry can be regarded as $R_L \parallel 1/(j\omega(c_{b'c} + c_{ce}))$. When frequency is increased both the output and input reactance magnitude is reduced while the resistive part stays fairly constant. When approaching and passing 100MHz, especially the input impedance changes in a non-uniform way. This is because the feedback and thus reduced gain changes the Miller effect.

If the transistor is to be used in the upper frequency range, simulation is definitely recommended and then with the actual transistor model instead of the current source equivalents used here. The main benefit of the current source approach is that it is possible to do hand calculations to verify the result, or at least its credibility within an order of magnitude.

6.2 Tuned and Compensated Amplifiers

In the previous section it was shown how the capacitive effects of the transistor could be disregarded at low frequency but caused great impact in the MHz range. To design for example a 77 MHz IF amplifier with good performance, one in some way must compensate for these effects to avoid mismatch and thereby gain reduction. One effective way to reduce these effects is simply to tune out them using a corresponding reactance but with the opposite sign. If the transistor in example 6.1 was loaded not only by the resistance R_L on the output but also with an inductance of suitable size, the capacitance would be neutralized and the gain would increase at the resonant frequency, see figure 6.9

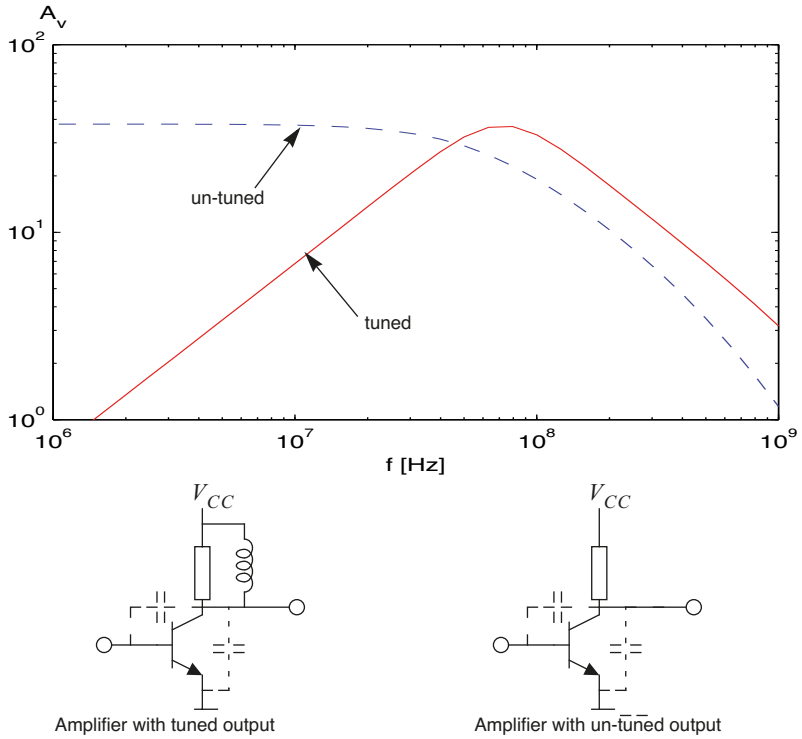


Figure 6.9 Gain of transistor amplifier with tuned and un-tuned output.

By tuning the output it is thus possible to extend the maximum gain to a higher frequency. The side effect is that the amplifier will have a band pass frequency response. The bandwidth of this response is actually the same as the single side bandwidth in the un-tuned case. Consider the tuned circuit in figure 6.9 with resistor in parallel with a capacitance and an inductance. The Q value of that circuit will be

$$Q = \frac{f_o}{B_{3dB}} = \frac{R}{X_L} = \frac{R}{X_c} = R\omega_0 C \Rightarrow B_{3dB} = \frac{1}{2\pi RC} \tag{6.4}$$

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{6.5}$$

By changing the parallel inductance L, the centre frequency will be moved but the 3 dB bandwidth B_{3dB}, being a function of R and C, will remain the same. In the extreme case, the inductance will be infinite and we will end up with the un-tuned case in ($f_0 = 0$) figure 6.9.

The tuned amplifier in figures 6.9 has only an inductance to neutralise the capacitance and thus tune the output of the transistor. A more common practice is to add both one inductance and one capacitance. This to be able to more freely choose the values of the components as well as making the tuning less sensitive to the internal capacitances of the transistor.

Example 6.2

Design an output network that has a resonant (=centre) frequency of 10.7 MHz for a transistor amplifier. Use the parameters from example 6.1.

The output capacitance for the transistor is $C_{tot} = c_{ce} + c_{b'c} = 1.5\text{pF}$ and this can be tuned by a parallel inductance of $L = 1/(\omega_o^2 C) = 147\mu\text{H}$. This may solve the problem in theory but designing a coil with such a large inductance at 10.7MHz will be quite difficult if one wants to avoid self-resonance in the coil. A better solution is to add a parallel capacitance of say 47pF to the already existing 1.5pF, thereby reducing the inductance to a much more feasible $4.6\mu\text{H}$. With this solution there is also the possibility of making the external capacitor tuneable, either mechanically or electrically.

In the same way as in the example with the output capacitance, the input and feedback capacitances can be compensated. By adding an inductance in parallel with the feedback capacitor $c_{b'c}$ the feedback now forms a parallel resonant circuit with an impedance maximum at ω_0. This technique enables a cancellation of the feedback in a limited frequency range and may be effective as such. However, determining the magnitude of the inductance as well as connecting it in a way so stray effects are avoided is a quite difficult task. The new feedback path is also prone to cause instability far away from the resonant frequency.

6.3 Choice of Topology

If one takes apart a radio from the 70:s or early 80:s one will notice that the device is almost completely built from discrete components and there are lots and lots of components to adjust, both resistive and reactive. Beside the fact that it is almost impossible *not* to turn on at least one of these screws to see if there is any effect it was a technique possible when volumes were small and almost each radio application in each country had its own frequency range, bandwidth and type of modulation. Today when many radio systems are standardised on a worldwide basis and radio equipment are made in much larger volumes, there has also emerged a need for different types of amplifiers.

1. Specialised amplifiers, often integrated circuits that are designed for maximum performance in a certain system with a strict set of demands. This can be power amplifiers for wireless local-area-networks (LAN) and low noise amplifiers for mobile radio front-ends. Often these are integrated together with other radio functions as mixers to make the solution more cost effective.
2. General wideband amplifiers with almost optimal performance regarding noise, output power and gain over a large bandwidth, usually from DC up to the GHz range. These may sometimes be programmable or part of a larger series to further extend the application area. The principle here is to sell few items to many applications as opposed to the previous class. These are also usually integrated amplifiers or hybrids, discretely built up amplifiers in a common package.
3. Custom amplifiers designed to get absolute maximum performance in one certain area. Combined with the fact that these amplifiers may not go into production in any large volumes, this makes it possible to still build these amplifiers from discrete components and have them individually tuned. It is a costly procedure but sometimes necessary when specifications are hard to fulfil.

In this chapter we will mainly focus on the third group. With the knowledge of how to design high frequency amplifiers “from scratch” one will also gain knowledge when it is possible and more cost effective to use a general solution. One can also see what trade-offs that have to be made between the different amplifier parameters.

6.4 Discrete Amplifiers

A discrete transistor amplifier can be designed using a common emitter (CE), common base (CB) or common collector (CC) coupled transistor. The CC coupling is less common in high frequency design since the load on the emitter provides a feedback that is fixed to a voltage gain of unity. Besides instability is a large problem since the emitter is not directly connected to ground (this is actually a way of making oscillators-adding a reactive component between device and ground to cause instability). The common base amplifier is widely used in high frequency design thanks to its low input capacitance at high voltage gain. The current gain is however always unity due to feedback. The only coupling not providing any feedback from either load or source is

CE (it still has feedback from the internal transistor paths but this is not related to the coupling). The advantage of this is that current and voltage gain can be determined by the designer. There is also a much wider variety of making external feedback loops.

Comparing the CB and CE amplifiers for different parameters give that the CE amplifier is preferred for most applications but we will encounter cases where the CB amplifier has its benefits.

Table 6.2 Comparing CE and CB amplifier performance.

Parameter	Common emitter	Common base
Voltage gain	$-g_m R_L$	$g_m R_L$
Current gain	β	1
Output power	Limited by supply voltage and collector quiescent current	As CE but output voltage swing is limited by input voltage swing
Noise figure	Very good. Often determined by thermal noise from $r_{bb'}$	Fair. Often determined by shot noise in collector current
Input impedance	High, with high input capacitance due to Miller effect.	Low, but with low input capacitance since it is non-inverting
Output impedance	High	Very high due to feedback
Main application	All sorts of amplifiers	High frequency amplifiers and oscillators

6.4.1 Choosing the Operating Point

The design of an HF-amplifier is often aimed at fulfilling a certain requirement, usually gain or power or in some tricky cases: both. No matter which, the crucial point is still the operating point when no AC-signal is applied on the input. For a CE amplifier with a a parallel resonant circuit as collector load as seen in figure 6.10 there are two requirements regarding I_C.

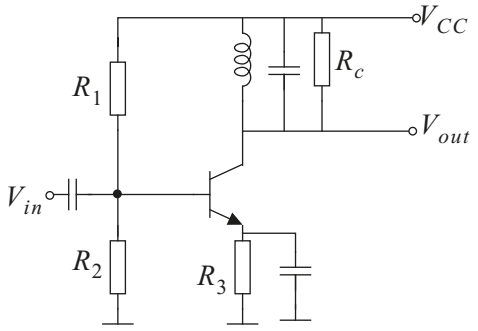


Figure 6.10 CE amplifier with resonant collector circuit.

$$A_v = -g_m \cdot R_c = -\frac{I_C}{v_T} \cdot R_c \tag{6.6}$$

$$max(P_{R_c}) = min\left(\frac{I_C^2 \cdot R_c}{2}, \frac{(V_{CE} - V_{CEmin})^2}{2R_c}\right) \tag{6.7}$$

Equation (6.6) shows that given a certain load R_c on the collector, g_m and thereby I_c must be chosen to give a certain amplification. In equation (6.7),where V_{CEmin} is the collector-emitter saturation voltage, states that to provide a certain output power in the load, the operating point values I_C and V_{CE} should be large enough. The maximum expression indicates that the output power could either be limited by the collector current or the collector-emitter bias voltage. To assure that the output power is large enough both have to be evaluated.

By transforming the connected load via a network like for example a tap or transformer one can yield an arbitrary load impedance. This is however still an under-determined system where there are more parameters than equations. To fulfil the design one has to consider less obvious aspects such as power consumption, component availability, available area on the circuit board and cost. If there are no such additional demands on the design, the designer has to determine one parameter and calculate the others based on this. The design process can under those circumstances be iterative before one has come up with a solution with reasonable parameters and component values.

The bottom line is that there is no perfect or single solution given a specification. The designer must compare pros and cons and make the best possible amplifier under the circumstances.

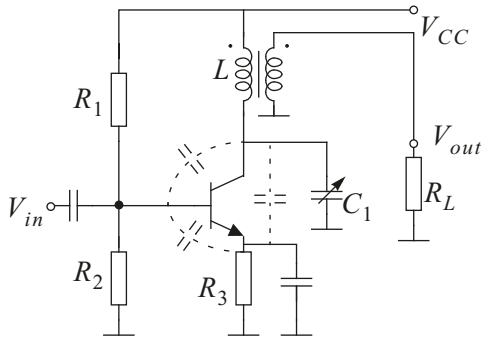


Figure 6.11 CE coupled small signal amplifier with tuned collector circuit.

Example 6.3 Design of a 10.7MHz amplifier

Design a small signal amplifier with $f_0 = 10.7\text{MHz}$ and $B_{3dB} = 500\text{kHz}$ using the circuitry in figure 6.11. The amplifier should have an output power of at least 0.5 mW and a gain of at least 10 assuming that the load impedance is 50Ω The transistor parameters are $\beta_0 = 100$, $c_{b'c} = 1\text{pF}$, $c_{b'e} = 10\text{pF}$, $c_{ce} = 0.5\text{pF}$, $V_A = 50\text{V}$.

One possible solution

The Q-value of the loaded collector circuit is $f_0/B = 21.4$. Since the output should be matched for maximum power gain, the transformed load R_L' should be equal of that already at the collector. This leaves us with some initial demands on the design.

- $QX_L = R_c = R_{pL} \| r_o \| R_L'$ to assure that bandwidth demands are met
- $R_{pL} \| r_o = R_l'$ makes the output matched.
- $g_m \cdot R_c \cdot \frac{n_s}{n_p} = 10$ is the specified gain and finally
- $P_{amp} \geq 2\text{mW}$, this since there are two equal loads on the collector, the output impedance $R_{pL} \| r_o$ and the transformed load R_L'. The efficiency of the transistor is maximum 50%[1] and that leaves 25% for each of the loads.

At this point one must make the now famous decision to set one or a few parameters and calculate the rest. It might be useful to have some kind of math/spreadsheet computer program to assist in making the calculations since many parameters interact and the design might be revised many times.

In this design the choice is to keep the total collector load in the area of $1\text{k}\Omega$. This means that it will be possible to use standard-Q coils having a loss equalling a parallel resistance in the area of $3 - 5\text{k}\Omega$ and an unloaded Q-value $Q_u \approx 30$. The load R_L should now be transformed to $2\text{k}\Omega$. This gives a transformer turn ratio n_p/n_s of

6.3. From this, g_m can be calculated as

$$g_m = \frac{10n_p}{R_c n_s} = 63\text{mA/V} \Rightarrow I_c = g_m \cdot V_T = 1.64\text{mA}$$

which is a realistic value for a design like this. Checking the amplifier output power shows that the amplifier is simultaneously current and voltage limited if V_{CE} is 2.7V. To get the specified output power V_{CE} can be as low as 2V. With either configuration the amplifier can run off a 3.3V supply and there will still be enough voltage for the emitter resistor. Here we chose $V_{CE} = 2.7\text{V}$ and

$$V_{Re} = 0.6\text{V} \Rightarrow R_e = \frac{V_{Re}}{I_c(1 + 1/\beta_0)} = 362\Omega$$

With the operating point and g_m determined one can also calculate $r_o = V_A/I_c = 30.5\text{k}\Omega$. This means that the parallel resistance from the coil must be slightly more than $2\text{k}\Omega$ to make $R_{P_L} \| r_0 = 2\text{k}\Omega$.

$$R_{p_L} = Q_u \cdot \omega_0 L = 2.15\text{k}\Omega \Rightarrow L = 2.15\times10^3/(\omega_0 \cdot Q_u) = 1.06\mu\text{H}$$

This gives us a value of the inductance provided that its Q_u is known. Another possibility if Q_u cannot be chosen arbitrarily is to add a resistance R_{cc} in parallel with the primary coil of the transformer so $R_{pL} \| r_o \| R_{cc} = 2\text{k}\Omega$.

The observant designer has by now noticed that many of the values are not exactly “off the shelf” but must be replaced by standard values. This is another reason why the design should be made in a spreadsheet environment since one can have the design updated for each parameter choice and checked that performance meets the initial specifications.

The next step in the design of the collector circuit is to determine the tuning capacitors. With $f_0 = 1/(2\pi\sqrt{LC})$ where L is already determined, the total collector capacitance must be $C = C_t + c_{b'c} + c_{ce} = 221\text{pF}$.
It is wise to accomplish this by using one 200pF fixed capacitor and another 20-50pF tunable one in parallel.

Making the bias network is rather straightforward since $V_B = V_E + 0.7\text{V} = 1.3\text{V}$ and the high β_0 makes the base current I_B low, only $16.4\mu A$. If we assure that the current through R_1 and R_2 is much larger than the base current the voltage divider can be treated as unloaded.

$$V_{CC} \cdot \frac{R_1}{R_1 + R_2} = 1.3\text{V}$$

We previously realised that $V_{CC} = 3.3\text{V}$ would be sufficient for the power demand and this together with $I_{R_{1,2}} \gg I_B$ yields that $R_1 = 3.9\text{k}\Omega$ and $R_2 = 6.1\text{k}\Omega$. These are neither crucial values,

they can often be a factor 2 larger or smaller without any severe effect. As long as the transistor input impedance is low enough, only a negligible part of the input signal is lost in the bias network.

The final step in the design is to determine the coupling capacitors. A node can be considered as decoupled if the impedance of the capacitor connected to the node is significantly smaller than the nodes resistive part. For the decoupling capacitor on the emitter this means that

$$X_{C_e} = \frac{1}{\omega_0 C_e} \ll \frac{1}{g_m} \parallel R_e = 15.1\Omega$$

resulting in $C_e = 10\text{nF}$. For the base bypass capacitor the calculation is somewhat more complicated. The input impedance should be $r_\pi \parallel X_{c_{b'e}}$ but since there is also c_{bc} that is magnified through the Miller effect we will have an almost purely capacitive input. In this example,

$$Z_{in} = r_\pi \parallel jX_{C_{b'e}} \parallel jX_{C_{b'c}}(1 + |A_u|) = 1580\Omega \parallel -j349\Omega$$

Choosing a input capacitor of 1nF should be sufficiently large. Another brute force method is to determine which is the largest capacitor available that has a resonant frequency well above f_0 and use that one. It may though result in an unnecessary large design. It is often advisable to decoupled the emitter with another capacitor 100 times smaller than the first one to assure that even very high frequencies are decoupled properly.

6.4.2 Common Base Amplifier

As seen from the CE amplifier calculations, the base-collector capacitance can appear as very large and yield a low input impedance at high frequencies. To prevent this there is the possibility to vary the base-emitter voltage through the emitter instead of the base. This configuration, pictured in figure 6.12 is called common base (CB) configuration. It has good high frequency characteristics but the low input impedance and high output impedance makes it less common in IF stages where one often wants to stack several similar stages after each other with low interstage loss.

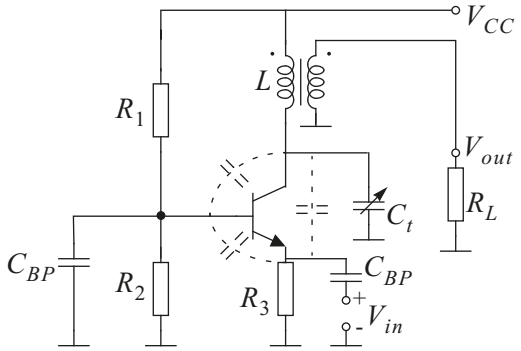


Figure 6.12 Common base amplifier with transformer coupled collector.

The transistor hybrid-π model in a CB configuration is depicted in figure 6.13

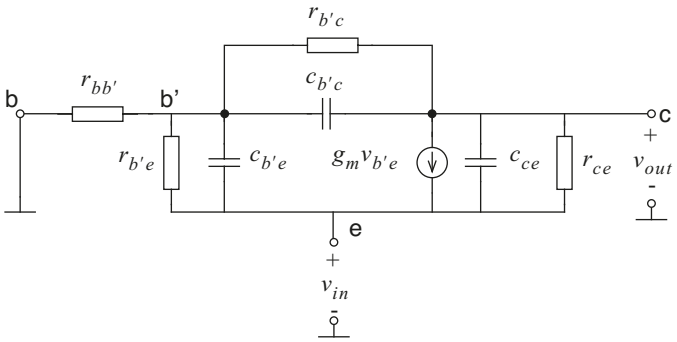


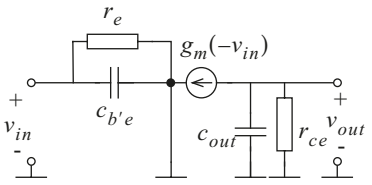
Figure 6.13 Bipolar transistor in CB configuration.

What differs most between the CB- and CE configuration is the input impedance and that a CB amplifier is non-inverting. The input impedance can be calculated assuming that r_{ce} is large and C_{ce} negligible. Calculating the input current yields that one finds that

$$I_{in} = g_m \cdot V_{in} + V_{in}/r_{b'e} + V_{in} j\omega c_{b'e} \tag{6.8}$$

$$Z_{in} = 1/(g_m + g_{b'e} + j\omega c_{b'e}) \tag{6.9}$$

This implies that one can make a condensed model of the CB amplifier according to figure 6.14, based on a model suggested by [2].



$$r_e = \frac{1}{g_m}\left(\frac{\beta_0}{1+\beta_0}\right) \qquad c_{out} = c_{ce} + c_{b'c}$$

Figure 6.14 Condensed model of a transistor in CB configuration.

6.5 Designing with Amplifier Modules

In radio designs of today with a higher and higher degree of integration it is more likely that one will use some kind of integrated amplifier that consist of a few transistors that can be configured and biased in different ways. Depending on the demands on the amplifier one may choose to use a circuit with everything fixed and no external components. The disadvantage is that these ready made general purpose amplifiers have somewhat lower performance than if you make the full design by yourself.

6.5.1 The CA3028 Differential Amplifier/Cascode

An example of a integrated circuit well suited for IF amplifier designs is CA3028 from Harris semiconductor[3]. The amplifier, pictured in figure 6.15, has a differential stage on top of a single transistor. In addition there are resistors for biasing. With a few external components it can be configured as a differential amplifier with current generator on the emitter or, as in figure 6.15, as a cascode amplifier with voltage controlled gain (VCG) capability. The cascode is a two transistor amplifier where the first transistor Q_1 is in CE configuration and the second Q_2 in CB. The other transistor in the differential pair Q_3 should be regarded as disconnected for the time being but will contribute in the VGC -mode described later in this chapter.

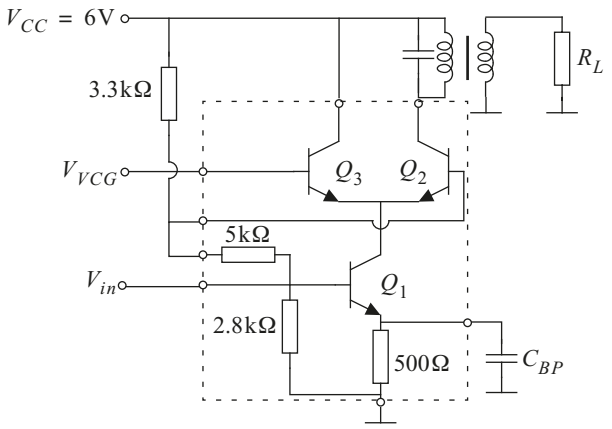


Figure 6.15 IF amplifier CA3028 connected as cascode with voltage controlled gain capability. Internal components are within the dashed box.

The main benefits of using a cascode instead of a single CE stage is not as one might think gain but that the input impedance is not affected by the Miller effect. Excellent isolation between input and output also makes the amplifier stable. For a CE stage, the input impedance is $Z_{in} = r_\pi \| jX_{Cbe} \| jX_{Cbc}(1 + |A_u|)$. This means that at high gain the influence of C_{bc} will be significant. Since the gain of the CE stage is $-g_{m1}R_c$ one must analyse the load on the collector to calculate the gain and thereby the input impedance. The same collector current will run in both the transistors leading to an input impedance on $Q_2's$ emitter of $1/g_{m2} = 1/g_{m1}$ making the voltage gain for Q_1 unity. With this low voltage gain the influence from C_{bc} can often be disregarded compared to C_{be}. For the second stage the voltage gain is $g_{m2}R_c$ and in addition to this the turns ratio of the transformer n_p/n_s.

Example 6.4 Voltage gain for a cascode amplifier

Calculate the voltage gain of the cascode in figure 6.15 if $R_L = 300\Omega$ and the turns ratio of the loss-less transformer is 3:1.

First calculate the base voltage and from this the emitter voltage and current

$$V_{B1} = 6 \cdot \frac{2.8\text{k}\Omega}{2.8\text{k}\Omega + 5\text{k}\Omega + 3.3\text{k}\Omega} = 1.51\text{V}$$

$$V_e = V_B - 0.7\text{V} \Rightarrow I_C \cong I_E = \frac{0.81}{500} = 1.6\text{mA}$$

The transformed load is the only load on the collector since the cascode configuration has increased output impedance compared to a CE stage. This gives the gain as

$$A_u = \frac{I_C}{v_T} \cdot (n^2 \cdot R_L) \cdot \frac{1}{n} = 0.062 \cdot 300 \cdot 3 = 55.8 = 35\text{dB}$$

To use the amplifier in VCG mode the full differential pair must be used. If the gain control voltage V_{VGC} is the same as that of the base voltage of Q_2, then the collector currents in Q_2 and Q_3 will be equal. The emitter currents of Q_2 and Q_3 will then add up as collector current of Q_1.

$$I_{C1} = I_{C2} \tag{6.10}$$

$$I_{C1} = I_{C2} + I_{C3} \Rightarrow g_{m1} = \frac{g_{m2}}{2} = \frac{g_{m3}}{2}$$

The load on the collector of the input stage will be the same, see equation (6.11), but since only half of the collector current I_{C1} runs in Q_2, the gain will now be 50% lower than if the VCG voltage was zero.

$$\frac{1}{g_{m2}} \parallel \frac{1}{g_{m3}} = \frac{1}{g_{m1}} \tag{6.11}$$

By increasing the VCG voltage on Q_3 the current I_{C3} will increase and thereby cause a decrease in I_{C2} so that equations (6.10) and (6.11) are still valid. This causes the gain to fall and we will thus have a voltage controlled gain. This voltage may be derived from a detector in a radio receiver and thereby keep the signal within limits to avoid saturation in the amplifiers. This function is called automatic gain control (AGC) and is necessary as soon as there is information in the signal amplitude.

Many times even this integrated amplifier has too low integration to meet the demands on the latest radio equipment. Often the manufacturer tries to incorporate several IF amplifiers and also a detector in the same circuit (for example Philips 3189 IF amp and FM detector). The advantage is that the footprint will be much smaller but on the other hand the circuit will be system specific. For portable radio equipment such as mobile telephones, the demands on size

are very strict and production series so large that one usually makes a set of custom chips for each specific system to achieve maximum performance. Mobile telephone systems like GSM and DECT as well as PAL TV systems all have their own chip-sets providing just the necessary system functions to reduce cost and size.

6.5.2 Mini-Circuits™ MAR family

Another example of integrated amplifiers in the RF area are the MAR amplifiers from Mini-Circuits™. This is a family of amplifiers that are designed to meet certain specifications regarding output power, gain and noise.

MONOLITHIC AMPLIFIERS 50 Ω

BROADBAND DC to 2 GHz

MAR

up to +12.5 dBm output

MODEL NO.	❁FREQ. MHz f_L	f_U	GAIN, dB Typical (at MHz) 100	500	1000	2000	note 1 MIN.	MAXIMUM POWER, dBm Typ. Output (1 dB Comp.)	Input (no damage)	DYNAMIC RANGE NF dB Typ.	IP3 dBm Typ.	VSWR Typ. (:1) In	Out	ABSOLUTE MAXIMUM RATING[7] (25 °C) I (mA)	P (mW)	DC POWER at Pin 3 Current (mA)	Volt Typ.	THERMAL RESISTANCE[6] θjc °C/W	CAPD DATA (see RF/IF Designer Handbook) Page	Case Style Note B	CONNECTION	Price $ (Qty 30)
❑ MAR-1	DC	1000	18.5	17.5	15.5	—	13.0	+1.5	+13	5.5	+14.0	1.3	1.3	40	200	17	5.00	105	3-22	VV105	cb	0.99
❑ MAR-2	DC	2000	12.5	12.3	12.0	11.0	8.5	+4.5	+13	6.5	+17.0	1.3	1.4	60	325	25	5.00	95	3-22	VV105	cb	1.12
❑ MAR-3	DC	2000	12.5	12.2	12.0	10.5	8.0	+10.0	+13	6.0	+23.0	1.5	1.7	70	400	35	5.00	105	3-22	VV105	cb	1.19
❑ MAR-4	DC	1000	8.3	8.2	8.0	—	7.0	+12.5	+13	6.5	+25.5	1.6	2.0	85	500	50	5.25	90	3-23	VV105	cb	1.29
❑ MAR-6	DC	2000	20.0	18.5	16.0	11.0	9.0	+2.0	+13	3.0	+14.5	1.5	1.4	50	200	16	3.50	110	3-23	VV105	cb	1.66
❑ MAR-7	DC	2000	13.5	13.1	12.5	11.0	8.5	+5.5	+13	5.0	+19.0	1.4	1.5	60	275	22	4.00	110	3-23	VV105	cb	1.31
❑ MAR-8	DC	1000	32.5	28.0	22.5	—	19.0	+12.5	+13	3.3	+27.0	#	#	65	500	36	7.80	130	3-24	VV105	cb	1.27

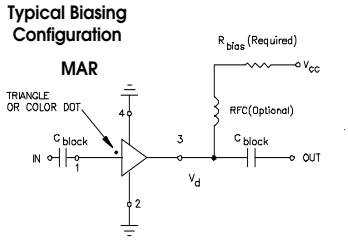


Figure 6.16 Data sheet for the MAR amplifier family (Reprinted by permission of Mini-Circuits™).

The design phase here is to pick a member of the family with data that meet ones specifications and in case they all can not be met, compromise. One major advantage of these amplifiers (as with any pre-made module - there are lots of others) is that they demand very few external components, usually just a biasing resistor and two bypass capacitors for input and output, see figure 6.17. For many applications they are a convenient way of achieving amplification when one does not have extreme demands. The amplifiers are fairly cheap and are excellent for prototyping and evaluation purposes. In the final design it is likely that many components are instead integrated on one chip, but before one is at that stage, there might be lots of investigations and alterations. Since the internal resistors are fixed all one has to do is to attach a proper resistor to the output. From the data sheets in figure 6.16 one can find what DC voltage and current that should be applied to the output pin for

proper operation. If one uses MAR1 in a 12V system the data sheet specifies that the base voltage should be 5V and the current 17mA. The resistor is then calculated as $R_{bias} = (V_{CC} - V_{DC})/I_{DC} = 410\Omega$. The coil is a RF choke to prevent loss of the output signal in the bias resistor. This is chosen to be as large as possible. One upper limit is that a coil goes into resonance at high frequencies but as long as the impedance is large it does not matter if it is capacitive or inductive. The coupling capacitors on the input and output should be large enough to provide an impedance of only a couple of Ω:s at the selected frequency. Even though the modules are wideband one should always include some input and output filtering to reduce noise and stop interfering signals from driving the amplifier into saturation.

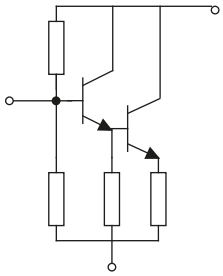


Figure 6.17 MAR amplifier circuit layout. (Reprinted by permission of Mini-Circuits™).

There is little need for matching the input and output as long as the biasing follows the specifications. A typical value of the input and output standing wave ratio (SWR) is 1:1.5 - 1.8 giving a loss below 0.5 dB per port from DC-1GHz which is sufficient for most applications but very low noise amplifiers. One of the reasons that the amplifier shows so good SWR over such a large frequency range is that there are no frequency selective elements other than stray effects inside the packaging and that the amplifier has both serial and parallel feedback, see Figure 6.17.

The wideband properties puts fairly high demands on circuit layout. If it is recommended to use ground planes and well decoupled DC voltages with a “normal” amplifier it is absolutely necessary here. Since the amplifiers have gain up to several GHz, only a few pF of stray capacitance may cause instability. Further demands and design recommendations can be found in [4]

6.6 Alternatives to Compensating

In this chapter we have assumed that unwanted capacitive effect could be compensated out, at least for a limited frequency range, leaving an almost ideal transistor. But we have also encountered how the gain and impedance levels fall off rapidly when frequency is increased. If we further replace our transistor model with a more complex one like the Gummel-Poon models used by, e.g., SPICE we will have something like 60 transistor parameters that all depend on operating point as well as each other in a delicate way. When frequency is so high that all these parameters become significant, it will be an impossible task to do effective amplifier design by trying to compensate out one transistor component. By doing so we will inevitably also affect a lot of other parameters, parameters that we had no intention of chang-

ing and the amplifier design runs the risk of becoming a huge iterative process. When it comes to MOS transistors the situation is even more complicated since each parameter set has a limited operating range. Depending on the layout of the transistor not only the parameters but also the model type must be updated. Even with qualified simulation software it is hard to achieve a valid result due to the complicated process along with many parasitic effects.

Some unwanted stray effects may not be possible to neutralise simply because one can not get direct access via the bonding pads on the device. For example, neutralising the base-collector capacitance seems like an achievable task but if $r_{bb'}$ along with stray inductances in the bond wires must considered, as in the GHz range, it becomes much more complicated. Another problem is to extract the hybrid-π parameters for the device. Some can be found in data books and some, but definitely not all, can be measured directly but what about the others? Of course, something is always better than nothing but making simulations with a model that has some many flaws that the result must seriously be doubted is not an effective way of designing radio circuits.

Since it is so hard to get a general overview on what to expect from a device at such high frequencies an alternative is needed. Most transistor manufacturers supply their devices with various two-port parameters, usually S, y or z parameters. These parameters say nothing about the actual device and its physical properties. it may be a bipolar, MOS or whatever device in any packaging. The important thing is the parameter set between the different nodes of the device. Using such a parameter set makes microwave design easier but gives little information of what to expect if the conditions where the parameters were obtained changes. The best solution may be to have a good knowledge of the interior process of the device combined with any contribution from the package. From this, deviations from the expected result can be explained and hopefully corrected. The parameter approach can then be used in a much more effective way than if the device is just considered to be a “black box”.

6.7 The Bottom Line

How the IF amplifier should be designed and what integration level one will choose is much a matter of application. Doing the design from scratch gives many degrees of freedom and the possibility of including system specific considerations. The drawback is that it may take a lot of time and experimental work. The ready made amplifier IC/hybrids of today are often sufficient for most applications since there is a variety to choose from. A design often has its goals in a specific gain or power level but even things as noise, size, power consumption are important. If the amplifier should be mass produced one also have to think about cost, testability and reliability.

The important part is that the designer knows about the topologies available and chooses the best one according to specifications, well aware that there might be another solution with somewhat better performance if one just had that little extra time.

6.8 References

[1] P. H. Young, *Electronic Communication Techniques*, Charles E. Merill Publishing Company, 1994.

[2] P. Gray and R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, third edition, John Wiley & Sons Inc. 1993.

[3] Harris Semiconductor, *Linear & Telecom ICs for analog signal processing applications*, 1991

[4] Mini Circuits, *RF/IF Designers Handbook*, 1992.

Chapter 7

High Frequency Transistors

As the frequency of operation reaches 1GHz and beyond transistors are typically operated close to their frequency limit. Even though there are transistor technologies commercially available that can produce useful power at 100GHz or more using better devices than necessary can not be economically justified. Cost increases rapidly with increasing frequency capabilities.

A plethora of various transistor technologies are available from the very basic silicon bipolar transistors (BJT) and field-effect transistors (FET) to more advanced and complex technologies such as high-electron-mobility transistors (HEMT). Today the evolution of semiconductor technology has even pushed the capabilities of the basic transistor technologies such that there are now products commercially available that can operate at 5GHz or more and the development continues.

There are basically two ways to increase the maximum frequency of operation of transistors. One is the downscaling of critical structures of the intrinsic transistor. Simplified, for bipolar type of devices it is the width of base that controls the frequency capabilities whereas for the FET devices it is the gate length. The second approach is to use more clever transistor structures and materials to increase the mobility of the charge carriers, reduce parasitic elements and obtain high saturation drift velocity.

The previous chapter serves as an introduction to the design of high frequency amplifiers where the starting point is low frequency design based on the hybrid-π model for the bipolar transistor. As the frequency increases it is demonstrated that the influence of the reactive components increases accordingly. To some extent this can be solved by introducing tuning and hand calculations are still reasonably accurate.

The next chapter also deals with the design of high frequency amplifiers but the approach is completely different. For one thing we do not longer rely on hybrid-π models or the like. The reason is that there will be too many significant elements in the models. These does not only come from the intrinsic transistor but also from the actual package carrying the transistor chip. A simplification is necessary but with a preserved accuracy. Another issue is the fact that circuit elements like the structures on the circuit board must be considered as distributed which complicates the design work even more when using the traditional low frequency methods.

The purpose of this chapter is to present an overview of the most common transistor technologies available for radio and microwave frequencies as well as presenting measures used to rate frequency performance of transistors. Also, the influence of the package will be discussed. From this knowledge together with the previous chapter the design approaches discussed in the next chapter will come as no surprise to the reader. Here, we rely on the complex behaviour of a transistor, including the package if necessary, being captured in a black box model represented by just a few parameters.

7.1 Transistor Technologies

Two fundamentally different transistor technologies exist, the bipolar junction transistor (BJT) and the field-effect transistor (FET). In the BJT positive and negative charge carriers diffuse through the two junctions (collector-base and base-emitter) between the collector and the emitter controlled by an external current into the base. In a FET a channel is acting as a current source between drain and source controlled by a gate voltage. The channel can be of enhancement type or depletion type, i.e., a gate voltage must be applied to build up a channel or to deplete an existing channel, respectively.

For each basic technology one can choose materials different from silicon, otherwise by far the dominant semiconductor material. Another possibility is to mix material for different regions of the transistor. When only one material is used for all regions such as silicon-to-silicon or germanium-to-germanium (varying only in doping profiles and densities) the transistor is said to be a homojunction transistor. When two completely different materials are used such as germanium-to-gallium-arsenide the term heterojunction is used. In addition to material options, there are many ways to realise the actual structure of a transistor, especially when it comes to FET devices with the actual channel and the control thereof.

The most common technologies will be briefly described below as well as their relation to other technologies. However, it is not the purpose of this text to treat semiconductor physics of these devices nor their properties in more detail. Only a qualitative description will be given. For more information about semiconductor physics refer to, e.g., [1][2] and for microwave devices in particular [3]. A more designer-oriented presentation can be found in [4].

7.1.1 Bipolar Devices

The principle geometry of a bipolar device is always the same. However, in practice the geometry varies with fabrication process, the intended application and whether the transistor is a discrete transistor or integrated on a chip. A simple structure is shown for a discrete npn BJT in figure 7.1. Typically, emitter and base contact stripes are interdigitated as shown in figure 7.2 to increase the perimeter and thereby increasing the current handling capabilities and reducing parasitic resistors and capacitors.

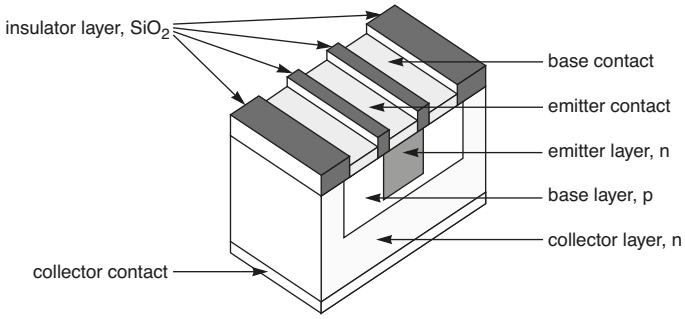


Figure 7.1 Simplified structure of a planar n-p-n discrete BJT.

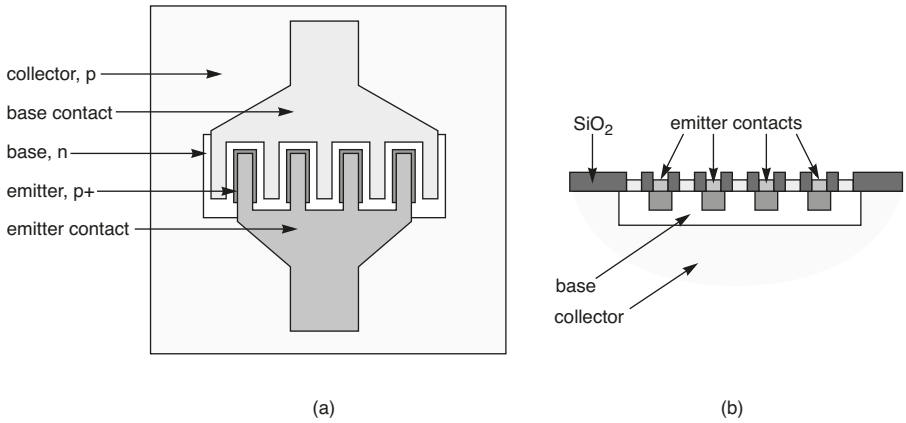


Figure 7.2 (a) Surface layout of transistor with interdigitated emitter and base contacts and (b) cross section.

Heterojunction bipolar transistors are usually referred to as HBTs. Various combinations of materials can be conceived to form such a transistor but it is important that the lattice constants are close. One example is to use n-type germanium and p-type gallium arsenide. The advantage of HBTs over regular BJTs can be expressed in terms of semiconductor physics:

1. Higher emitter efficiency (affecting the current gain) because emitter minority carriers flowing from the base to the emitter are blocked by higher barrier in the valence band.
2. Less base resistance because the base can be heavily doped without sacrificing emitter efficiency.
3. Less emitter current crowding because of low voltage drop along the emitter-base junction. Emitter current crowding has, however, the desirable effect to reduce the base resistance but it can also reduce the emitter efficiency.
4. Improved frequency response because of higher current gain and lower base resistance.

In addition to this, some of the materials that can be used for HBTs have high-temperature capabilities. As an example, a transistor formed by an n-type AlGaAs emitter, a p-type GaAs base and a n-type GaAs collector can be operated above 300°C. The performance of HBTs is comparable with HEMTs described in the next section.

7.1.2 FET Devices

In this chapter we briefly consider four different types of FET devices, the junction field-effect transistor (JFET), the metal-semiconductor field-effect transistor (MESFET), the metal-oxide-semiconductor field-effect transistor (MOSFET) and finally the high electron-mobility transistor (HEMT). They are all based on controlling the cross-sectional area of a channel available for current flow between source and drain with a voltage applied to the gate. The operational principle of the various technologies are described below and the basic structure of each type of FET device is exemplified in figure 7.3. In practice the drain, gate and source are interdigitated similar to the bipolar devices.

The JFET, see figure 7.3a, contains a conductive channel with a direct interface to the gate having opposite doping to form pn-junction with the channel. By applying a gate voltage (backward biased) the depletion layer in the junction can be controlled thereby changing the cross-sectional area of the channel. The JFET was the first FET device that was invented. However, today the JFET is not a competitive technology for higher frequencies.

The MESFET is identical in operation to the JFET but instead of a pn-junction there is a metal-semiconductor rectifying junction, i.e. a schottky barrier, established between the gate contact and the doped channel. The performance of a GaAs MESFET device compared with Si JFET or MOSFET is evident from significantly higher mobility and saturation drift velocity.

The MOSFET structure illustrated in figure 7.3c is an n-channel enhancement mode device where a positive gate voltage attracts negative charge carriers to form a channel between the source and the drain just beneath the gate insulator. Not illustrated here is the depletion mode device where the channel is n-doped to conduct without applying a gate voltage. Instead, a negative gate voltage must be applied to deplete the channel. Despite speed limitations MOSFETs are attractive compared with MESFETs and JFETs in power amplifiers because they exhibit better linearity and has a larger non-destructive gate voltage range since there is no junction that can be forward-biased. The MOSFET has many acronyms including IGFET (insulated field-effect transistor), MISFET (metal-insulator-semiconductor field-effect transistor) and MOST (metal-oxide-semiconductor transistor).

HEMTs, sometimes also referred to as heterojunction MESFET or HFET, is quite different from the other FET devices. One example of a HEMT structure is shown in figure 7.3d where a sequence of GaAs and AlGaAs layers forces an electron gas to form in the interface of the undoped GaAs layer and the undoped AlGaAs layer. The mobility of the electrons in this gas is much higher than for electrons in the conduction band of a doped material of the same type. Also, for lower temperatures the mobility in HEMTs increases dramatically. As is the case with MOSFETs, both enhancement mode and depletion mode HEMTs can be fabricated.

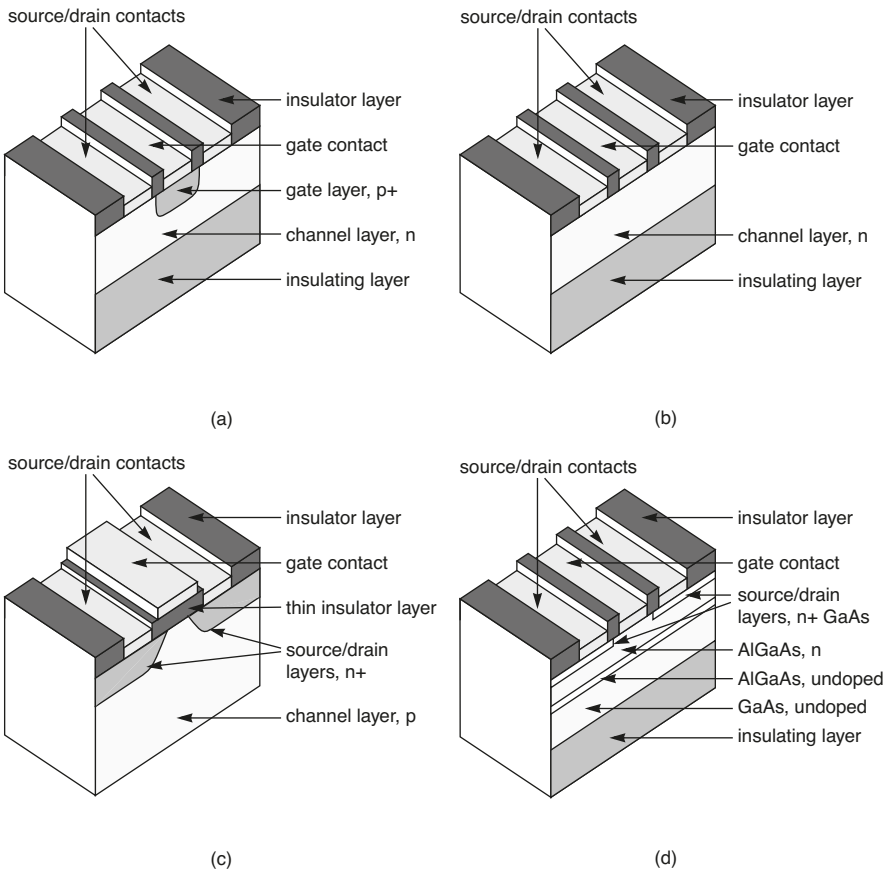


Figure 7.3 Simplified structures
(a) n-channel JFET (b) n-channel MESFET
(c) n-channel MOSFET (d) HEMT

The properties of the various transistor technologies available for RF and microwave applications are summarised and compared in table 7.1.

Table 7.1 Performance comparison between various transistor technologies (derived from [3])

Technology	Speed	Power	Noise
Si BJT	***	**	**
Si MOSFET	*	****	**
GaAs-AlGaAs HBT	*****	***	***
GaAs MESFET	***	***	***
HEMT	*****	****	****

7.2 Transistors in High Frequency Operation

The behaviour of a transistor operating at high frequencies is to large extent determined by capacitive elements in the transistor. These elements determine the maximum frequency of operation for the device. Actually, it is common for radio and microwave transistors to operate at frequencies a tenth of this maximum frequency (f_{max}) or more. This means that the margins are small and care must be taken in the design to avoid waste of transistor performance. The common-emitter (common-source) configuration is mostly used and that configuration will be assumed to be used throughout this section. To understand what limits the transistor performance in terms of frequency we need an intuitive model as a starting point.

In circuit simulation software, e.g. Spice etc., fairly complex models are used to model the intrinsic behaviour of transistors. Up to 60 parameters can be specified and simulation results can be very accurate provided that the parameters are correctly specified. Such a complex model takes into account many aspects of the transistor operation including the nonlinear behaviour, but the model is also difficult to understand. For analysis and hand-calculation purposes we need to model the transistor with as few parameters as possible. The simulation model is far to complex for these purposes. The hybrid-π model is a small-signal model for bipolar transistors that is suitable for our needs. Most of the model element values can be calculated directly or can be found in data sheets. Another alternative is to use a simulator with an accurate model to extract the hybrid-π element values for a given operating point. We can also develop an equivalent model for FET devices that is equally useful. These models will be used below to demonstrate the qualitative behaviour of transistors at high frequencies.

Two figures of merit are commonly used to rate transistors in terms of frequency capability. These are:

1. f_T - the transition or gain-bandwidth frequency. This is the frequency where the short-circuit current gain becomes unity.
2. f_{max} - the maximum frequency of oscillation. This is the frequency where the maximum available power gain (MAG) becomes unity and theoretically indicates the highest frequency at which the device will be able to oscillate.

The transition frequency, f_T, is mostly used to rate transistors for low frequency applications and transistors in digital circuits whereas f_{max} is used to rate radio and microwave transistors. Both these two figure of merits will be investigated below for bipolar and FET devices and how they relate to elements in simplified models for these devices.

7.2.1 Bipolar Transistors

A hybrid-π model of an intrinsic bipolar transistor is shown in figure 7.4. Some elements are more dominating than others, especially for high frequencies. The collector-emitter resistor, r_{ce}, is typically very large and the reactance of the base-collector capacitor, $X_{c_{b'c}}$, is small in comparison to the base-collector resistance. Sometimes transistor models are used that model the distributed nature of the base resistor and the base-collector capacitor.

Such a model splits the capacitor in two parts (or more) as shown in figure 7.4. Omitting the insignificant elements and the distributed capacitor we get the model in figure 7.5.

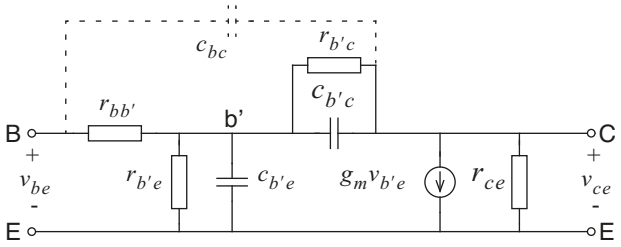


Figure 7.4 Hybrid-π model.

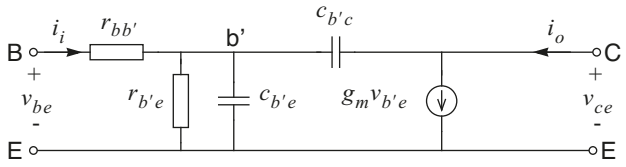


Figure 7.5 Simplified hybrid-π model.

To begin with, consider the frequency dependency of the current gain. Short-circuiting the output of the model in figure 7.5 and omitting the base resistor, $r_{bb'}$ the current gain becomes

$$\beta(\omega) = \frac{i_o}{i_i}(\omega) = \frac{g_m r_{b'e}}{1 + j\omega r_{b'e}(c_{b'e} + c_{b'c})} \tag{7.1}$$

and since $g_m r_{b'e}$ equals the DC current gain β_0 we can write

$$\beta(\omega) = \frac{\beta_0}{1 + j\omega\beta_0 \dfrac{(c_{b'e} + c_{b'c})}{g_m}} \approx \frac{g_m}{j\omega(c_{b'e} + c_{b'c})} \tag{7.2}$$

The last approximation is only valid for high frequencies. To obtain the transition frequency we solve (7.2) for $|\beta(\omega)| = 1$ which gives

$$f_T = \frac{\omega_T}{2\pi} = \frac{1}{2\pi} \cdot \frac{g_m}{c_{b'e} + c_{b'c}} \tag{7.3}$$

The transconductance, g_m, and the base-emitter capacitance, $c_{b'e}$, are proportional to the collector current. However, for small collector currents the base-collector capacitance, $c_{b'c}$, dominates over $c_{b'e}$, which means that f_T will increase with increasing current up to the point where $c_{b'e}$ is of similar

size as $c_{b'c}$. For even higher currents f_T decreases again. This effect is however not covered by this simple theory. The reduction of f_T is due to high-level injection and the Kirk effect, which out of the scope to treat in this text.

The conditions for deriving f_{max} is different from that of f_T. For f_{max} we must have conjugate matching both at the input and the output in contrast to the short-circuit output that was used for deriving the transition frequency. It can be shown that the maximum frequency of oscillation is given by

$$f_{max} = \sqrt{\frac{f_T}{8\pi r_{b'b} c_{b'c}}} = f_T \cdot \sqrt{\frac{c_{b'e} + c_{b'c}}{4 r_{bb'} c_{b'c} g_m}} \tag{7.4}$$

From this expression we see that under certain circumstances the maximum frequency of oscillation, f_{max} can be larger than the transition frequency. Even though the current gain is unity or less the transistor can still be capable of power amplification if the input impedance is smaller than the output impedance of the device.

The transition frequency can also be expressed in terms of the charge carrier transit time, τ_c, such that

$$f_T = \frac{1}{2\pi\tau_c} \tag{7.5}$$

The transit time in turn is composed of several terms but it is dominated by the transit time through the base and the collector depletion layer. Therefore, to obtain a high transition frequency the thickness of the base and the collector depletion layer should be small. The former is possible with state-of-the-art bipolar process technology that improves continuously. The latter is a matter of doping the collector and the base properly.

However, there are some fundamental limits that will also limit the performance of transistors no matter how small geometries that can be fabricated. One limitation is the that there is a maximum possible velocity of charge carriers in a semiconductor, the saturation drift velocity v_s, and is on the order of 6×10^6 cm/s for both electrons and holes in silicon and in germanium. There is also a maximum electric field that can be sustained in a semiconductor without having dielectric breakdown. For silicon this field is about 2×10^5 V/cm and 1×10^6 V/cm for germanium.

The transit time can be expressed as

$$\tau_c = \frac{L}{v} \tag{7.6}$$

where L represents the effective emitter-collector distance and v the charge carrier velocity. If the carriers are moving at the saturation drift velocity, v_s, the transit time can only be reduced by reducing L. But there is a lower limit on L since the electric field will increase accordingly and finally cause dielectric breakdown. This in turn can be circumvented by reducing the voltage applied to the device but this also reduces the maximum attainable power.

It should be noted that the maximum obtainable frequency that can be estimated from these simple relations are too optimistic. In practice the velocity and the electric field intensity will not be uniform in the transistor structure which will reduce the attainable frequency quite substantially. A more detailed treatment of physical limitations on the performance of transistors is found in [5].

Example 7.1 Performance of Silicon BJT

Calculate the transition frequency and maximum frequency of oscillation for a silicon BJT with $r_{bb'} = 10\Omega$, $c_{b'e} = 10\text{pF}$, $c_{b'c} = 0.35\text{pF}$ $\beta_{DC} = 120$ and $I_C = 20\text{mA}$.

For the transition frequency we have

$$f_T = \frac{1}{2\pi} \cdot \frac{g_m}{c_{b'e} + c_{b'c}}$$

Here the transconductance is given by $g_m = qI_C/kT = 0.773$ S and we get

$$f_T = \frac{1}{2\pi} \cdot \frac{0.773}{10\times10^{-12} + 0.35\times10^{-12}} = 11.9\text{GHz}$$

For the maximum frequency of oscillation we have

$$f_{max} = f_T \cdot \sqrt{\frac{c_{b'e} + c_{b'c}}{4r_{bb'}c_{b'c}g_m}} = 11.6\text{GHz}$$

7.2.2 FET Devices

A small signal equivalent model for the intrinsic part of FET devices is shown in figure 7.6. This model is accurate enough to describe the qualitative behaviour of all FET devices. The gate-source resistor r_i represents a small resistor from the gate through the channel to source. For a real device the actual resistance of the gate structure must also be included since it may be in same range as the intrinsic part.

In many cases the feedback capacitor c_{gd} can be disregarded in a coarse estimation of performance. In this case f_T and f_{max} are easily derived from the model schematic and it is found that

$$f_T = \frac{g_m}{2\pi c_{gs}} \text{ and} \tag{7.7}$$

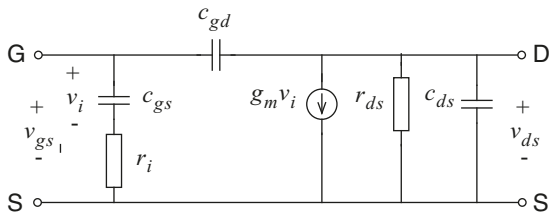


Figure 7.6 Small signal equivalent model of FET device.

$$f_{max} = \frac{f_T}{2} \cdot \sqrt{\frac{r_{ds}}{r_i}} \tag{7.8}$$

The transition frequency can also be expressed in terms of the charge carrier transit time (τ_c) through the channel (similar to the BJT):

$$f_T = \frac{1}{2\pi\tau_c} \tag{7.9}$$

Given the gate length, L, we can associate the charge carriers with a velocity, v, such that

$$\tau_c = \frac{L}{v} \tag{7.10}$$

Due to the small geometries of high performance transistors very large electric fields are present in the channel forcing the charge carriers into velocity saturation. For devices with small structures we can therefore replace the velocity, v, with the charge carrier saturation drift velocity, v_s, and the transition frequency then becomes

$$f_T = \frac{v_s}{2\pi L} \tag{7.11}$$

Example 7.2 Performance of GaAs MESFET

Calculate the transition frequency and maximum frequency of oscillation for an GaAs MESFET device in room temperature. The effective gate length L is 0.50μm. The total gate-source resistance equals 7Ω and the drain-source resistance equal 350Ω.

The drift velocity for electrons in GaAs has a peak at 3×10^3 V/cm and equals 2×10^7 cm/s [2]. For higher electric fields the saturation drift velocity converge towards 5×10^6 cm/s. Assuming that the device is operated with the latter figure we get

$$f_T = \frac{v_s}{2\pi L} = \frac{5\times10^4\,\text{m/s}}{2\pi\cdot 0.5\times10^{-6}\,\text{m}} = 15.9\,\text{GHz}$$

and

$$f_{max} = \frac{f_T}{2}\cdot\sqrt{\frac{r_{ds}}{r_i}} = \frac{15.9\,\text{GHz}}{2}\cdot\sqrt{\frac{350}{7}} = 56.2\,\text{GHz}$$

7.3 Package Modelling

The simple hybrid-π model and the equivalents for field-effect transistors give satisfactory accuracy for discrete realisation only when designing circuits for low frequencies. There is no well-defined limit but problems can be expected at a few hundred MHz or less. This is due to the influence of the pins and the package that houses the transistor chip.

The transistor chip is encapsulated in a package and connected through bonding wires to the external pins. Bonding wires are modelled as inductors sometimes in series with small resistors. Bonding wires that are close to each other and the pins that they are connected to exhibit more or less inductive and capacitive coupling.

The equivalent circuit for the surface mount plastic package SOT23 housing the BFR520 transistor is shown in figure 7.7 as well the physical dimensions of the package. The resistors in series with the inductors are fairly small in this case and can be represented by a Q value for each inductor. In this example the Q value is about 50 at 1GHz.

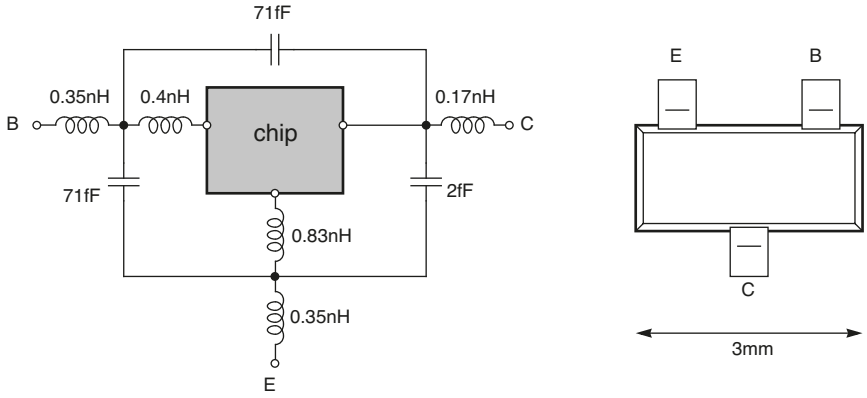


Figure 7.7 Package equivalent circuit for SOT23 and physical dimensions.

As is obvious from figure 7.7 the complexity of the transistor increases substantially when the package is taken into account. As will be described in the next chapter a two-port representation can be used for the complex circuit modelling a discrete transistor. Instead of specifying all circuit elements both

within the intrinsic device and the package, four complex-valued parameters are specified for each combination of frequency and operating point of interest. Examples of such parameters are S parameters and y parameters. Measured S or y parameters are usually provided by the manufacturers of high frequency transistors. In some cases small-signal equivalent model parameters are also given. These are however typically best-fit values that have been fitted to the measured S-parameters.

7.4 References

[1] B. G. Streetman, *Solid state electronic devices*, 3rd edition, Prentice Hall, 1990.

[2] S. M. Sze, *Semiconductor devices - physics and technology*, John Wiley & Sons, 1985.

[3] S. Y. Liao, *Microwave devices and circuits*, 3rd edition, Prentice Hall, 1990.

[4] P. G. Gray and R. G. Meyer, *Analysis and design of analog integrated circuits*, 3rd edition, John Wiley & Sons, 1993.

[5] E. O. Johnson, "Physical limitations on frequency and power parameters of transistors", *RCA Rev.*, vol. 26, no. 6, pp. 163-177, juni 1965.

Chapter 8

Amplifier Design Using Two-Port Network Representation

This chapter deals with the design of high frequency amplifiers for small signal applications where the amplifying devices, the transistors, are characterised by two-port networks. Such a two-port is a rather abstract representation given by a set of a few parameters that have no direct connection with the physical behaviour of the device. One parameter set is valid for one fixed frequency and one fixed operating point only.

The design methods presented in this chapter differ substantially from the methods within the discipline of traditional analog circuit design. The latter relies on using models like the hybrid-π model for bipolar transistors. Such a model consists of circuit elements that are related to the physical behaviour of the transistor. As long as these models can be kept simple by disregarding insignificant elements in the models, amplifiers can be designed using powerful methods that provides an overall picture of the amplifier behaviour. With such a method we do not necessarily design for a specific frequency, instead we can control the complete frequency response through the manipulation of poles and zeros that are related to the component values in the circuit. For discrete realisation these methods are applicable below a few hundred MHz, sometimes much less depending on the type of components used and the accuracy that is required.

In the case of higher frequencies the behaviour of a transistor must be modelled with more elements including the package. Most of these element are reactive elements, i.e., capacitors and inductors, that make the traditional methods difficult or impossible to use because of the increased complexity. This is where the two-port representation manifests itself, a black-box representation with a set of parameters for each frequency and operating point. Since a complex circuit is compressed into a model with a small number of parameters for a fixed operating condition, design methods based on such a model do no provide the insight that for example pole-zero analysis and the hybrid-π model provide.

The chapter contains two parts. The first serves as an introduction to two-port representations where y, z and ABCD parameters and their applications are introduced. From this perspective power gain and stability concepts are defined. This forms the basis for second part that deals with S parameters and design of amplifiers using S parameters.

8.1 Y, Z and ABCD Parameters

To be able to define the different two-port representations consistently we need to define the two-port. Consider the diagram in figure 8.1.

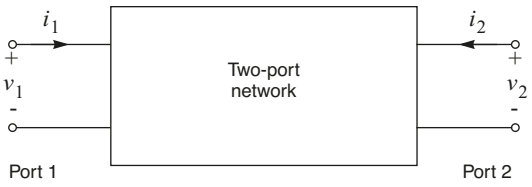


Figure 8.1 Two-port network representation.

Each port is associated with a current and a voltage with directions as indicated. If the two-port is linear it can be described using four complex-valued elements for one frequency and one operating point. That is, four elements are required to relate the two current and the two voltages to each other. The admittance parameters are probably the most intuitive because the equivalent schematic reminds of a stripped-down hybrid-π model. The schematic is shown in figure 8.2 and the corresponding equations in matrix form is given by

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{8.1}$$

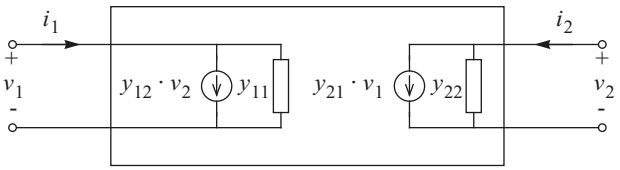


Figure 8.2 Two-port schematic based on y parameters.

In addition to the admittance (y) parameters the most commonly used parameters (except for S parameters that will be treated separately in this chapter) are the impedance (z) parameters and the chain or ABCD parameters. They are defined as follows:

z parameters:

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{8.2}$$

ABCD parameters:

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} v_2 \\ -i_2 \end{bmatrix} \tag{8.3}$$

An equivalent schematic for the z parameters is illustrated in figure 8.3.

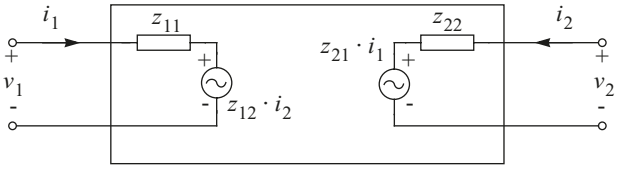


Figure 8.3 Two-port schematic based on z parameters.

For completeness we also show one possible equivalent schematic for the ABCD parameters in figure 8.4. Although, the ABCD parameters are probably better understood by the mathematical definition (8.3).

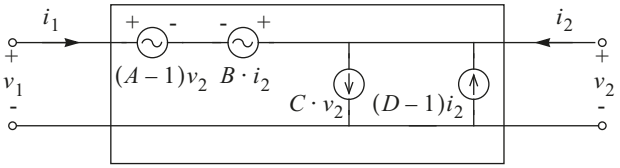


Figure 8.4 Two-port schematic based on ABCD parameters.

The y, z and ABCD parameters are attractive because they are simple to measure at low frequencies. Consider for example y_{11} in (8.1). By applying a signal to the input port and setting $v_2 = 0$, i.e., AC short-circuiting the output of the device being measured, we get

$$y_{11} = \left.\frac{i_1}{v_1}\right|_{v_2 = 0} \tag{8.4}$$

All y, z and ABCD parameters can be measured in this way with short- or open-circuit terminations. It goes without saying that we can transform one set of parameters to another set of parameters, refer to tables 8.2 and 8.3.

If two two-ports are connected to form a new two-port, parameters for the new two-port is easily calculated from the individual parameter sets if the correct type of parameters is chosen. To begin with, again consider the y parameters and corresponding schematic diagrams for two two-ports in shunt connection in figure 8.5.

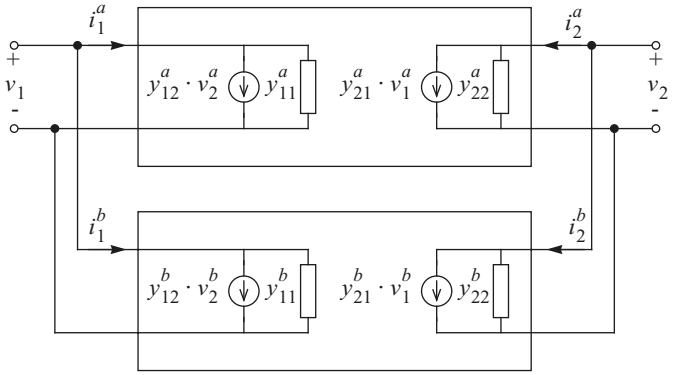


Figure 8.5 Shunt connection using y parameters.

From the schematic it is obvious that the new two-port is represented by the sum of the matrices of the individual two-ports when the two-ports are shunt connected, i.e.,

$$\begin{bmatrix} i_1 \\ i_2 \end{bmatrix} = \begin{bmatrix} i_1^a + i_1^b \\ i_2^a + i_2^b \end{bmatrix} = \begin{bmatrix} y_{11}^a + y_{11}^b & y_{12}^a + y_{12}^b \\ y_{21}^a + y_{21}^b & y_{22}^a + y_{22}^b \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \tag{8.5}$$

Similarly, if two two-ports are connected in series as shown in figure 8.6 the z parameters are the most appropriate and again the new matrix is given by the sum of the matrices of the individual two-ports, i.e.,

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} v_1^a + v_1^b \\ v_2^a + v_2^b \end{bmatrix} = \begin{bmatrix} z_{11}^a + z_{11}^b & z_{12}^a + z_{12}^b \\ z_{21}^a + z_{21}^b & z_{22}^a + z_{22}^b \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \end{bmatrix} \tag{8.6}$$

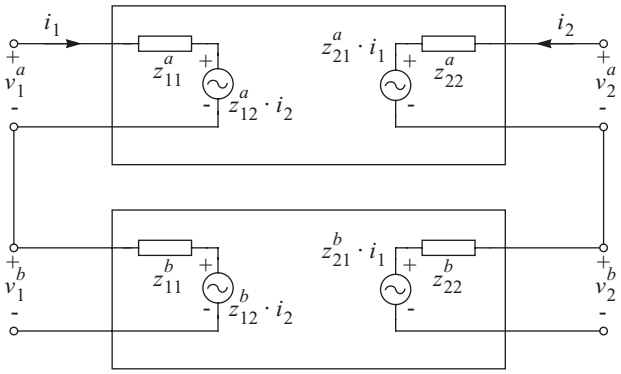


Figure 8.6 Series connection using z parameters.

Another important case is cascading of two-ports as in figure 8.7, and this is where ABCD parameters becomes handy.

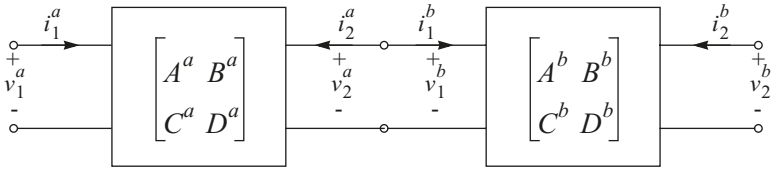


Figure 8.7 Cascade connection using ABCD parameters.

The new parameters are given by matrix multiplication, i.e.,

$$\begin{bmatrix} v_1 \\ i_1 \end{bmatrix} = \begin{bmatrix} v_1^a \\ i_1^a \end{bmatrix} = \begin{bmatrix} A^a & B^a \\ C^a & D^a \end{bmatrix} \begin{bmatrix} v_2^a \\ -i_2^a \end{bmatrix} = \begin{bmatrix} A^a & B^a \\ C^a & D^a \end{bmatrix} \begin{bmatrix} A^b & B^b \\ C^b & D^b \end{bmatrix} \begin{bmatrix} v_2^b \\ -i_2^b \end{bmatrix} \tag{8.7}$$

It was demonstrated how the parameters can be measured by simply AC short or open-circuit the input or the output of a device. However, well-defined open-circuit or short-circuit tests are very difficult to achieve at high frequencies, especially over a broad band. Also, it is very common that radio frequency transistors oscillates due to internal feedback if any of the two ports are short or open-circuited. This of course makes the measurement impossible. Instead, for measuring purposes at high frequencies, S parameters are preferred because they rely on connecting a reasonable resistive load to both the input and output when measuring a device. Please note that this fact does not imply that the parameters described earlier are useless for higher frequencies. They are still valid for design and analysis no matter what the frequency is. It is perfectly alright to measure a device with a network analyser to obtain S parameters and convert these parameters to y parameters for example, if this is more convenient for the designer.

Parameter sets are typically given for transistors in common-emitter (or common-source) configuration. If we have a parameter set for a transistor in one of the other configurations, common-base (common-gate) or common-collector (common-drain), we can convert these parameters to correspond to another configuration. Formulas are found in table 8.1.

Table 8.1
Conversion between y parameters for various transistor configurations.
CE - common-emitter CB - common-base CC - common-collector

	from CE	from CB	from CC
to CE	y_{11e} y_{12e} y_{21e} y_{22e}	$y_{11b}+y_{12b}+y_{21b}+y_{22b}$ $-(y_{12b}+y_{22b})$ $-(y_{21b}+y_{22b})$ y_{22b}	y_{11c} $-(y_{11c}+y_{12c})$ $-(y_{11c}+y_{21c})$ $y_{11c}+y_{12c}+y_{21c}+y_{22c}$
to CB	$y_{11e}+y_{12e}+y_{21e}+y_{22e}$ $-(y_{12e}+y_{22e})$ $-(y_{21e}+y_{22e})$ y_{22e}	y_{11b} y_{12b} y_{21b} y_{22b}	y_{22c} $-(y_{21c}+y_{22c})$ $-(y_{12c}+y_{22c})$ $y_{11c}+y_{12c}+y_{21c}+y_{22c}$
to CC	y_{11e} $-(y_{11e}+y_{12e})$ $-(y_{11e}+y_{21e})$ $y_{11e}+y_{12e}+y_{21e}+y_{22e}$	$y_{11b}+y_{12b}+y_{21b}+y_{22b}$ $-(y_{11b}+y_{21b})$ $-(y_{11b}+y_{12b})$ y_{11b}	y_{11c} y_{12c} y_{21c} y_{22c}

Table 8.2
Conversion from S and z parameters to S, z, y, and ABCD parameters.

	S	z
S	S_{11}	$S_{11} = \frac{(z_{11n}-1)(z_{22n}+1)-z_{12n}z_{21n}}{\Delta_1}$
	S_{12}	$S_{12} = \frac{2z_{12n}}{\Delta_1}$
	S_{21}	$S_{21} = \frac{2z_{21n}}{\Delta_1}$
	S_{22}	$S_{22} = \frac{(z_{11n}+1)(z_{22n}-1)-z_{12n}z_{21n}}{\Delta_1}$
z	$z_{11n} = \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{\Delta_5}$	z_{11}
	$z_{12n} = \frac{2S_{12}}{\Delta_5}$	z_{12}
	$z_{21n} = \frac{2S_{21}}{\Delta_5}$	z_{21}
	$z_{22n} = \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{\Delta_5}$	z_{22}
y	$y_{11n} = \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{\Delta_6}$	$y_{11} = \frac{z_{22}}{\lvert z\rvert}$
	$y_{12n} = \frac{-2S_{12}}{\Delta_6}$	$y_{12} = \frac{-z_{12}}{\lvert z\rvert}$
	$y_{21n} = \frac{-2S_{21}}{\Delta_6}$	$y_{21} = \frac{-z_{21}}{\lvert z\rvert}$
	$y_{22n} = \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{\Delta_6}$	$y_{22} = \frac{z_{11}}{\lvert z\rvert}$
A B C D	$A_n = \frac{(1+S_{11})(1-S_{22})+S_{12}S_{21}}{2S_{21}}$ $B_n = \frac{(1+S_{11})(1+S_{22})-S_{12}S_{21}}{2S_{21}}$ $C_n = \frac{(1-S_{11})(1-S_{22})-S_{12}S_{21}}{2S_{21}}$ $D_n = \frac{(1-S_{11})(1+S_{22})+S_{12}S_{21}}{2S_{21}}$	$A = \frac{z_{11}}{z_{21}}$ $B = \frac{\lvert z\rvert}{z_{21}}$ $C = \frac{1}{z_{21}}$ $D = \frac{z_{22}}{z_{21}}$

$\Delta_1 = (z_{11n}+1)(z_{22n}+1)-z_{12n}z_{21n}$
$\Delta_5 = (1-S_{11})(1-S_{22})-S_{12}S_{21}$
$\Delta_6 = (1+S_{11})(1+S_{22})-S_{12}S_{21}$
$z_{ijn} = z_{ij}/Z_0 \qquad y_{ijn} = y_{ij}Z_0$
$A_n = A \qquad B_n = B/Z_0 \qquad C_n = CZ_0 \qquad D_n = D$
$\lvert z\rvert = z_{11}z_{22}-z_{12}z_{21}$

Table 8.3
Conversion from y and ABCD parameters to S, z, y, and ABCD parameters.

	y	ABCD
S	$S_{11} = \frac{(1-y_{11n})(1+y_{22n})+y_{12n}y_{21n}}{\Delta_2}$ $S_{12} = \frac{-2y_{12n}}{\Delta_2}$ $S_{21} = \frac{-2y_{21n}}{\Delta_2}$ $S_{22} = \frac{(1+y_{11n})(1-y_{22n})+y_{12n}y_{21n}}{\Delta_2}$	$S_{11} = \frac{A_n+B_n-C_n-D_n}{\Delta_4}$ $S_{12} = \frac{2(A_nD_n-B_nC_n)}{\Delta_4}$ $S_{21} = \frac{2}{\Delta_4}$ $S_{22} = \frac{-A_n+B_n+D_n-C_n}{\Delta_4}$
z	$z_{11} = \frac{y_{22}}{\lvert y\rvert}$ $z_{12} = \frac{-y_{12}}{\lvert y\rvert}$ $z_{21} = \frac{-y_{21}}{\lvert y\rvert}$ $z_{22} = \frac{y_{11}}{\lvert y\rvert}$	$z_{11} = \frac{A}{C}$ $z_{12} = \frac{\Delta_8}{C}$ $z_{21} = \frac{1}{C}$ $z_{22} = \frac{D}{C}$
y	y_{11} y_{12} y_{21} y_{22}	$y_{11} = \frac{D}{B}$ $y_{12} = \frac{-\Delta_8}{B}$ $y_{21} = \frac{-1}{B}$ $y_{22} = \frac{A}{B}$
A B C D	$A = \frac{-y_{22}}{y_{21}}$ $B = \frac{-1}{y_{21}}$ $C = \frac{-\lvert y\rvert}{y_{21}}$ $D = \frac{-y_{11}}{y_{21}}$	A B C D

$\Delta_2 = (1+y_{11n})(1+y_{22n}) - y_{12n}y_{21n}$
$\Delta_4 = A_n + B_n + C_n + D_n$
$\Delta_8 = AD - BC$
$z_{ijn} = z_{ij}/Z_0 \qquad y_{ijn} = y_{ij}Z_0$
$A_n = A \qquad B_n = B/Z_0 \qquad C_n = CZ_0 \qquad D_n = D$
$\lvert y\rvert = y_{11}y_{22} - y_{12}y_{21}$

8.2 Stability

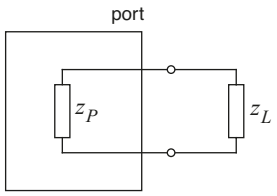
An amplifier needs an input signal to produce an output signal. This is not the case for an oscillator. The amplifier has been designed for stable operation. The oscillator has been deliberately designed to be unstable. The cause of instability is positive feedback within the circuit. Therefore, feedback is a problem in amplifier design and a prerequisite in oscillator design.

If a circuit is designed by considering a transfer function in terms of poles and zeros we know that the poles must be located in the left-half plane for stable operation. Other measures for stability analysis that are closely related to pole-zero patterns are amplitude and phase margin that can be found from a Bode plot or a Nyquist plot. Basically, poles and zeros give the designer an overall picture of the frequency response. This is not case for a two-port represented by a parameter set. Here, the condition for stability must be defined differently because we only consider one frequency at the time.

8.2.1 Negative Resistance

Stability properties of a two-port can be investigated by looking at the port impedances. If the resistive part of a port impedance is negative the two-port will be able to oscillate, it all depends on how the port is loaded. A negative resistance may not be a property that is easy to accept or understand. The concept of travelling waves provides another way of looking at it because we can think of a one-port that has a reflection coefficient that is larger than unity which actually corresponds to a negative port resistance. This means that an incident wave is reflected by the port with more power than the incident wave.

For oscillation to occur, a negative resistance is not sufficient. The circuit must include reactive elements that must be in resonance at the desired frequency of oscillation. A general treatment of this topic can be found in the chapter on oscillators. In this special case where we consider a port with a negative resistance the criteria for oscillation can be formulated using the loop impedance, see figure 8.8. Oscillation will occur if the loop resistance is negative and the loop reactance is zero, i.e., the load resistance must not be larger than the absolute value of the port resistance and the load reactance must cancel the port reactance.



$$r_{loop} = \mathrm{Re}[z_L] + \mathrm{Re}[z_P]$$
$$x_{loop} = \mathrm{Im}[z_L] + \mathrm{Im}[z_P]$$

Figure 8.8 Loop impedance for port with load.

In the case of amplifier design it is advisable to have a fair margin to avoid oscillation. If possible, negative port resistances should be avoided completely, see section 8.2.3. However, if necessary this rule can be stretched but care must be taken to include effects of component tolerances, aging and temperature when the margins become smaller.

8.2.2 Port Impedances

Since the port impedances need to be investigated to determine the stability properties of a two-port we need expressions for the port impedances as functions of the two-port parameters as well as the source and load impedances.

In general the source and load ports affects each other. The output impedance depends on the source impedance and the input impedance depends on the load impedance. Using y parameters, it is easy to show that the input and the output admittances are given by:

$$y_{in} = y_{11} - \frac{y_{12} \cdot y_{21}}{y_{22} + y_L} \tag{8.8}$$

$$y_{out} = y_{22} - \frac{y_{12} \cdot y_{21}}{y_{11} + y_S} \tag{8.9}$$

where y_S is the source admittance and y_L the load admittance.

8.2.3 Unconditional Stability

A two-port is defined to be unconditionally stable if no source or load impedance results in a negative output resistance or negative input resistance, respectively. Here, we assume passive impedances that by definition have a positive resistive part. Note that if this condition is not fulfilled it does not necessarily mean that the amplifier will oscillate. Instead, if for example the input impedance is negative for some load impedance we must investigate the loop impedance before we can draw any conclusion on the behaviour of the amplifier. However, a negative port resistance is undesirable and is usually avoided.

To formulate the unconditional stability with equations we need to investigate the input and output impedances, see (8.8) and (8.9). We can apply boundary conditions for these equations so that the input, output, load and source resistances must be larger than zero. From this analysis we can derive an explicit expression for the stability referred to as a stability factor. One commonly used stability factor is the Linvill stability factor,

$$C_L = \frac{|y_{12} \cdot y_{21}|}{2 \cdot g_{11} \cdot g_{22} - \mathrm{Re}[y_{12} \cdot y_{21}]} \tag{8.10}$$

If $0 < C_L < 1$ the two-port is unconditionally stable. g_{11} and g_{22} denote the conductance part of the corresponding admittance, i.e., $g_{11} = \mathrm{Re}[y_{11}]$ and $g_{22} = \mathrm{Re}[y_{22}]$. Another well known stability factor is the Stern's stability factor,

$$K_S = \frac{2(g_{11} + g_{input})(g_{22} + g_{output})}{|y_{12} \cdot y_{21}| + \mathrm{Re}[y_{12} \cdot y_{21}]} \tag{8.11}$$

If $K_S > 1$ the two-port is unconditionally stable. Stern's stability factor differs from the Linvill stability factor in that it takes into account additional resistances (g_{input} and g_{output}) that are shunt-connected at the input and at the output. This is discussed in the next section.

8.2.4 Conditional Stability

Even though a two-port is not unconditionally stable it can be forced to stable operation with a proper selection of source and load impedances. Such a two-port is referred to as a conditionally stable two-port. However, the most useful methodologies that are available for this case are based on S parameters rather than y parameters, see next chapter. A few approaches that are not based on S parameters are mentioned below.

The easiest way to make a conditionally stable two-port unconditionally stable is by resistive shunt-loading at the input or at the output or even both. The effect of such loading is most easily investigated using Stern's stability factor since it provides a way to include source and load resistors. By proper selection of these we can force $K_S > 1$ and thereby obtain an unconditionally stable two-port. The disadvantage with this technique is that the gain is reduced because power is dissipated in the resistors.

Neutralisation is another way of making a two-port more stable. The objective is to reduce or even neutralise the effect of the feedback capacitor between the input and the output port, e.g. between the base and the collector in a bipolar device. The principle is that an inductor is connected in parallel with the capacitor. The value of the inductor is chosen in such a way that the resonance frequency of the LC circuit becomes equal to the operating frequency and thereby the feedback path will be neutralised.

8.3 Power Gain

Power is a more versatile unit than voltage and current when we have the option to include inductors, capacitors and transmission lines in the circuit. These circuit elements have one important contribution to the circuit performance. We can preserve the power that is available from a source and transfer it to the load without significant loss. This is achieved with conjugate matching. This is exactly what is needed when an amplifier has to be designed to operate close to f_{max}. For obvious reasons power gain is preferred over voltage or current gain under these circumstances. Below, three different power gain definitions are introduced that will be used extensively from now on.

8.3.1 Power Gain Definitions

Consider the two-port with source and load connected to it in figure 8.9. If we would like to calculate the power gain of this circuit it is justifiable to question what we mean by power gain.

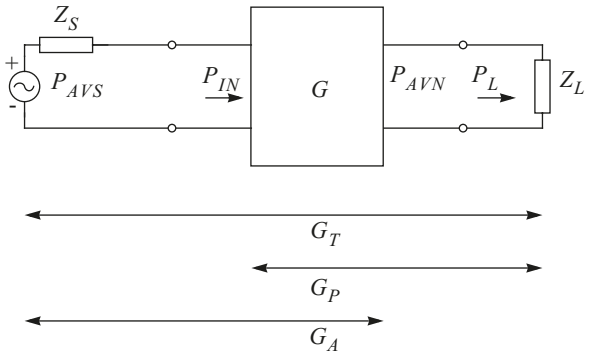


Figure 8.9 Power gain definitions and related power quantities in a two-port.

We know that somehow the power relation between the input and the output should be found. A closer look reveals that we can define several types of power gains for different purposes. One definition is

$$G_P = \frac{P_L}{P_{IN}} \tag{8.12}$$

where P_L is the power delivered to the load and P_{IN} the power delivered to the input of the two-port. G_P is called operating gain. It is readily understood that this definition is independent of the source impedance - we only consider how much power that reaches the input of the two-port but not what forces this power into the two-port. On the output side we consider how much power that is actually delivered to the load and for this reason G_P is dependent on the load impedance.

If the source impedance is given we know that maximum power is transferred to the load (in this case the input of the two-port) if the source and the load is conjugately matched. The power that is transferred in this case is referred to as available power from source, P_{AVS}. We can use this property to introduce a second power gain definition,

$$G_T = \frac{P_L}{P_{AVS}} \tag{8.13}$$

G_T is called transducer gain. Since the available power from the source is required G_T becomes dependent on the source impedance.

These two power gain definitions were obtained by deciding how much of the circuit at the input side of the two-port that should be part of the gain definition (see figure 8.9). We can do the same on the output side by defining avail-

able power from network (the output of two-port), P_{AVN}. This property results in two additional power gain definitions, but only one is relevant and it is

$$G_A = \frac{P_{AVN}}{P_{AVS}} \tag{8.14}$$

G_A is called available gain. Note that this definition is not dependent on the load.

All three power gain definitions are commonly used for different purposes. This will be evident later in this chapter. However, figure 8.10 will be used to demonstrate one simple case where two different gain definitions are required.

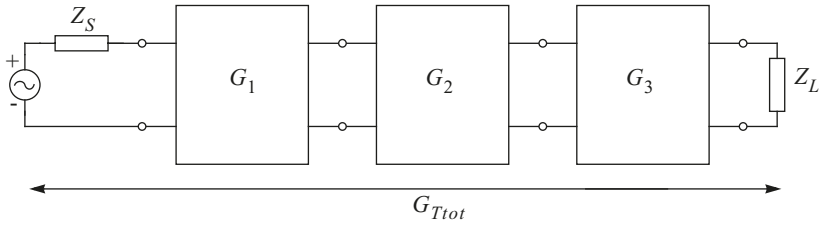


Figure 8.10 Cascade of amplifiers where two different gain definitions are required to calculate the total gain.

A two-port consists of three cascaded two-ports with source and load connected at the ends. The question is how the gain of each individual two-port should be specified to be able to calculate the total gain, lets say the transducer gain, G_{Ttot}, as indicated in figure 8.10, as a product of the individual power gains. There are three solutions and we pick one as an example. The total gain equals

$$G_{Ttot} = \frac{P_L}{P_{AVS}}$$

This expression can be extended and arranged as follows

$$G_{Ttot} = \frac{P_L}{P_{AVS}} = \frac{P_L}{P_{IN3}} \cdot \frac{P_{OUT2}}{P_{IN2}} \cdot \frac{P_{OUT1}}{P_{AVS}} = G_{P3} \cdot G_{P2} \cdot G_{T1}$$

where P_{OUT1} and P_{OUT2} refer to the actual power delivered by stage 1 and 2 respectively to stage 2 and three respectively, i.e., $P_{OUT1} = P_{IN2}$ and $P_{OUT2} = P_{IN3}$. That is, the total gain can be written as the product of the operating gain of stage 2 and 3 and transducer gain of stage 1. The other solutions are left as an exercise for the reader.

8.3.2 Gain with Conjugate-Matched Input and Output

Maximum gain is obtained with conjugate-matched input and output. We make a distinction between two cases. The first case corresponds to that y_{12} equals zero, i.e., there is no feedback from the output to the input. A two-port with this property is said to be unilateral. In practice, y_{12} is never exactly zero but if y_{12} is small enough a simplified design approach is applicable. The case is illustrated in figure 8.11.

If the two-port is fixed the source and the load must satisfy $y_S = y_{IN}^*$ and $y_L = y_{OUT}^*$ respectively. Of course, the source and the load impedance do usually not satisfy these equalities. Matching network must be placed in between. The unilateral case is simple because we have that the input admittance of the two-port equals y_{11} and the output admittance equals y_{22}, i.e., the input and the output can be matched separately.

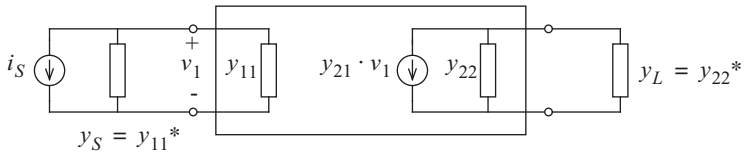


Figure 8.11 Unilateral two-port with conjugate matched source and load.

If a two-port cannot be considered as unilateral, i.e. it is bilateral, the mathematical treatment of the conjugate-match case becomes much more difficult because of the coupling between the input and the output. The input impedance will depend on the load connected to the output and the output impedance will depend on the source connected to the input, see (8.8) and (8.9). This case is referred to as simultaneous conjugate match. To obtain simultaneous conjugate match the source impedance is given by

$$y_S = \frac{1}{2g_{22}}\sqrt{(2g_{11}g_{22} - \mathrm{Re}[y_{12}y_{21}])^2 - |y_{12}y_{21}|^2} + j\left(\frac{\mathrm{Im}[y_{12}y_{21}]}{2g_{22}} - b_{11}\right) \tag{8.15}$$

and the load impedance

$$y_L = \frac{1}{2g_{11}}\sqrt{(2g_{11}g_{22} - \mathrm{Re}[y_{12}y_{21}])^2 - |y_{12}y_{21}|^2} + j\left(\frac{\mathrm{Im}[y_{12}y_{21}]}{2g_{11}} - b_{22}\right) \tag{8.16}$$

Here g_{ii} and b_{ii} denote the conductance and the susceptance part of the corresponding admittance, i.e., $g_{ii} = \mathrm{Re}[y_{ii}]$ and $b_{ii} = \mathrm{Im}[y_{ii}]$.

For this specific case maximum gain is obtained and is given by

$$G_{max} = \frac{2|y_{21}|^2}{\left(\sqrt{2g_{11}g_{22} - \mathrm{Re}[y_{12}y_{21}] + |y_{12}y_{21}|} + \sqrt{2g_{11}g_{22} - \mathrm{Re}[y_{12}y_{21}] - |y_{12}y_{21}|}\right)^2} \tag{8.17}$$

Typically, amplifiers are designed for maximum power gain if possible. However, it can be shown that this is only possible when the two-port is unconditionally stable.

It is readily understood that all power gain definition become equal when we have conjugate match on both the input side and the output side because the delivered power equals the available power at both ports, i.e.,

$$G_T = \frac{P_L}{P_{AVS}} = G_P = \frac{P_L}{P_{IN}} = G_A = \frac{P_{AVN}}{P_{AVS}}$$

8.3.3 Gain with Arbitrary Source and Load

When a two-port is not unconditionally stable we cannot design for maximum gain as discussed in the previous section. In some cases you might design for a specific gain rather than maximum gain. Both cases correspond to mismatch at the input or at the output or both. However, we will not present any design methodologies based on y parameters for this purpose here. Instead, we only present the expressions for the various gain definitions for arbitrary source and load impedances and wait with the more comprehensive design methodologies until the next chapter. Thus, the operating gain can be expressed as

$$G_P = \frac{g_L|y_{21}|^2}{|y_L + y_{22}|^2 \mathrm{Re}\left[y_{11} - \frac{y_{12}y_{21}}{y_L + y_{22}}\right]} \tag{8.18}$$

and the transducer power gain is given by

$$G_T = \frac{4g_S g_L|y_{21}|^2}{|(y_S + y_{11})(y_L + y_{22}) - y_{12}y_{21}|^2} \tag{8.19}$$

and finally the available power gain is given by

$$G_A = \frac{g_S|y_{21}|^2}{|y_S + y_{11}|^2 \mathrm{Re}\left[y_{22} - \frac{y_{12}y_{21}}{y_S + y_{11}}\right]} \tag{8.20}$$

8.4 S-Parameters

At this point the reader is well acquainted with the two-port concept using parameter sets like y and z parameters. In previous chapters travelling waves have been introduced to facilitate analysis and design of distributed circuit elements. The reflection coefficient is a quantity derived from this theory that is used to characterise one-port networks. The concept of reflection coefficients can be extended to networks with any number of ports although we will deal with no more than 2 ports in this text. So, instead of voltages and currents at a port we will think in terms of incident waves and reflected waves. Of course, since waves can scatter from one port to another the term reflection coefficient is not appropriate any more. When we have more than one port we instead refer to Scattering (S) parameters. These will be discussed in more detail in the following sections. We will also look into stability and gain concepts based on S parameters rather than y parameters. This is followed by a structured step-by-step design methodology that cover all aspects of designing an amplifier with respect to gain. Noise is another important topic that is brought up. Noise parameters are discussed followed by a design methodology for minimum or specific noise. The fact that we have to compromise between noise and gain is also discussed.

The strength of a wave that propagates in a transmission line can be described by the voltage or the current amplitude for example. If the characteristic impedance of the transmission line is known we can calculate the corresponding current or voltage amplitude, respectively. However, in connection with S parameters neither voltage nor current is used. Instead a more convenient normalised notation related to the power carried by the wave is used as we will see in the next section.

8.4.1 Definition of S-Parameters

A reflection coefficient Γ describes the properties of a linear one-port network. This concept can be extended to include linear networks with any number of ports. One incident wave, a_i, and one reflected wave, b_i, are associated with each port denoted by index i. In this text we will only deal with two-ports for the sake of clarity. This is sufficient for most applications. The treatment is easily generalised to N-port networks.

Consider the two-port in figure 8.12. For the input port there is an incident wave, a_1, that in the general case contributes to the reflected wave, b_1, at the same port but it also scatters through the two-port and contribute to the reflected wave, b_2, at the output port. The same reasoning goes for the output port.

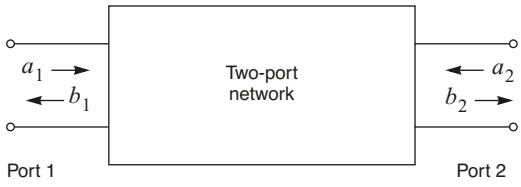


Figure 8.12 Two-port network with incident and reflected waves.

We conclude that we can define four (complex valued) parameters that characterise the two-port network, just as was the case for the y, z and ABCD parameters that were presented previously. Since incident waves can scatter from one port to another these parameters are referred to as scatter (S) parameters that are defined by

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2 \\ b_2 &= S_{21}a_1 + S_{22}a_2 \end{aligned} \tag{8.21}$$

or in matrix form

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} . \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{8.22}$$

So far we have not dealt with the travelling waves denoted a_i and b_i. As a matter of fact is does not matter if the waves are defined to be voltage travelling waves (V_i^+ and V_i^- respectively) or current travelling waves (I_i^+ and I_i^- respectively). Equation (8.22) is valid for both cases. However, in connection with S parameters these quantities are defined neither to be voltage waves nor current waves. Instead, a normalised notation is used.

Consider an incident wave V^+ and a reflected wave V^- at one port of a two-port. At this port we know that the total voltage equals

$$V = V^+ + V^- \tag{8.23}$$

and the net current equals

$$I = I^+ - I^- = \frac{V^+}{Z_0} - \frac{V^-}{Z_0} \tag{8.24}$$

where Z_0 is a reference impedance. The reflection coefficient for the port is given by

$$\Gamma = \frac{V^-}{V^+} \tag{8.25}$$

We now introduce a normalised notation such that

$$
\begin{aligned}
v &= \frac{V}{\sqrt{2Z_0}} \\
i &= \sqrt{Z_0/2}I \\
a &= \frac{V^+}{\sqrt{2Z_0}} \\
b &= \frac{V^-}{\sqrt{2Z_0}}
\end{aligned}
$$

where a and b are referred to as power waves, which is the definition used in connection with S parameters. Equations (8.23) to (8.25) can now be written as

$$
\begin{aligned}
v &= a + b \\
i &= a - b \\
b &= \Gamma a
\end{aligned}
\tag{8.26}
$$

This means that both the "normalised" voltage and current at a specific location can be expressed by the power waves a and b. Note that the normalisation does not affect the definition of the reflection coefficient, Γ.

The term power wave stems from the fact that the absolute value of the wave squared equals the power associated with that wave, i.e.,

$$
\begin{aligned}
P^+ &= \frac{\left|V^+\right|^2}{2Z_0} = |a|^2 \\
P^- &= \frac{\left|V^-\right|^2}{2Z_0} = |b|^2
\end{aligned}
\tag{8.27}
$$

Here, the wave quantities V^+ and V^- are given as peak values.

Thus the power delivered to a port can be express as the difference between the power carried by the incident and the reflected waves, respectively, i.e.,

$$
P_{IN} = P^+ - P^- = |a|^2 - |b|^2 \tag{8.28}
$$

We conclude this section by saying that the S parameters are defined based on a reference impedance Z_0 that is 50Ω by convention if nothing else is stated.

8.4.2 Properties of S-Parameters

From the definition of the S parameters in (8.21) we can define the individual parameters as follows:

$$S_{11} = \left.\frac{b_1}{a_1}\right|_{a_2 = 0} \qquad S_{12} = \left.\frac{b_1}{a_2}\right|_{a_1 = 0}$$

$$S_{21} = \left.\frac{b_2}{a_1}\right|_{a_2 = 0} \qquad S_{22} = \left.\frac{b_2}{a_2}\right|_{a_1 = 0}$$

This definition suggests how we should proceed to measure these quantities if we can measure the incident and reflected waves associated with each port. For example, S_{11} is obtained as the ratio of the reflected wave, b_1, at port 1 to the incident wave a_1 at the same port. Port 2 is assumed to be properly terminated. With properly terminated we mean that whatever port 2 transmit to its load nothing should be reflected back by the load again as an incident wave to port 2 because this would then propagate all the way back through S_{12} to port 1 and interfere with the measurement. Consequently, the load connected to port 2 should have a reflection coefficient equal to 0 which corresponds to 50Ω.

In contrast to this type of port termination we saw that to measure y, z and ABCD parameters we had to either short-circuit or open-circuit the ports to measure a parameter. This is not feasible at high frequencies. Actually, it is very difficult to realise well-defined broadband short-circuit and open-circuit terminations. This is not the case for a 50Ω termination. Moreover, even if it would be possible to realise well-defined short-circuit and open-circuit terminations they would in many cases force the measured device to oscillate. This is typically not the case with a 50Ω termination.

.

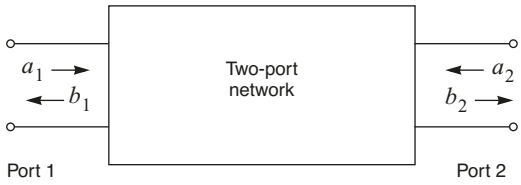


Figure 8.13 Two-port network with incident and reflected waves.

Once the S parameters have been measured there is no restriction in transforming them to other parameter sets. For example, y parameters are practical if we want to calculate the total y parameter set for two two-ports connected in parallel and ABCD parameters can be used when calculating the parameters of two two-ports in cascade. No information is lost when a parameter set is transformed to another type of parameters set.

Note that when transforming y, z and ABCD parameters to S parameters the parameters must be normalised with the reference impedance (50Ω) prior to the transformation. And the other way around, when going from S parameters to one of the other parameter sets the result will be normalised parameters that should be denormalised.

The chain scattering parameters or scattering transfer parameters (or T parameters for short) is a parameter set that is closely related to the S parameters. They are useful in analysis of cascaded networks. The total T parameter

matrix of a number of two-ports in cascade is equal to the matrix product of the T parameter matrices of all the individual two-ports. The T parameters are defined as

$$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \cdot \begin{bmatrix} b_2 \\ a_2 \end{bmatrix} \tag{8.29}$$

We note that the waves associated with port 1, a_1 and b_1, are the dependent variables whereas the waves associated with port 2, a_2 and b_2, are the independent variables. Combining (8.22) and (8.29) we get the relationship between the S parameters and the T parameters,

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{S_{21}} & -\frac{S_{22}}{S_{21}} \\ \frac{S_{11}}{S_{21}} & S_{12} - \frac{S_{11}S_{22}}{S_{21}} \end{bmatrix} \tag{8.30}$$

and

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} \frac{T_{21}}{T_{11}} & T_{22} - \frac{T_{21}T_{12}}{T_{11}} \\ \frac{1}{T_{11}} & -\frac{T_{12}}{T_{11}} \end{bmatrix} \tag{8.31}$$

A network analyser is used for S parameter measurements. A network analyser can be calibrated to eliminate the influence of the circuitry (cables, matching networks etc.) between the instrument and the device or circuitry that should be measured. This operation is usually referred to as moving the measurement reference plane. There is one simple and special case where it is very simple to perform this operation with hand-calculations that also demonstrates one property of the S parameters.

Assume that we have measured S parameters for the complete circuit in figure 8.14 and want S parameters for the two-port network without the transmission lines. The S parameters for the two-port network is defined by

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{8.32}$$

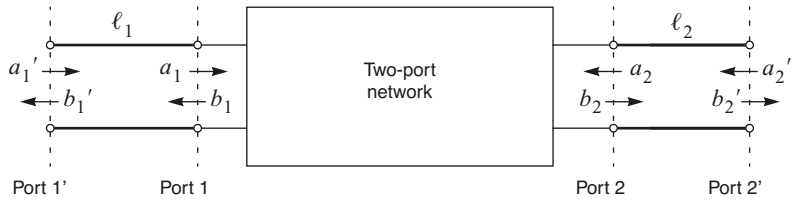


Figure 8.14 Different reference planes for a two-port network with transmission lines.

We also assume that the characteristic impedance of the transmission lines is $Z_0 = 50\Omega$ and that we know the complex propagation constant, γ, for the transmission line. In most cases the transmission line can be considered as lossless in which case we have that $\gamma = j\beta$. The measurement equipment can be calibrated to have reference planes at the connectors that we connect to port 1' and port 2'. The S parameters for the whole circuit are defined by

$$\begin{bmatrix} b_1' \\ b_2' \end{bmatrix} = \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \cdot \begin{bmatrix} a_1' \\ a_2' \end{bmatrix} \tag{8.33}$$

If we know the properties of the transmission lines we can calculate the S parameters for the two-port network itself since we can relate the travelling waves at port 1 and port 2 to the travelling waves at port 1' and port 2', respectively. Thus we can write

$$\begin{aligned} b_1 &= b_1' \cdot e^{\gamma \ell_1} \\ a_1 &= a_1' \cdot e^{-\gamma \ell_1} \\ b_2 &= b_2' \cdot e^{\gamma \ell_2} \\ a_2 &= a_2' \cdot e^{-\gamma \ell_2} \end{aligned} \tag{8.34}$$

If we substitute (8.34) into (8.33) we get

$$\begin{bmatrix} b_1 \cdot e^{-\gamma \ell_1} \\ b_2 \cdot e^{-\gamma \ell_2} \end{bmatrix} = \begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} \cdot \begin{bmatrix} a_1 \cdot e^{\gamma \ell_1} \\ a_2 \cdot e^{\gamma \ell_2} \end{bmatrix} \tag{8.35}$$

and rearranging gives

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11}' \cdot e^{2\gamma \ell_1} & S_{12}' \cdot e^{\gamma \ell_1 + \gamma \ell_2} \\ S_{21}' \cdot e^{\gamma \ell_1 + \gamma \ell_2} & S_{22}' \cdot e^{2\gamma \ell_2} \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{8.36}$$

, i.e.,

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} S_{11}' \cdot e^{2\gamma \ell_1} & S_{12}' \cdot e^{\gamma \ell_1 + \gamma \ell_2} \\ S_{21}' \cdot e^{\gamma \ell_1 + \gamma \ell_2} & S_{22}' \cdot e^{2\gamma \ell_2} \end{bmatrix} \tag{8.37}$$

Similarly, we can derive the relations the other way around which yields

$$\begin{bmatrix} S_{11}' & S_{12}' \\ S_{21}' & S_{22}' \end{bmatrix} = \begin{bmatrix} S_{11} \cdot e^{-2\gamma \ell_1} & S_{12} \cdot e^{-\gamma \ell_1 - \gamma \ell_2} \\ S_{21} \cdot e^{-\gamma \ell_1 - \gamma \ell_2} & S_{22} \cdot e^{-2\gamma \ell_2} \end{bmatrix} \tag{8.38}$$

8.4.3 Signal Flow Graphs

For y, z and ABCD parameters we can draw schematics that corresponds to the different parameter sets to make these representations less abstract. Unfortunately, this is not easily done for the S parameters. S parameters relies on travelling waves rather that node voltages or branch currents. For this reason we need a graphical representation that separates waves that travel through the same node or along the same transmission line but with opposite directions.

Signal flow graphs can be used in analysis of circuits with travelling waves. The concept is that a variable, a wave, is assigned a node. S parameters (for multi-port networks) and reflection coefficients (for one-port networks) are represented by branches between nodes and thus represent transfer functions from one node to a contribution to another node. Branches emanate from independent nodes and enter dependent nodes. Independent nodes correspond to incident waves whereas dependent nodes correspond to reflected or scattered waves. A signal flow graph for a one-port is shown in figure 8.15.

Figure 8.15 A one-port network and its corresponding signal flow graph.

From equation (8.21) we can directly develop the signal flow graph for a two-port represented by S parameters, see figure 8.16. The incident and independent wave at port 1, a_1, is partly reflected back again by S_{11} and contributes to the reflected and dependent wave at the input, b_1. It is also transmitted or scattered to port 2 through S_{21} and contributes to the reflected and dependent wave at the output, b_2. The same reasoning applies to the incident wave at the output, a_2.

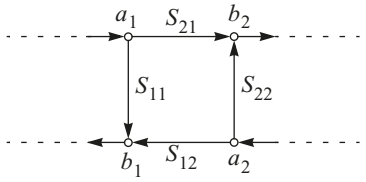


Figure 8.16 A two-port network represented by a signal flow graph.

What we have not discussed so far is how to represent a signal source. Clearly, a signal source is an independent variable. Consider the voltage signal source in figure 8.17. In this diagram there are two node voltages and one branch current. To transform this into a signal flow graph we have to reformulate these quantities to waves.

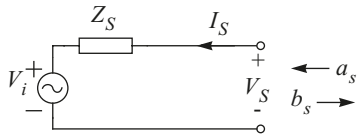


Figure 8.17 Voltage signal source.

Kirchoff's voltage law (KVL) gives

$$V_S = V_i + I_S Z_S \tag{8.39}$$

If we represent the source output voltage, V_S, and the current, I_S, with travelling waves we get

$$V_S^+ + V_S^- = V_i + (I_S^+ - I_S^-)Z_S = V_i + \left(\frac{V_S^+}{Z_0} - \frac{V_S^-}{Z_0}\right)Z_S \tag{8.40}$$

Solving for V_S^- we get

$$V_S^- = V_i \cdot \frac{Z_0}{Z_0 + Z_S} + V_S^+ \cdot \frac{Z_S - Z_0}{Z_0 + Z_S} \tag{8.41}$$

Finally, the travelling waves are normalised to power waves and we get

$$b_S = b_i + a_S \cdot \Gamma_S \tag{8.42}$$

where

$$b_S = \frac{V_S}{\sqrt{2Z_0}} \qquad a_S = \frac{V_S^+}{\sqrt{2Z_0}}$$

$$b_i = \frac{V_i}{\sqrt{2}} \frac{\sqrt{Z_0}}{Z_0 + Z_S} \qquad \Gamma_S = \frac{Z_S - Z_0}{Z_0 + Z_S}$$

and the signal flow graph can be drawn directly from equation (8.42) as shown in figure 8.18.

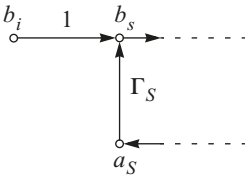


Figure 8.18 Signal source represented by a signal flow graph.

We have now presented how to develop signal flow graphs for three basic blocks. These can be combined to investigate a two-port with a source and a load connected to it, see figure 8.19.

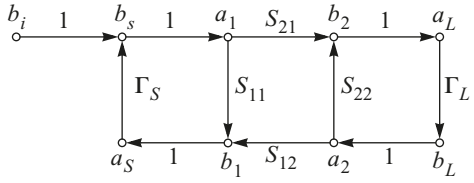


Figure 8.19 Signal flow graph for a two-port with signal source and load.

The three blocks were connected using unit value branches (1). For example, the wave that emanate from the source, b_S, is also the incident wave to the two-port. With the same reasoning applied to all other unit branches (except for the source, b_i, that is an independent variable) we can simplify the signal flow graph to the graph shown in figure 8.20.

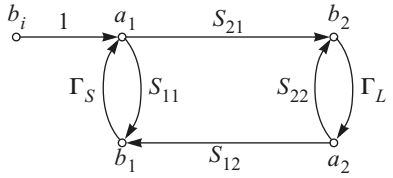


Figure 8.20 Simplified signal flow graph for a two-port with signal source and load.

Signal flow graphs are convenient for analysing systems with travelling waves. Once a signal flow graph has been developed Mason's rule (1 can be applied to derive the ratio of a dependent to an independent wave and indirectly therefore also the ratio of any variables. For example, it is possible derive expressions for various types of gain and reflection coefficients for any port.

8.5 Stability Analysis Using S-Parameters

We know from above that stability characteristics of a two-port network can be investigated by looking at the sign of the input and output port conductance (or resistance). If both the output and the input resistances are positive for all passive source and load impedances the two-port is said to be unconditionally stable. If either the output or the input resistance or both are negative for some configurations of source and load impedances the two-port is said to be conditionally stable. We also know that a negative resistive part of an impedance corresponds to a reflection coefficient with a magnitude larger than unity. Therefore, the conditions for stability can be reformulated using reflection coefficients as follows:

A two-port network is unconditionally stable if $|\Gamma_{IN}| < 1$ *and* $|\Gamma_{OUT}| < 1$ *for all* Γ_S *and* Γ_L *that satisfy* $|\Gamma_S| < 1$ *and* $|\Gamma_L| < 1$.

If the two-port is unilateral there is no coupling between the ports and we know that $\Gamma_{IN} = S_{11}$ and $\Gamma_{OUT} = S_{22}$. Thus, it is very easy to determine whether an unilateral two-port is unconditionally stable or not. A unilateral two-port is unconditionally stable if $|S_{11}| < 1$ and $|S_{22}| < 1$.

We can rewrite the condition above using S parameters. The input and output reflection coefficients for a two-port are given by

$$\Gamma_{IN} = S_{11} + S_{12}S_{21}\frac{\Gamma_L}{1 - S_{22}\Gamma_L} \tag{8.43}$$

and

$$\Gamma_{OUT} = S_{22} + S_{12}S_{21}\frac{\Gamma_S}{1 - S_{11}\Gamma_S} \tag{8.44}$$

Thus, the condition for a two-port to be unconditionally stable is ($|\Gamma_S| < 1$ and $|\Gamma_L| < 1$ assumed):

$$\begin{cases} |\Gamma_{IN}| = \left| S_{11} + S_{12}S_{21}\dfrac{\Gamma_L}{1 - S_{22}\Gamma_L} \right| < 1 \\ |\Gamma_{OUT}| = \left| S_{22} + S_{12}S_{21}\dfrac{\Gamma_S}{1 - S_{11}\Gamma_S} \right| < 1 \end{cases} \tag{8.45}$$

It is difficult to tell whether a given two-port is unconditionally stable or not from these inequalities. However, it can be shown that these inequalities can be reformulated to

$$|\Delta| < 1 \text{ and } K > 1 \tag{8.46}$$

where

$$\Delta = S_{11}S_{22} - S_{12}S_{21} \tag{8.47}$$

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{12}S_{21}|} \tag{8.48}$$

Thus, a simple stability check can be performed for a given device in the very beginning of the design work. The outcome of this stability check will determine the design methodology that should be used.

If the two-port is potentially unstable it means that some source or/and load impedances will result in an input or/and an output reflection coefficient whose magnitude is larger than unity. If this is case the device is still useful but care should be taken to avoid the source and load impedances that result in reflection coefficients larger than unity.

By solving $|\Gamma_{IN}| = 1$ and $|\Gamma_{OUT}| = 1$ for Γ_L and Γ_S, respectively, we obtain boundaries that divide the source and load reflection coefficients into two sets each. One set corresponds to $|\Gamma_{IN}| < 1$ or $|\Gamma_{OUT}| < 1$ (the stable region) and the other set corresponds to $|\Gamma_{IN}| > 1$ or $|\Gamma_{OUT}| > 1$ (the potentially unstable region). Fortunately, the boundaries are always defined by circles in the Smith chart and they are easy to calculate. The boundary in the Γ_S-plane that gives $|\Gamma_{OUT}| = 1$ is referred to as the input stability circle:

Radius of input stability circle:

$$r_S = \frac{|S_{12}S_{21}|}{\left||S_{11}|^2 - |\Delta|^2\right|} \tag{8.49}$$

Centre of input stability circle:

$$\Gamma_{SO} = \frac{S_{11}^* - \Delta^* S_{22}}{|S_{11}|^2 - |\Delta|^2} \tag{8.50}$$

and similarly the boundary in the Γ_L-plane that gives $|\Gamma_{IN}| = 1$ is referred to as the output stability circle:

Radius of output stability circle:

$$r_L = \frac{|S_{12}S_{21}|}{\left||S_{22}|^2 - |\Delta|^2\right|} \tag{8.51}$$

Centre of output stability circle:

$$\Gamma_{LO} = \frac{S_{22}^* - \Delta^* S_{11}}{|S_{22}|^2 - |\Delta|^2} \tag{8.52}$$

Figure 8.21 exemplifies what these stability circles might look like in the case of a potentially unstable two-port.

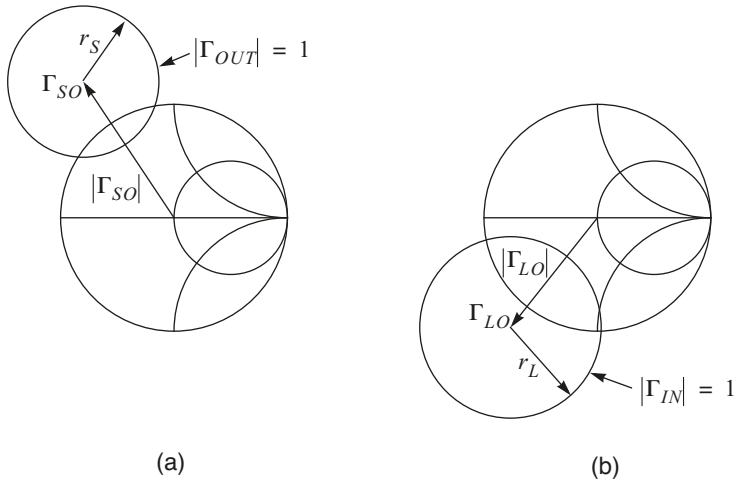


Figure 8.21 Example of (a) input stability circle (Γ_S-plane) and (b) output stability circle (Γ_L-plane).

We now know how to calculate the boundary but we do not know whether the stable region is within the circle or outside the circle.

It is readily understood that if we can prove that one arbitrary point is part of the stable set or the potentially unstable set, then this holds for all other points belonging to the same set. Lets consider the input stability circle as an example. We can pick an arbitrary Γ_S and calculate the corresponding Γ_{OUT} using (8.44). If $|\Gamma_{OUT}| < 1$ for this point in the Γ_S-plane we know that it belongs to the stable region and vice versa if $|\Gamma_{OUT}| > 1$. However, there is an even more simple method.

Again consider the input stability circle, that is in the Γ_S-plane. If we set $\Gamma_S = 0$ (the centre of the Smith-chart) then Γ_{OUT} becomes equal to S_{22}, i.e., we do not have to calculate Γ_{OUT}. This is easily shown by (8.44).

So, if $|S_{22}| < 1$ then $\Gamma_S = 0$ lies in the stable region but also all other values of Γ_S that are on the same side of the boundary, i.e., the input stability circle. However, if $|S_{22}| > 1$ then $\Gamma_S = 0$ lies in the potentially unstable region and consequently the other side of the stability circle constitutes the stable region.

Of course, the same reasoning goes for the output stability circle; If $|S_{11}| < 1$, the centre of the Smith-chart is a part of the stable region of Γ_L-values and the other way around if $|S_{11}| > 1$. This test is also illustrated in figures 8.22 and 8.23.

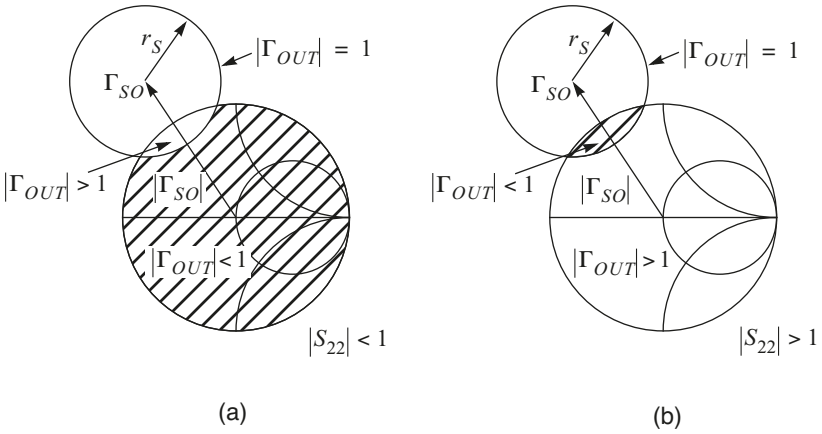


Figure 8.22 Input stability circles and stable and potentially unstable regions for (a) $|S_{22}| < 1$ and (b) $|S_{22}| > 1$. Stable region dashed.

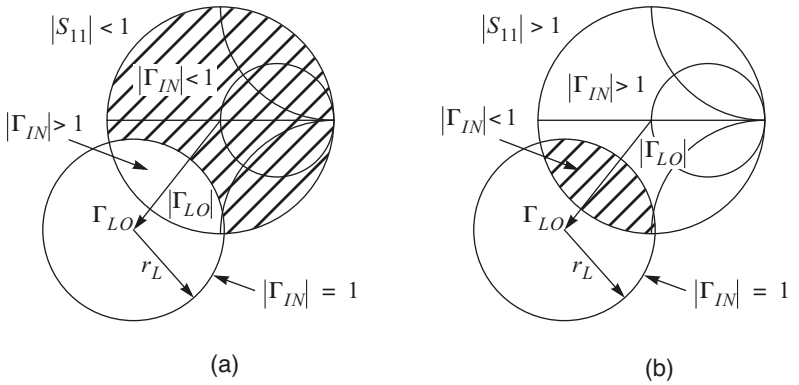


Figure 8.23 Output stability circles and stable and potentially unstable regions for (a) $|S_{11}| < 1$ and (b) $|S_{11}| > 1$. Stable region dashed.

Example 8.1 Stability properties of a transistor

Investigate the stability properties of BFR520 (Philips Semiconductors) with $I_C = 20\text{mA}$ and $V_{CE} = 6\text{V}$, at 300MHz and 2GHz. We get the following S parameters from the data book:

300MHz:

$S_{11} = 0.34\angle{-82.4°}$ $S_{12} = 0.043\angle{68.6°}$

$S_{21} = 14.18\angle{110.6°}$ $S_{22} = 0.50\angle{-33.4°}$

2GHz:

$S_{11} = 0.10\angle{163.6°}$ $S_{12} = 0.22\angle{69.6°}$

$S_{21} = 2.57\angle{65.1°}$ $S_{22} = 0.34\angle{-33.4°}$

To begin with, investigate whether the transistor is unconditionally stable or not at these frequencies by calculating $|\Delta|$ (8.47) and K (8.48).

300MHz: $|\Delta| = 0.56$, $K = 0.77$

2GHz: $|\Delta| = 0.52$, $K = 1.03$

A two-port network is only stable if $|\Delta| < 1$ and $K > 1$. Thus, the transistor is unconditionally stable at 2GHz but not at 300MHz.

The next step is to draw the stability circles at 300MHz to determine which source and load impedances that lie in the stable region (we already know that we can choose any passive source and load impedance for 2GHz so we do not need the stability circles in this case). We use equations (8.49) to (8.52) to calculate the centre and the radius of the circles:

Input stability circle, 300MHz:
centre: $\Gamma_{SO} = 1.148 - j2.138 = 2.427\angle{-61.8°}$
radius: $r_S = 3.115$

Output stability circle, 300MHz:
centre: $\Gamma_{LO} = -5.705 - j7.556 = 9.467\angle{-127.1°}$
radius: $r_L = 10.219$

Since $|S_{11}| < 1$ and $|S_{22}| < 1$ we know that the centre of the Smith-chart is stable both for the input and the output.

In the 2GHz case we saw that the transistor is unconditionally stable and we are free to choose any passive source and load impedance without violating the stability criteria. What does this means in terms of stability circles?

Input stability circle, 2GHz:
centre: $\Gamma_{SO} = 1.038 + j0.246 = 1.067\angle{13.4°}$
radius: $r_S = 2.133$

Output stability circle, 2GHz:
centre: $\Gamma_{LO} = -2.127 - j1.368 = 2.529\angle{-147.2°}$
radius: $r_L = 3.577$

We note that the difference between the radius and the length of the circle centre vector is more than one both for the input and the output stability circles. This means that the stability circles encircles the Smith-chart completely. Since $|S_{11}| < 1$ and $|S_{22}| < 1$ the whole Smith-chart is stable which we already knew from calculating $|\Delta|$ and K above. A Smith-chart with the stability circles is shown in figure 8.24.

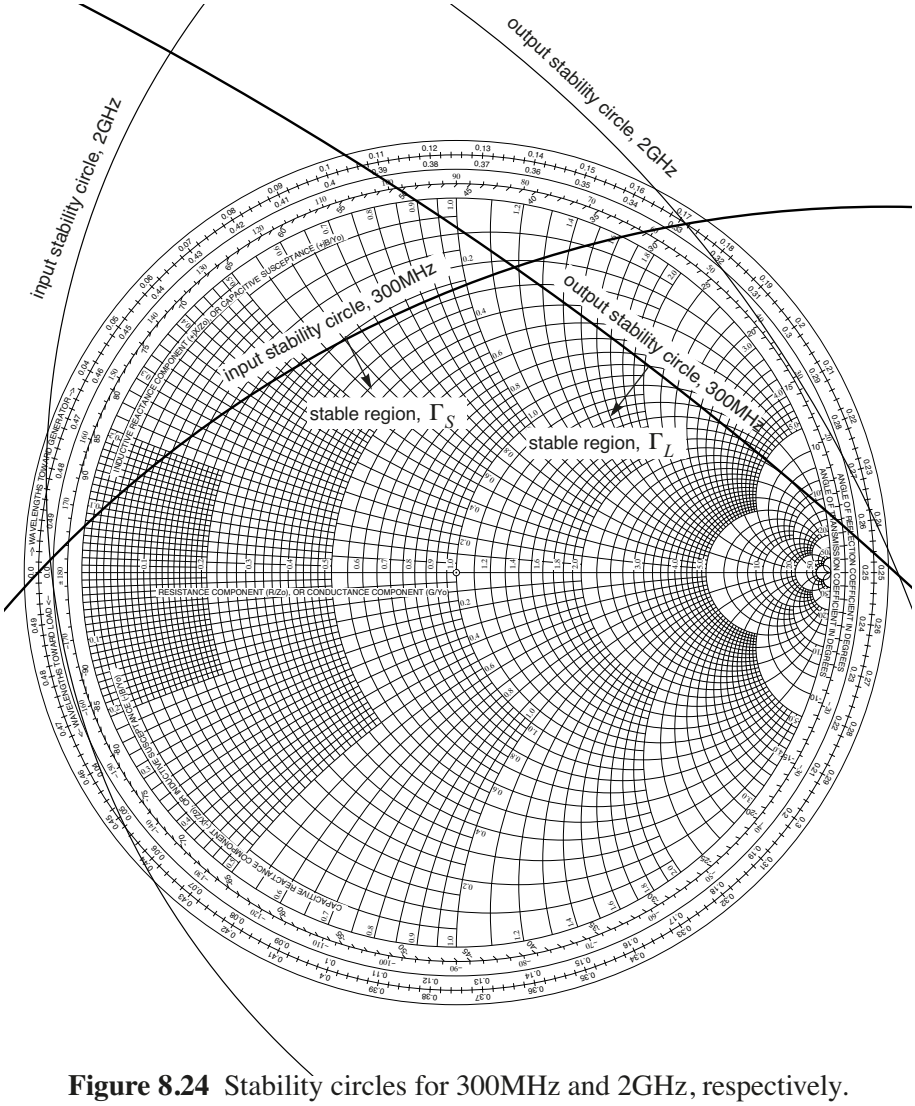


Figure 8.24 Stability circles for 300MHz and 2GHz, respectively.

This example demonstrates one problem in the design of amplifiers. Assume that we would use this transistor to design an amplifier operating at 2GHz. Once the source and load impedances have been determined to achieve, let say, a certain gain at 2GHz it is justified to ask if this amplifier will work properly at other frequencies as well. As a matter of fact, we cannot predict the behaviour of the amplifier at other frequencies than 2GHz based on the 2GHz parameters. As we saw above the transistor is conditionally stable at 300MHz and the source and load impedances might very well lie in the potentially unstable region at 300MHz. If this is the case, the amplifier might become an oscillator. To summarise, once the designer has completed the design for a specific frequency other frequencies must be investigated as well to ensure that the design will be stable for all frequencies below f_{max}.

We conclude this section by stressing the fact that we actually have two levels of stability requirements. If possible, the designer should strive for port impedances with a positive resistive part, i.e., stay within the stable regions bounded by the stability circles for all frequencies. If this is not possible, the designer should try to obtain a positive loop resistance.

8.6 Designing for Gain with S-Parameters

This section deals with the design of amplifiers with respect to gain using lossless (mis)matching networks to obtain the desired result. Typically, the source and the load that we want to connect to the two-port are fixed. If you would connect them directly to the two-port you would not get the desired gain, noise or whatever the design criteria was. Instead, (mis)matching networks are connected between the source and the two-port and between the load and the two-port, respectively, as shown in figure 8.25. Note that in this case we must make a distinction between the actual source and load and the source and load that the two-port sees. As illustrated in figure 8.25 Γ_S and Γ_L refer to the reflection coefficients as seen by the two-port whereas the actual source and load reflection coefficients are denoted $\Gamma_{S'}$ and $\Gamma_{L'}$.

The design methodologies described below determine the values on Γ_S and Γ_L to obtain the desired gain but not how the matching networks should be realised.

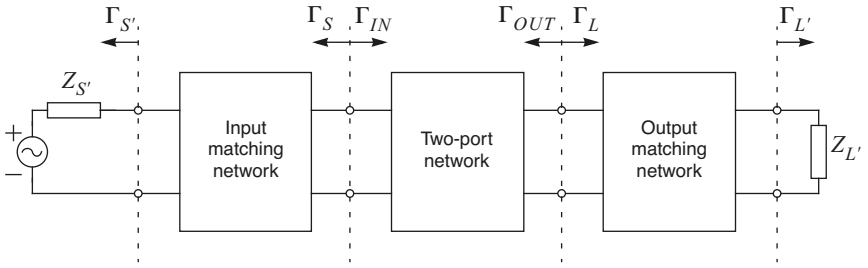


Figure 8.25 Two-port with source, load and matching networks.

Different methods must be applied depending on the characteristics of the two-port that will be used and whether we design for maximum gain or not. All three power gain definitions that were introduced previously will be used. Below, formulas are given for these power gains using S parameters. These formulas can be derived using Mason's rule. Note the partitioning of the formulas. They all contain three factors, one that represent the input side and one that represent the output side and finally one factor that equals forward scattering parameter, S_{21}, squared.

Operating Gain:

$$G_P = \frac{P_L}{P_{IN}} = \frac{1}{1-|\Gamma_{IN}|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \tag{8.53}$$

Transducer Gain:

$$\begin{aligned} G_T = \frac{P_L}{P_{AVS}} &= \frac{1-|\Gamma_S|^2}{|1-\Gamma_{IN}\Gamma_S|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \\ &= \frac{1-|\Gamma_S|^2}{|1-S_{11}\Gamma_S|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-\Gamma_{OUT}\Gamma_L|^2} \end{aligned} \tag{8.54}$$

Available Gain:

$$G_A = \frac{P_{AVN}}{P_{AVS}} = \frac{1-|\Gamma_S|^2}{|1-S_{11}\Gamma_S|^2} \cdot |S_{21}|^2 \cdot \frac{1}{1-|\Gamma_{OUT}|^2} \tag{8.55}$$

Here, Γ_{IN} and Γ_{OUT} are given by (8.43) and (8.44), respectively. For the transducer gain we note that if both the input and the output are terminated with 50Ω the gain becomes equal to $|S_{21}|^2$.

8.6.1 Design Methodologies

All in all there are four different cases that must be treated separately depending on whether the two-port can be considered to be unilateral or not and whether we design for maximum gain or for a specific gain, see figure 8.26. Design methodologies for these four cases will now be described in detail but first as a support for the unilateral cases the unilateral figure of merit will be defined.

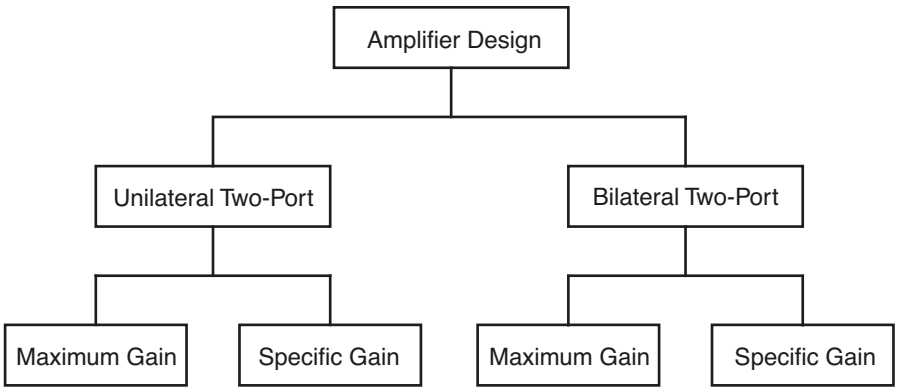


Figure 8.26 Four design cases where gain is the only design criteria.

8.6.1.1 Unilateral Figure of Merit

Before we start to present the various cases and their respective design methodologies we present a method to decide whether a two-port can be considered to be unilateral or not. You will hardly find a two-port with S_{12} being exactly zero but you may find a two-port where S_{12} is small enough to justify an approximation. The error introduced by the approximation is defined as the ratio of the bilateral transducer gain to the unilateral transducer gain and it can be shown that we can write this as

$$\frac{G_T}{G_{TU}} = \frac{1}{|1-X|^2} \tag{8.56}$$

where

$$X = \frac{S_{12}S_{21}\Gamma_S\Gamma_L}{(1-S_{11}\Gamma_S)(1-S_{22}\Gamma_L)} \tag{8.57}$$

We see that the ratio is bounded by

$$\frac{1}{(1+|X|)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1-|X|)^2} \tag{8.58}$$

If we design for maximum gain we would chose $\Gamma_S = S_{11}^*$ and $\Gamma_L = S_{22}^*$. In this special case the ratio becomes

$$\frac{1}{(1+U)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1-U)^2} \tag{8.59}$$

where

$$U = \frac{|S_{12}||S_{21}||S_{11}||S_{22}|}{(1-|S_{11}|^2)(1-|S_{22}|^2)} \tag{8.60}$$

This boundary is known as the unilateral figure of merit.

Example 8.2 Calculating unilateral figure of merit

Investigate if the transistor BFR520 (same as in example 8.1) with $I_C = 20\text{mA}$ and $V_{CE} = 6\text{V}$ can be considered as unilateral at 300MHz and 2GHz, respectively. The S parameters are

300MHz:

$S_{11} = 0.34\angle{-82.4°}$ $\quad$ $S_{12} = 0.043\angle{68.6°}$

$S_{21} = 14.18\angle{110.6°}$ $\quad$ $S_{22} = 0.50\angle{-33.4°}$

2GHz:

$S_{11} = 0.10\angle 163.6°$ $\quad S_{12} = 0.22\angle 69.6°$

$S_{21} = 2.57\angle 65.1°$ $\quad S_{22} = 0.34\angle -33.4°$

Calculate U (8.60) for 300MHz:

$$U = \frac{|0.043||14.18||0.34||0.5|}{(1-|0.34|^2)(1-|0.5|^2)} = 156\times 10^{-3}$$

Use (8.59) to calculate the unilateral figure of merit

$$0.748 < \frac{G_T}{G_{TU}} < 1.405 \text{ or } -1.26\text{dB} < \frac{G_T}{G_{TU}} < 1.48\text{dB}$$

Calculate U (8.60) for 2GHz:

$$U = \frac{|0.22||2.57||0.1||0.34|}{(1-|0.1|^2)(1-|0.34|^2)} = 22.0\times 10^{-3}$$

Calculate the unilateral figure of merit

$$0.958 < \frac{G_T}{G_{TU}} < 1.045 \text{ or } -0.19\text{dB} < \frac{G_T}{G_{TU}} < 0.19\text{dB}$$

Thus, we see in this case that the difference between the unilateral gain and bilateral gain is much larger for 300MHz than for 2GHz. As a rule of thumb a few tenths of dB in error is reasonable because we can hardly build an amplifier with higher accuracy anyway.

8.6.1.2 Case A: Unilateral Two-Port and Maximum Gain

Maximum power transfer and therefore also maximum gain is obtained with conjugate matching of the input and the output ports, i.e., $\Gamma_S = S_{11}^*$ and $\Gamma_L = S_{22}^*$. Substituting Γ_S with S_{11}^* and Γ_L with S_{22}^* in (8.54) gives:

Maximum Unilateral Transducer Gain:

$$G_{TUM} = \frac{1}{1-|S_{11}|^2} \cdot |S_{21}|^2 \cdot \frac{1}{1-|S_{22}|^2} \tag{8.61}$$

Conjugate matching of the ports works under the assumption that $S_{11} < 1$ and $S_{22} < 1$. When this is not the case the design becomes somewhat more complicated. We do not have the option to alter the load or the source impedance in such a way that the magnitude of Γ_{IN} and Γ_{OUT} becomes less than unity because they are fixed and given by S_{11} and S_{22}. This means that we must accept a negative port resistance and instead ensure that the loop resistance is positive. However, in this case there is no well-defined maximum gain. Theoretically the gain goes to infinity when the loop resistance goes to zero. While

this might sound as an interesting amplifier scheme one should keep in mind that the Q-value of the circuit also goes to infinity and consequently the bandwidth goes to zero, i.e., such an amplifier will not be able to amplify a modulated signal with a certain bandwidth. The reader might already have concluded that what we really get is an oscillator. So, in practice the design involves compromising between gain and bandwidth. Since the maximum gain is not well-defined we consider this to be a case of designing for a specific gain, i.e., case B that is treated below.

Example 8.3 Maximum gain of unilateral two-port

In example 8.2 it was demonstrated that the BFR520 transistor could be considered as being unilateral at 2GHz with an error of around 0.2dB having conjugate match of both ports.

Calculate the maximum unilateral gain and determine the reflection coefficients for the source and the load. Also calculate the bilateral gain using (8.54) to verify size of the discrepancy. The S parameters are

$$S_{11} = 0.10\angle 163.6^\circ \qquad S_{12} = 0.22\angle 69.6^\circ$$
$$S_{21} = 2.57\angle 65.1^\circ \qquad S_{22} = 0.34\angle -33.4^\circ$$

The maximum unilateral gain is given by (8.61):

$$G_{TUM} = \frac{1}{1-|S_{11}|^2}\cdot|S_{21}|^2\cdot\frac{1}{1-|S_{22}|^2}$$
$$= \frac{1}{1-|0.10|^2}\cdot|2.57|^2\cdot\frac{1}{1-|0.34|^2} \approx 7.54 \approx 8.8\text{dB}$$

The reflection coefficients for the source and the load are the conjugates of S_{11} and S_{22}, respectively, i.e., $\Gamma_S = 0.10\angle -163.6^\circ$ and $\Gamma_L = 0.34\angle 33.4^\circ$.

Use the following expression to calculate the bilateral gain (8.54):

$$G_T = \frac{1-|\Gamma_S|^2}{|1-\Gamma_{IN}\Gamma_S|^2}\cdot|S_{21}|^2\cdot\frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2}$$

we note that we have to calculate Γ_{IN} first. Equation (8.43) gives

$$\Gamma_{IN} = S_{11} + S_{12}S_{21}\frac{\Gamma_L}{1-S_{22}\Gamma_L} = 0.313\angle 166.6^\circ$$

and finally the bilateral transducer gain becomes

$$G_T = \frac{1-0.10^2}{|1-0.031\angle -3.0^\circ|^2}\cdot 2.57^2\cdot\frac{1-0.34^2}{|1-0.116|^2}$$
$$\approx 7.90 \approx 9.0\text{dB}$$

As expected the difference between the G_T and G_{TU} is the same as we calculated in example 8.2.

8.6.1.3 Case B: Unilateral Two-Port and Specific Gain

When we want to design for a specific gain (less than maximum gain, see case A) we must introduce mismatch at least at one of the ports to reduce the gain. To begin with we will assume $|S_{11}| < 1$ and $|S_{22}| < 1$. The transducer gain is given by (8.54) and for a unilateral two-port the transducer gain can be written as:

Unilateral Transducer Gain:

$$G_{TU} = \frac{1-|\Gamma_S|^2}{|1-S_{11}\Gamma_S|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2} \tag{8.62}$$

It is readily seen that the expression contains one factor that is associated with the input port and another factor that is associated with the output port. For convenience, we introduce the following notation:

$$G_{TU} = \alpha_S \cdot |S_{21}|^2 \cdot \alpha_L \tag{8.63}$$

where α_S and α_L corresponds to the first and third factor in (8.62), respectively. Also, from (8.61) we have that

$$\max(\alpha_S) = \frac{1}{1-|S_{11}|^2} \text{ and } \max(\alpha_L) = \frac{1}{1-|S_{22}|^2}.$$

Thus, we can rewrite the transducer gain as

$$\begin{aligned} G_{TU} &= \frac{\alpha_S}{\max(\alpha_S)} \cdot \max(\alpha_S)|S_{21}|^2\max(\alpha_L) \cdot \frac{\alpha_L}{\max(\alpha_L)} \\ &= \frac{\alpha_S}{\max(\alpha_S)} \cdot G_{TUM} \cdot \frac{\alpha_L}{\max(\alpha_L)} \\ &= g_S \cdot G_{TUM} \cdot g_L \end{aligned} \tag{8.64}$$

where $0 \le g_S \le 1$ and $0 \le g_L \le 1$. This means that we can relate our design goal to the maximum transducer gain, which is easily calculated, and apportion the gain backoff, g_S and g_L, to the ports. One question remains, when g_S and g_L have been calculated how do we go by to determine the values for Γ_S and Γ_L? By solving g_S and g_L for Γ_S and Γ_L, respectively we obtain circles in the Γ_S-plane and the Γ_L-plane that satisfy fixed g_S and g_L. These circles are given by:

Radius of input gain circle:

$$r_S = \frac{\sqrt{1-g_S}\left(1-|S_{11}|^2\right)}{1-|S_{11}|^2(1-g_S)} \tag{8.65}$$

Centre of input gain circle:

$$\Gamma_{SO} = \frac{g_S|S_{11}|}{1-|S_{11}|^2(1-g_S)} \cdot e^{j\arg(S_{11}^*)} \tag{8.66}$$

Radius of output gain circle:

$$r_L = \frac{\sqrt{1-g_L}\left(1-|S_{22}|^2\right)}{1-|S_{22}|^2(1-g_L)} \tag{8.67}$$

Centre of output gain circle:

$$\Gamma_{LO} = \frac{g_L|S_{22}|}{1-|S_{22}|^2(1-g_L)} \cdot e^{j\arg(S_{22}^*)} \tag{8.68}$$

For a given circle, higher gain is obtained if you choose a point within the circle and consequently less gain is obtained if a point outside the circle is chosen.

In the beginning of this section we introduced a restriction; $|S_{11}| < 1$ and $|S_{22}| < 1$. When the magnitude of at least one of S_{11} and S_{22} is larger than unity there is no well-defined maximum gain to relate to. As a matter of fact, the maximum gain goes to infinity. This is easily understood from (8.62) where we see that the gain factors associated with the input and the output, α_S and α_L, goes to infinity when

$$\Gamma_S = 1/S_{11} \tag{8.69}$$

and

$$\Gamma_L = 1/S_{22} \tag{8.70}$$

respectively. In terms of impedances, these relations basically state that the resistive part of the source and load equals the magnitude of the negative resistive parts of S_{11} and S_{22}, respectively. Thus, we should consider the loop resistance instead. Since the loop resistances should be positive the stable regions are given by

$$\mathrm{Re}[Z_S] > |\mathrm{Re}[Z_{IN}]| \tag{8.71}$$

and

$$\mathrm{Re}[Z_L] > \left|\mathrm{Re}[Z_{OUT}]\right| \tag{8.72}$$

It is fairly easy to mark the stable regions in the Smith-chart even if we think in terms of reflection coefficients. The question is how we go by to mark S_{11} or S_{22} in a regular Smith-chart, which only covers positive resistances.

Actually, it is possible to use the normal Smith-chart even for reflection coefficients with a magnitude larger than unity. By locating $1/\Gamma^*$ (i.e. invert the magnitude and negate the phase) in the normal Smith-chart the resistance circles are interpreted as negative as the reactance circles as labelled. This is easy to derive. Assume that Γ is associated with a normalised impedance $z = r + jx$ where r is negative. Then we can write

$$\begin{aligned}\frac{1}{\Gamma^*} &= \left(\frac{z+1}{z-1}\right)^* = \frac{z^*+1}{z^*-1} \\ &= \frac{r-jx+1}{r-jx-1} = \frac{-r+jx-1}{-r+jx+1} \\ &= \frac{z'-1}{z'+1} = \Gamma'\end{aligned} \tag{8.73}$$

where z' has the reactance as z but the opposite sign of the resistance. Γ' is the corresponding reflection coefficient. Thus, when, lets say, $1/S_{11}^*$ is plotted in the Smith-chart the stable region for Γ_S is given by the inner side of the resistance circle which $1/S_{11}^*$ lies upon, see figure 8.27. Of course, the same reasoning holds for S_{22}.

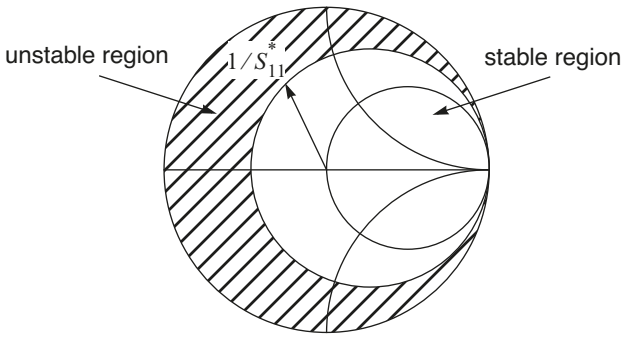


Figure 8.27 Stable and unstable regions for Γ_S when $\left|S_{11}\right| > 1$.

So far only the stability issue has been addressed for this case. When it comes to designing for a specific gain it can be shown that the gain circles defined by (8.65) to (8.68) can be used without modifications.

Example 8.4 Amplifier with specific gain using unilateral approximation

Design an amplifier with 12dB gain at 1GHz using the BFR520 transistor with $I_C = 20\text{mA}$ and $V_{CE} = 6\text{V}$. We will assume that we can use the unilateral approximation, i.e., set S_{12} to zero. The S parameters are

$$S_{11} = 0.135\angle{-137.3°} \qquad S_{12} = 0.114\angle{73.2°}$$
$$S_{21} = 4.799\angle{83.4°} \qquad S_{22} = 0.365\angle{-30.5°}$$

Since $|S_{11}| < 1$ and $|S_{22}| < 1$ the transistor is unconditionally stable (assuming $S_{12} = 0$).

Begin by calculating the maximum unilateral gain:

$$G_{TUM} = \frac{1}{1-|S_{11}|^2} \cdot |S_{21}|^2 \cdot \frac{1}{1-|S_{22}|^2} \approx 27.06 \approx 14.3\,\text{dB}$$

Now we have the freedom to apportion the excess gain of 2.3dB to the input and output ports. Since we do not have any additional criteria that restrict the design we can do as we please. Lets say that we divide the gain reduction equally between the ports, i.e., 1.15dB:

$$g_S = g_L = -1.15\text{dB} = 0.767$$

From these quantities we can calculate the gain circles for the input and output using (8.65) to (8.68):

Input gain circle:

centre: $\Gamma_{SO} = -0.0764 + j0.0705 = 0.1040\angle{-137.3°}$
radius: $r_S = 0.4759$

Output gain circle:

centre: $\Gamma_{LO} = 0.2489 + j0.1466 = 0.2889\angle{-30.5°}$
radius: $r_L = 0.4318$

These circles are shown in the Smith-chart in figure 8.28. To be able to design matching networks we have to choose one point from each gain circle. There are no general rules that we can apply to choose the "right" points. Points can be selected so that a certain matching network topology can be used or such that a certain bandwidth is obtained etc.

Assume that we choose the points as marked in the Smith-chart below (under what circumstances are these points practical?), i.e.,

$\Gamma_S = 0.4336\angle{-115.6°}$ or $z_S = 0.5198 - j0.5008$

$\Gamma_L = 0.1827\angle{-101.0°}$ or $z_L = 0.8760 - j0.3251$

Now, since we have specific values on Γ_S and Γ_L we can verify if it is feasible to use the unilateral approximation or not. Use (8.57) and (8.58) to calculate the error:

$$X = \frac{S_{12}S_{21}\Gamma_S\Gamma_L}{(1-S_{11}\Gamma_S)(1-S_{22}\Gamma_L)} = 0.0206 + j0.0351 \quad \text{and thus}$$

$$\frac{1}{(1+|X|)^2} < \frac{G_T}{G_{TU}} < \frac{1}{(1-|X|)^2} \quad \rightarrow \quad -0.35\text{dB} < \frac{G_T}{G_{TU}} < 0.36\text{dB}$$

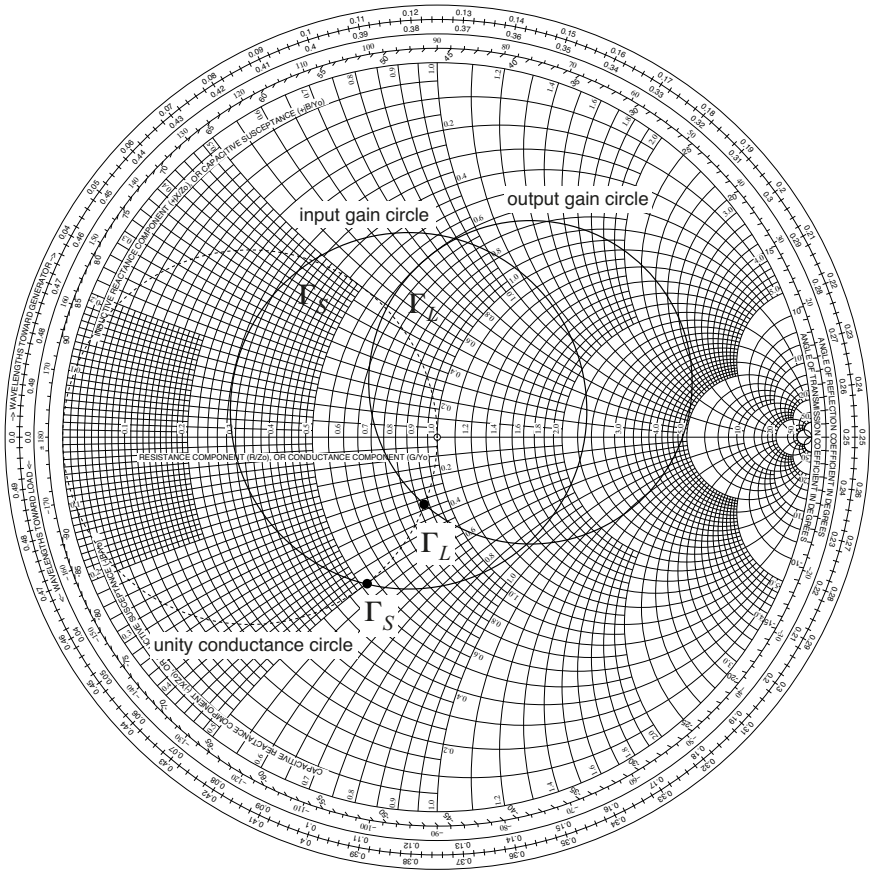


Figure 8.28 Gain circles for unilateral stage.

8.6.1.4 Case C: Bilateral Two-Port and Maximum Gain

When the two-port cannot be considered as unilateral the design procedure is somewhat more complicated. Due to the bidirectional coupling between the input port and the output port we cannot alter the conditions for one port without affecting the other port. In the case of maximum gain we know that we have to fulfil

$$\Gamma_S = \Gamma_{IN}^* \tag{8.74}$$

and

$$\Gamma_L = \Gamma_{OUT}^* \tag{8.75}$$

Due to the dependency between the ports this design requirement is referred to as simultaneous conjugate match and Γ_S and Γ_L must be determined simultaneously.

Substituting Γ_{IN} and Γ_{OUT} with (8.43) and (8.44), respectively we get

$$\Gamma_S = \left(S_{11} + S_{12}S_{21}\frac{\Gamma_L}{1 - S_{22}\Gamma_L}\right)^* \tag{8.76}$$

and

$$\Gamma_L = \left(S_{22} + S_{12}S_{21}\frac{\Gamma_S}{1 - S_{11}\Gamma_S}\right)^* \tag{8.77}$$

Thus, these two equations must be solved simultaneously. The solution is given by

$$\Gamma_{SM} = \frac{B_1 - \sqrt{B_1^2 - 4|C_1|^2}}{2C_1} \tag{8.78}$$

and

$$\Gamma_{LM} = \frac{B_2 - \sqrt{B_2^2 - 4|C_2|^2}}{2C_2} \tag{8.79}$$

where

$$B_1 = 1 + |S_{11}|^2 - |S_{22}|^2 - |\Delta|^2 \tag{8.80}$$

$$B_2 = 1 + |S_{22}|^2 - |S_{11}|^2 - |\Delta|^2 \tag{8.81}$$

$$C_1 = S_{11} - \Delta S_{22}^* \tag{8.82}$$

$$C_2 = S_{22} - \Delta S_{11}^* \tag{8.83}$$

It can be shown that simultaneous conjugate match is only possible if the two-port is unconditionally stable, i.e., $|\Delta| < 1$ and $K > 1$ must be fulfilled. It can also be shown that, under these conditions, the maximum transducer gain is simply given by

$$G_{TM} = \frac{|S_{21}|}{|S_{12}|}(K - \sqrt{K^2 - 1}) \tag{8.84}$$

From this expression we can derive another important quantity, namely the maximum stable gain which is equal to G_{TM} with K set to 1, i.e.,

$$G_{MSG} = \frac{|S_{21}|}{|S_{12}|} \tag{8.85}$$

For a conditionally stable two-port this quantity is a figure of merit that represents the maximum value that G_{TM} can have with proper resistive loading, i.e., the two-port is forced to be unconditionally stable by tapping some performance from the amplifying device. It can be shown that by having a shunt or series resistor on either or both ports it is possible to alter the value of K in (8.84) without changing the value of $|S_{21}/S_{12}|$.

In the event that we have a conditionally stable two-port there is no such thing as maximum gain, instead as we increase the gain we will get closer to the unstable region and finally end up with an oscillator. A pragmatic approach is used in this case. We start by calculating the maximum stable gain for the conditionally stable two-port using (8.85). From this gain we back off a few dB or so and design for that lower gain. Thus, we will design for a specific gain rather than for maximum gain that is not defined for this case. Therefore, when we have a conditionally stable two-port we should use the method described in the next section.

Example 8.5 Amplifier with maximum gain

Determine the source and the load reflection coefficients for maximum gain using the transistor configuration in example 8.4 (BFR520, 1GHz, 6V, 20mA). Also calculate the value of the maximum gain.

The S parameters are

$S_{11} = 0.135\angle{-137.3°}$ $\quad S_{12} = 0.114\angle{73.2°}$

$S_{21} = 4.799\angle{83.4°}$ $\quad S_{22} = 0.365\angle{-30.5°}$

Start by checking the stability properties to ensure that maximum gain (simultaneous conjugate matching of input and output) is achievable. We get

$|\Delta| = 0.5078$ and $K = 1.0112$

Since $|\Delta| < 1$ and $K > 1$ the transistor is unconditionally stable although K is very close to unity.

Equations (8.78) and (8.79) gives

$\Gamma_{SM} = -0.7181 + j0.2629 = 0.7647\angle{159.9°}$ and

$\Gamma_{SM} = 0.6632 + j0.4843 = 0.8236\angle{36.4°}$

The maximum gain is given by (8.84):

$$G_{TM} = \frac{|S_{21}|}{|S_{12}|}(K - \sqrt{K^2 - 1}) = 36.25 = 15.6\text{dB}$$

8.6.1.5 Case D: Bilateral Two-Port and Specific Gain

We saw previously that to design for a specific gain in the unilateral case we introduce more or less mismatch at the ports to meet the gain specification. This is simple because we start by calculating the maximum gain. Then, the difference between the wanted gain and the maximum gain (gain reduction) is apportioned as desired to both the ports. Finally, the matching networks for each port is designed separately. However, this is not possible with a bilateral two-port because of the coupling between the ports.

For bilateral two-ports the most common methods are based on using either the operating gain or the available gain definitions as a starting point because these are only dependent on what is connected to one of the ports (the load impedance for the operating gain (Γ_L) and the source impedance (Γ_S) for the available gain). Now, typically the specified gain is given as transducer gain, not operating or available gain. So what is the rationale in using operating gain or available gain when we want transducer gain? The answer to that is that these methods assume that mismatch is only introduced at the port considered in the gain definition (only one of Γ_L and Γ_S). The other port is assumed to be matched. This makes the operating and available gain equal to the transducer gain in the end.

Lets say that we design with operating gain, $G_P = P_L/P_{IN}$. Since we assume power matching of the input port the delivered power will be equal to the available power from the source, i.e., $P_{IN} = P_{AVS}$ and consequently $G_P = P_L/P_{IN} = P_L/P_{AVS} = G_T$. The same reasoning holds for available gain.

We will now describe the methods in more detail. To begin with, consider the operating gain given by

$$G_P = \frac{1}{1-|\Gamma_{IN}|^2} \cdot |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{|1-S_{22}\Gamma_L|^2}$$

As concluded earlier the operating gain is not dependent on Γ_S and we can rewrite and expand the expression in the following way:

$$G_P = |S_{21}|^2 \cdot \frac{1-|\Gamma_L|^2}{\left(1-\left|\frac{S_{11}-\Delta\Gamma_L}{1-S_{22}\Gamma_L}\right|^2\right)|1-S_{22}\Gamma_L|^2} = |S_{21}|^2 g_P \tag{8.86}$$

Thus, we set the operating gain to the specified gain and since $|S_{21}|^2$ is fixed we can calculate g_P, which is dependent on the known S parameters as well as the reflection coefficient for the load, Γ_L. At this stage we are used to circles in the Smith chart and this case is no exception. The set of Γ_L values that satisfies a given g_P forms a circle in the Γ_L-plane given by

Radius of operating gain circle:

$$r_L = \frac{\sqrt{1-2K|S_{12}S_{21}|g_P + |S_{12}S_{21}|^2 g_P^2}}{\left|1-g_P(|S_{22}|^2-|\Delta|^2)\right|} \tag{8.87}$$

Centre of operating gain circle:

$$\Gamma_{LO} = \frac{g_P C_L{}^*}{1+g_P(|S_{22}|^2-|\Delta|^2)} \tag{8.88}$$

where

$$C_L = S_{22} - \Delta S_{11}^* \tag{8.89}$$

Once the circle has been calculated a value for Γ_L is chosen from that circle. Next, the input reflection coefficient, Γ_{IN}, can be calculated and the source reflection coefficient is chosen to be $\Gamma_S = \Gamma_{IN}^*$.

Similarly, if we decide to have the mismatch at the input we should use available gain to determine the reflection coefficient for the source. The available gain can be written as

$$G_A = |S_{21}|^2 \cdot \frac{1-|\Gamma_S|^2}{\left(1-\left|\frac{S_{22}-\Delta\Gamma_S}{1-S_{11}\Gamma_S}\right|^2\right)|1-S_{11}\Gamma_S|^2} = |S_{21}|^2 g_A \tag{8.90}$$

The procedure is equivalent to the case with the operating gain above in that we calculate a parameter, g_A in thus case, as the ratio of the wanted gain (available gain in this case) to the forward S parameter, $|S_{21}|^2$. Once g_A is known we obtain a set of Γ_S values that satisfies a given g_A that forms a circle in the Γ_S-plane given by

Radius of available gain circle:

$$r_S = \frac{\sqrt{1-2K|S_{12}S_{21}|g_A + |S_{12}S_{21}|^2 g_A^2}}{\left|1-g_A(|S_{11}|^2-|\Delta|^2)\right|} \tag{8.91}$$

Centre of available gain circle:

$$\Gamma_{SO} = \frac{g_A C_S{}^*}{1+g_A(|S_{11}|^2-|\Delta|^2)} \tag{8.92}$$

where

$$C_S = S_{11} - \Delta S_{22}^* \tag{8.93}$$

So far we have not discussed how we should deal with the stability properties of the two-port. When the two-port is unconditionally stable we can design an amplifier using the method above without restrictions. However, when the two-port is only conditionally stable we know that there is no well-defined maximum gain and Γ_S and Γ_L cannot be chosen arbitrarily because they must stay within the stable regions.

For a conditionally stable two-port we start by calculating the maximum stable gain G_{MSG} given by (8.85). The realisable gain is a few dB or so below G_{MSG} so we have to make sure that the required gain is less. Depending on our preferences we choose to design for mismatch at either the output or the input as described above. In the case of mismatch at the output port we should choose a point, Γ_L, on the gain circle that lies in the stable region and not to close to the boundary of the stable region. After that, Γ_{IN} is calculated and from that we get $\Gamma_S = \Gamma_{IN}^*$, i.e., conjugate match of the input port. Finally, we must make sure that Γ_S also lies in the stable region. If it does not, a new point from the Γ_L-gain circle is chosen and the procedure is repeated. If this does not work either the gain must be reduced or another device must be considered for the design.

Example 8.6 Amplifier with specific gain

Determine the source and load reflection coefficients for an amplifier to obtain 16dB gain at 500MHz using the BFR520 transistor with $I_C = 5\text{mA}$ and $V_{CE} = 3\text{V}$. Apply mismatch at the output port and conjugate match at the input port.

According to data sheets the S parameters are

$$S_{11} = 0.473\angle{-84.8}^\circ \qquad S_{12} = 0.0760\angle{59.6}^\circ$$
$$S_{21} = 7.1790\angle{112.0}^\circ \qquad S_{22} = 0.5950\angle{-34.7}^\circ$$

We should verify that 20dB gain can be obtained by calculating maximum stable gain, G_{MSG}, if the transistor is conditionally stable or the maximum gain, G_{MAX}, if the transistor is unconditionally stable. Therefore, begin by checking the stability properties. Using the S-parameters above we get

$$|\Delta| = 0.5161 \text{ and } K = 0.6310$$

Since $K < 1$ the transistor is only conditionally stable and thus we should calculate the maximum stable gain

$$G_{MSG} = \left|\frac{S_{21}}{S_{12}}\right| = 94.5 = 19.75\text{dB}$$

This means that there is some headroom from the 16dB that we want. The next step is to draw the stability circles using equations (8.49) to (8.52):

Input stability circle:

centre: $\Gamma_{SO} = 6.181 - j10.518 = 12.200\angle{-59.6}^\circ$

radius: $r_S = 12.806$

Output stability circle:

centre: $\Gamma_{LO} = 3.638 + j5.859 = 6.897\angle 58.2°$

radius: $r_L = 6.222$

The stability circles are shown in the Smith-chart below. Since $|S_{11}| < 1$ and $|S_{22}| < 1$ we know that the centre of the Smith-chart is stable both for the input and the output.

now when we know what areas to avoid in the Smith-chart we should calculate the gain circle that gives 16dB total gain.

Finally, we should calculate the gain circle. Since we will control the gain by applying mismatch at the output port we should use the operating gain circle given by equations (8.87) and (8.88). The gain parameter g_P, see equation (8.86), is obtained from

$$G_P = |S_{21}|^2 g_P \text{ i.e. } g_P = G_P / |S_{21}|^2 = 39.81 / 51.54 = 0.772$$

and the gain circle becomes

centre: $\Gamma_{LO} = 0.2307 + j0.3715 = 0.4373\angle 58.2°$

radius: $r_L = 0.753$

We can choose any reflection coefficient for the load, Γ_L, that lies in the stable region and in this case we can for example choose the point marked in the Smith-chart below in figure 8.29, i.e.,

$$z_L = 0.6860 - j0.4635 \text{ or } \Gamma_L = 0.3202\angle -108.7°$$

Assuming that the load that is to be matched to the transistor output is 50Ω the matching network will only consist of an capacitor or a shunt stub (acting as an capacitor at 500MHz).

The final step is to choose the reflection coefficient for the source, Γ_S. The method implies that the input side should be conjugately matched when the output side is mismatched. Since Γ_L has been determined we can calculate Γ_{IN} using (8.43) and we get

$$\Gamma_{IN} = 0.1245 - j0.3443 = 0.3661\angle -70.1°$$

Thus, conjugate match means $\Gamma_S = \Gamma_{IN}^* = 0.3661\angle 70.1°$ as far as reflection coefficients are concerned.

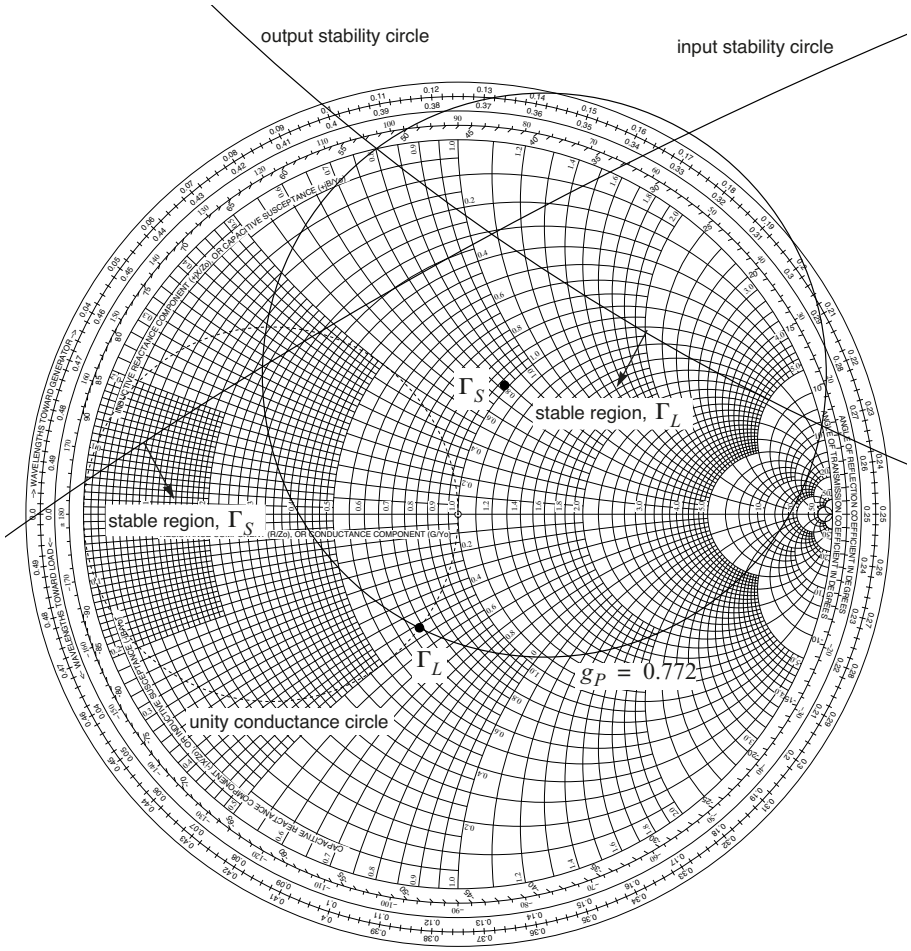


Figure 8.29 Gain circle and stability circles for bilateral stage.

8.7 Noise

One very important application for small-signal high frequency amplifiers is the first amplifier stage of a radio receiver. This amplifier must be designed to handle both strong signals and very weak signals. The capability to handle strong signals is limited by the linearity of the amplifier. If the input signal is sufficiently strong the amplifier will generate intermodulation distortion that will interfere with the desired signal. The capability to handle weak signals is limited by the noise generated by the amplifier itself that also will interfere with the desired signal. Below a method for designing low noise amplifiers is described. Linearity properties is discussed in the chapter on power amplifiers.

8.7.1 Noise in a Two-Port Network

Consider the two-port network in figure 8.30. Signal power, P_{Si}, generated by the source enters the two-port together with the noise power, P_{Ni}, also generated by the source. The ratio of the signal power to the noise power is referred to as signal-to-noise ratio, SNR. SNR is obviously a measure of the signal quality. We want the desired signal P_{Si} to be sufficiently large compared with the noise P_{Ni} so that whatever information that is carried by the desired signal it should be possible to retrieve it without significant loss of information.

The signal and the noise enter the input of the two-port network and are amplified with the same gain, G, before they reach the output load, Z_L. For a given source impedance the two-port network will also add a certain amount of noise power, P_{No}, to the output signal. Thus, the noise level increases more than the signal level and consequently the SNR will decrease.

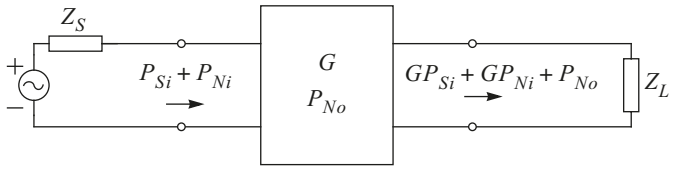


Figure 8.30 Noisy two-port network with noisy source.

The noise power, P_{No}, could be a useful measure of the noise performance for the two-port network but it is never used in practice. Instead the noise figure, denoted F, is used to characterise the noise properties of a two-port network. The noise figure is defined as the ratio of the total available noise power at the output of the two-port network to the available noise power at the output due to thermal noise from the source resistor having a specific temperature $T_0 = 300°\text{K}$.

The thermal noise arises due to thermal motions of the electrons. It is white in its nature, i.e., the spectral density is constant. This is true for frequencies below 10^{13} Hz or so. The thermal noise can be modelled as a voltage source in series with the resistor or as a current source in parallel with the resistor as shown in figure 8.31.

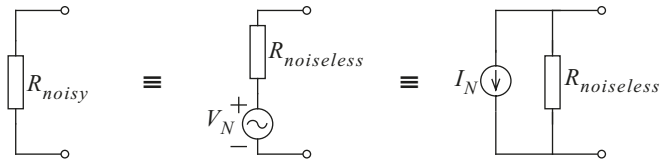


Figure 8.31 Modelling of the thermal noise in a resistor.

The rms value of the noise voltage is given by

$$V_N = \sqrt{4kTBR} \tag{8.94}$$

where k is Boltzmann's constant, T the temperature in the resistor and B is the noise bandwidth. The equivalent rms noise current is given by

$$I_N = \sqrt{4kTB/R} \tag{8.95}$$

The maximum available noise power from a source impedance with a resistive part, R, is transferred to a load if the load is conjugately matched. Thus, the maximum available noise power is given by

$$P_N = \frac{V_N^2}{4R} = kTB \tag{8.96}$$

, i.e. the maximum available power is not dependent on the value of the resistor.

As we saw above the thermal noise is directly coupled to the physical temperature of the resistor. However, the temperature term is also used in a more wider and abstract sense to quantify any noise source. We then refer to the effective noise temperature. This will be a fictitious temperature, not necessarily equal to any physical temperature within a given circuitry. Once the noise voltage, current or available noise power is known the effective noise temperature can be calculated from equations (8.94) to (8.96), respectively.

So, the noise figure is a ratio that describes how many times the noise power increases assuming a predefined noise source. As such the noise figure cannot be used directly to calculate the noise power at the output for an arbitrary source noise power simply by multiplying the source noise with the noise figure. For example, if the resistor temperature is $30°\mathrm{K}$ instead of $300°\mathrm{K}$ .the source noise power will be ten times lower whereas the noise contributed by the two-port network itself is still equal to the noise figure times the noise from a given resistor with a temperature of $300°\mathrm{K}$.

Maximum available noise power or equivalent noise temperature is used to specify the noise properties of a source. It is therefore natural to also use this quantity when we investigate the noise behaviour when we connect one or more stages in cascade with the source. That is we would like to be able to calculate the maximum available noise power at the output of the final stage. Consider the cascade of two two-port networks in figure 8.32.

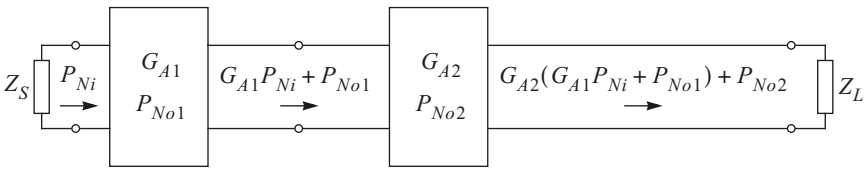


Figure 8.32 Cascade of two noisy two-port networks.

Assuming a $300°\mathrm{K}$ noise source the noise figure for the first stage can be written as

$$F_1 = \frac{P_{No1} + P_{Ni}G_{A1}}{P_{Ni}G_{A1}} \tag{8.97}$$

where P_{No1} is the available output noise generated by the first stage and $P_{Ni}G_{A1}$ the available output noise generated by the source. Exploiting the fact that the available gain is given by

$$G_{A1} = \frac{P_{So1}}{P_{Si1}} \tag{8.98}$$

where P_{So1} is the available output signal power and P_{Si1} the available source signal power, we can rewrite (8.97) as

$$F_1 = \frac{P_{Si1}/P_{Ni}}{P_{So1}/(P_{No1} + P_{Ni}G_{A1})} = \frac{SNR_i}{SNR_{o1}} \tag{8.99}$$

This means that the noise figure can also be defined as the reduction in signal-to-noise ratio from the input to the output side of a two-port network.

Continuing with the cascade circuit in figure 8.32 we have that the total available noise power at the output of the first stage equals $G_{A1}P_{Ni} + P_{No1}$ and similarly the total available noise power at the output of the second stage equals $G_{A2}(G_{A1}P_{Ni} + P_{No1}) + P_{No2}$. From this we can write the total noise figure as

$$\begin{aligned} F_{TOT} &= \frac{\text{total available noise power at the output}}{\text{available noise at the output originating from the source}} \\ &= \frac{G_{A2}(G_{A1}P_{Ni} + P_{No1}) + P_{No2}}{P_{Ni}G_{A1}G_{A2}} \\ &= 1 + \frac{P_{No1}}{P_{Ni}G_{A1}} + \frac{P_{No2}}{P_{Ni}G_{A1}G_{A2}} \end{aligned} \tag{8.100}$$

In the two last terms in this expression there is a ratio of the available noise generated by a two-port network (P_{No1} and P_{No2}, respectively) to the available noise of the source P_{Ni}. This indicates that we can write (8.100) as a function of the noise figures for the individual two-port networks:

$$F_{TOT} = F_1 + \frac{F_2 - 1}{G_{A1}} \tag{8.101}$$

where

$$F_1 = \frac{P_{No1} + P_{Ni}G_{A1}}{P_{Ni}G_{A1}} = 1 + \frac{P_{No1}}{P_{Ni}G_{A1}} \tag{8.102}$$

and

$$F_2 = \frac{P_{No2} + P_{Ni}G_{A2}}{P_{Ni}G_{A2}} = 1 + \frac{P_{No2}}{P_{Ni}G_{A2}} \tag{8.103}$$

In a generalised form equation (8.101) is known as Friis' formula and for N stages the formula can be written as

$$F_{TOT} = F_1 + \sum_{i=2}^{N} \frac{F_i - 1}{\prod_{j=1}^{i-1} G_{Aj}} \tag{8.104}$$

8.7.2 Design Methodology

The noise figure for a two-port network varies with the source impedance. So does the gain of course. Unfortunately, maximum gain is seldom obtained for the same source impedance that gives the minimum noise figure. It can be shown that the noise figure for a two-port can be written as

$$F = F_{min} + \frac{R_N}{G_S}|Y_S - Y_{opt}|^2 \quad ; \quad G_S = \mathrm{Re}[Y_S] \tag{8.105}$$

where F_{min}, R_N and Y_{opt} are parameters for a given two-port network. F_{min} is the minimum noise figure that is obtained when the source admittance Y_S is chosen to be Y_{opt}. A larger noise figure is obtained if $Y_S \neq Y_{opt}$. The equivalent noise resistance R_N determines how much the noise figure increases with the difference between Y_S and Y_{opt}. Equation (8.105) can also be written using normalised quantities:

$$F = F_{min} + \frac{r_N}{g_S}|y_S - y_{opt}|^2 \quad ; \quad g_S = \mathrm{Re}[y_S] \tag{8.106}$$

where $r_N = R_N/Z_0$, $y_S = Y_S/Z_0$ and $y_{opt} = Y_{opt}/Z_0$. From here it easy to express the noise figure in terms of reflections coefficients instead. We know that the relation between a given normalised admittance and its corresponding reflection coefficient is given by $y = (1-\Gamma)/(1+\Gamma)$. Thus (8.105) becomes

$$F = F_{min} + \frac{4r_N|\Gamma_S - \Gamma_{opt}|^2}{(1-|\Gamma_S|^2)|1+\Gamma_{opt}|^2} \tag{8.107}$$

F_{min}, r_N and Γ_{opt} are noise parameters that are usually given by transistor manufacturers. These parameters varies both with frequency and operating point of the device.

None of the equations (8.105) to (8.107) are suitable for design purposes. Since we usually want to design for a specific noise figure rather than for a specific source impedance we want to know what source impedances that satisfies a desired noise figure. Similar to the analysis behind the stability circles and the gain circles we can derive expressions for the set of source impedances that gives a certain noise figure, and of course, these sets are given by circles in the Γ_S-plane:

Radius of noise circle:

$$r_S = \frac{\sqrt{N_i^2 + N_i(1 - |\Gamma_{opt}|^2)}}{1 + N_i} \tag{8.108}$$

Centre of noise circle:

$$\Gamma_{SO} = \frac{\Gamma_{opt}}{1 + N_i} \tag{8.109}$$

where

$$N_i = (F - F_{min})\frac{|1 + \Gamma_{opt}|^2}{4r_N} \tag{8.110}$$

From the discussion above it is obvious that when we have a certain noise figure as our main design goal we will introduce more less mismatch in terms of gain at the input port. Since we want to keep track of the gain as well we can use the design methodologies in case B (unilateral, specific gain) or in case D (bilateral, specific gain) above. We can draw noise circles and gain circles in the same Smith-chart (Γ_S-plane). We will then be able to find intersections between the two that correspond to good compromises between noise and gain.

Example 8.7 Noise properties of a transistor

Draw the 1.5, 2 and 3dB noise circles in a Smith chart for the BFR520 transistor at 1GHz with $I_C = 5\text{mA}$ and $V_{CE} = 3\text{V}$. Also mark the centres for the circles as well as the reflection coefficient that gives the minimum noise figure. The following noise parameters are provided by the manufacturer:

$F_{min} = 1.15\text{dB} = 1.30$, $\Gamma_{opt} = 0.336\angle 49.0°$, $r_N = 0.250$

The noise circles are calculated using equations (8.108) to (8.110):

Noise circle for $F = 1.5\text{dB}$:
centre: $\Gamma_{SO} = 0.1884 + j0.2167 = 0.2872\angle 49.0°$
radius: $r_S = 0.3623$

Noise circle for $F = 2.0\text{dB}$:
centre: $\Gamma_{SO} = 0.1533 + j0.1764 = 0.2337\angle 49.0°$
radius: $r_S = 0.5297$

Noise circle for $F = 3.0\text{dB}$:

centre: $\Gamma_{SO} = 0.1062 + j0.1222 = 0.1619\angle 49.0°$

radius: $r_S = 0.7000$

The circles are shown in the Smith-chart below. Note that the centres of all circles has the same argument, i.e., they lie on the same radial line as shown below.

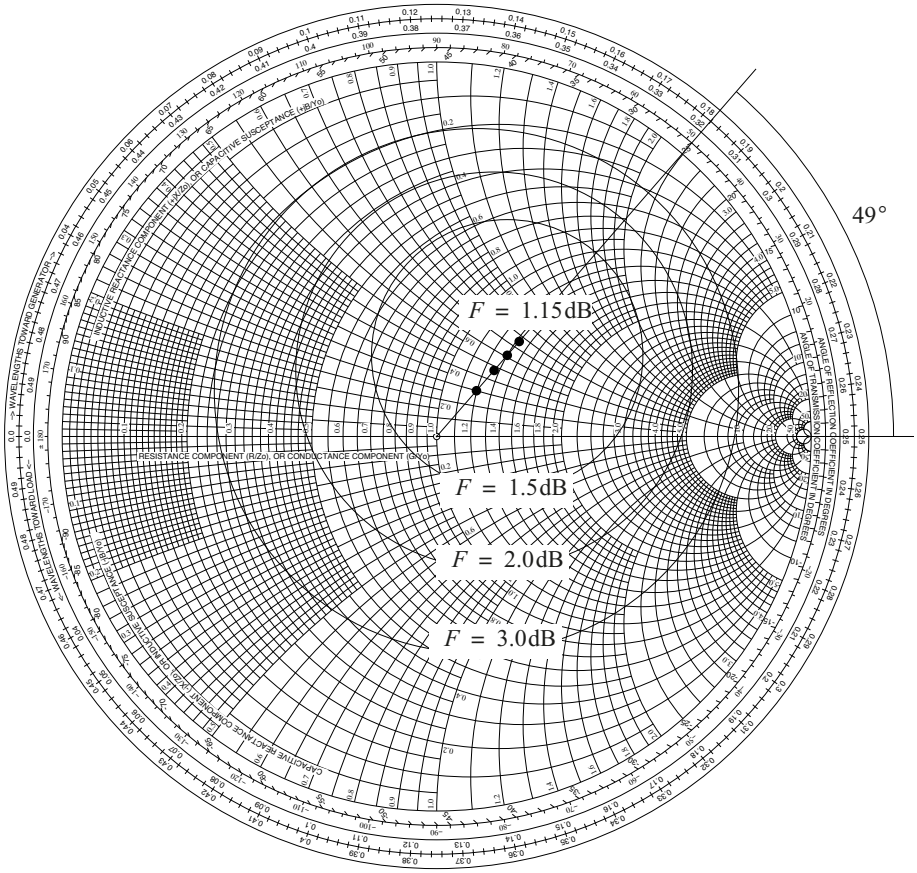


Figure 8.33 Reflection coefficient for optimal noise and noise circles for various noise figures.

8.8 Complete Design Example

This chapter ends with a design example that considers stability analysis, designing for specific gain and noise, broadband analysis and realisation of matching networks.

Example 8.8 Design of low noise amplifier

Design an amplifier with at least 18dB gain and a noise figure no higher than 1.3dB. Use the BFR520 transistor at 1GHz with $I_C = 5\text{mA}$ and $V_{CE} = 3\text{V}$. The terminating source and load are assumed to be purely resistive 50Ω, see figure 8.34 below. Design the matching networks using transmission line structures assuming lossless transmission lines. Also assume that we cannot consider the transistor to be unilateral.

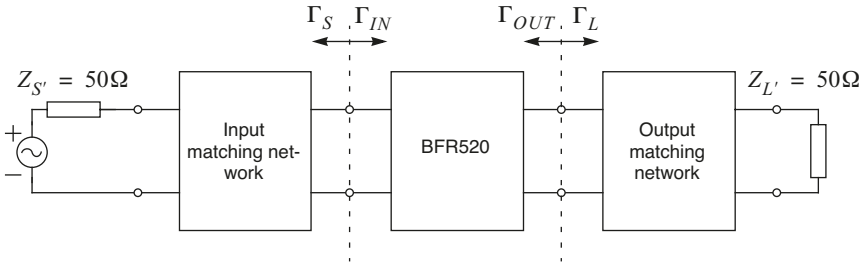


Figure 8.34 Amplifier topology to be designed for low noise.

The following noise parameters are provided by the manufacturer:

$S_{11} = 0.473\angle{-84.8°}$ $\quad S_{12} = 0.0760\angle{59.6°}$

$S_{21} = 7.1790\angle{112.0°}$ $\quad S_{22} = 0.5950\angle{-34.7°}$

$F_{min} = 0.90\text{dB} = 1.23$, $\Gamma_{opt} = 0.400\angle{26.0°}$, $r_N = 0.250$

The fact that the noise figure should be less than 1.3dB suggests that we must cater for mismatch at the input since the source reflection coefficient affects the noise performance of the amplifier. This implies that we should design with available gain, i.e., mismatch is introduced at the input to comply with the noise and the gain specification and then the output is conjugately matched so that the transducer gain will be equal to the available gain.

The first thing to do is to check the stability properties of the transistor. We get $|\Delta| = 0.5161$ and $K = 0.6310$.

Since $K < 1$ the transistor is only conditionally stable and thus we should calculate the maximum stable gain to ensure that the transistor is capable of providing the gain that was specified above

$$G_{MSG} = \left|\frac{S_{21}}{S_{12}}\right| = 94.5 = 19.75\text{dB}$$

This means that there is some headroom from the 18dB that we want. The next step is to draw the stability circles using equations (8.49) to (8.52):

Input stability circle:

centre: $\Gamma_{SO} = 6.181 - j10.518 = 12.200\angle{-59.6°}$

radius: $r_S = 12.806$

Output stability circle:

centre: $\Gamma_{LO} = 3.638 + j5.859 = 6.897\angle 58.2°$

radius: $r_L = 6.222$

The stability circles are shown in the Smith-chart below. Since $|S_{11}| < 1$ and $|S_{22}| < 1$ we know that the centre of the Smith-chart is stable both for the input and the output.

There is one additional check that we must perform before we continue with the design. We must verify that we can comply with the gain and noise requirements at the same time using the same source reflection coefficient. Thus, we should calculate the noise circle for a noise figure of 1.3dB and the available gain circle for 18dB gain. If no part of the noise circle is within the gain circle we will not be able to fulfil the specification. In practice this means that we should choose another transistor with better performance or loose the specification if possible. The noise and gain circles become:

Noise circle for $F = 1.3\,\text{dB}$: (equations (8.108) to (8.110))

centre: $\Gamma_{SO} = 0.2940 + j0.1434 = 0.3271\angle 26.0°$

radius: $r_S = 0.3981$

Available gain circle for $G_A = 18\,\text{dB}$:

($g_A = 1.2243$, equations (8.91) and (8.92))

centre: $\Gamma_{LO} = -0.3401 + j0.5788 = 0.6713\angle 120.4°$

radius: $r_L = 0.8194$

As can be seen from the Smith-chart below the 1.3dB noise circle is partly within the 18dB gain circle which means that we can continue with the specified transistor.

We are now at a stage where we can select a point for Γ_S such that the specification will be fulfilled. This example does not specify whether it is the noise or the gain that is the most important parameter so we can select the point arbitrarily within the intersection of the two circles. In this case we can actually select Γ_{opt} which gives minimum noise, i.e.,

$\Gamma_S = \Gamma_{opt} = 0.400\angle 26.0°$ or $z_S = 1.9049 + j0.7953$

With this reflection coefficient the available gain and the noise figure become $G_A = 18.1\,\text{dB}$ and $F = F_{min} = 0.9\,\text{dB}$, respectively.

Now we should verify that the output can be conjugately matched to obtain a transducer gain equal to the available gain that we headed for above. If this is not the case we can tolerate some degree of mismatch to obtain at least 18dB gain. The output reflection coefficient becomes $\Gamma_{out} = 0.4478\angle -55.6°$ and thus the load reflection coefficient should be $\Gamma_L = 0.4478\angle 55.6°$ for conjugate match and as can be seen in the Smith-chart this point is in the stable region for Γ_L but we also note that it is not far away for the stability circle. A practical design may possibly become too sensitive under these conditions. However, more information is required to analyse the stability under worst-case conditions. In this example we stick to the value that was calculated above.

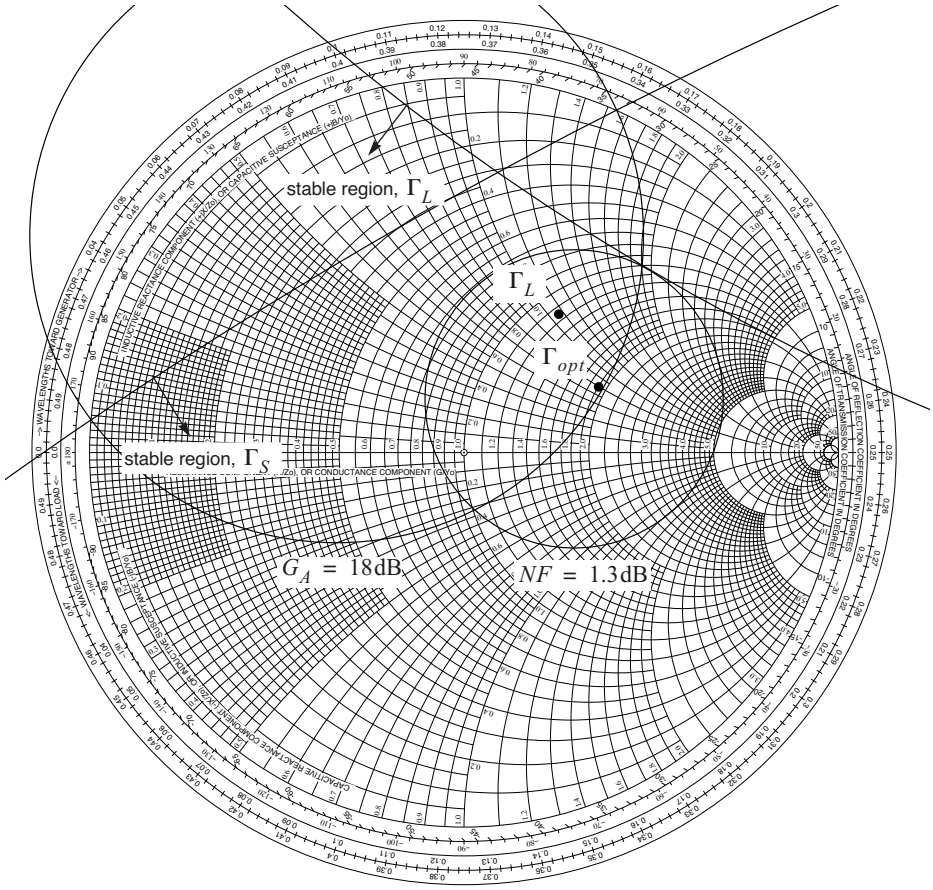


Figure 8.35 Compromise between high gain and low noise.

The reflection coefficients for the source and the load are now determined and the realisation of the matching networks remains.

The topology chosen for the matching networks is shown in the figure below. The transistor is connected to 50Ω series transmission lines that moves the admittance to the unity conductance circle. Short-circuited 50Ω stubs are used to eliminate the remaining susceptance. The dimensioning of the transmission lines are shown in the Smith-chart below.

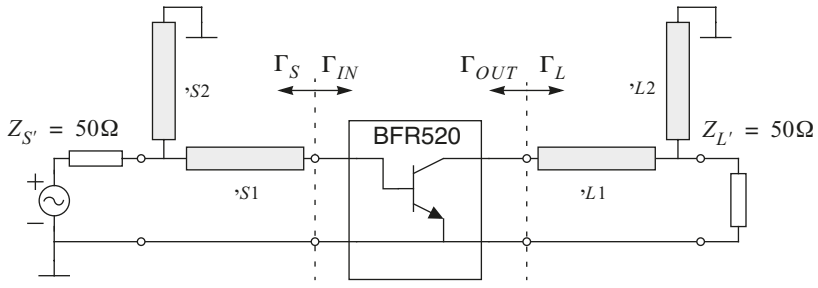


Figure 8.36 Matching networks for the amplifier.

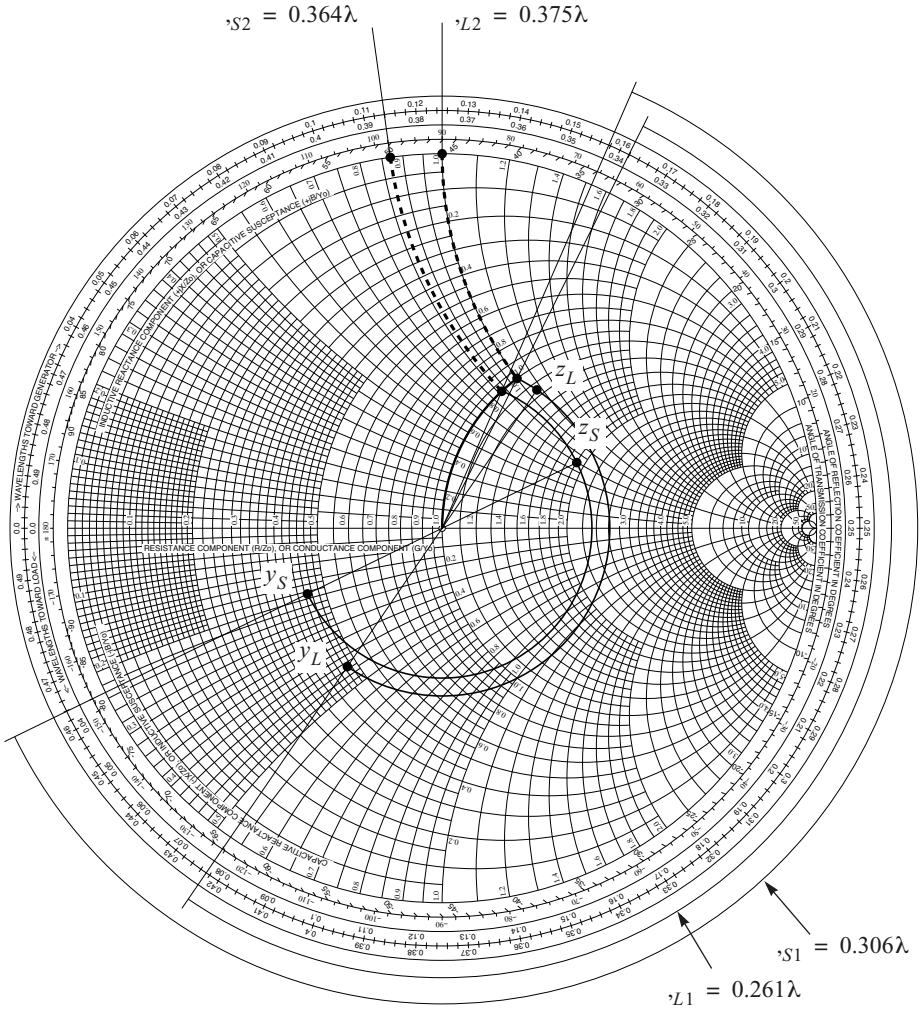


Figure 8.37 Design of matching networks.

For the physical realisation of the transmission lines with respect to geometry and material please refer to earlier chapters on these topics.

This design was carried out for one frequency only (1GHz). As matter of fact, we do not no much about the behaviour of this amplifier at other frequencies than in the neighbourhood of 1GHz. With access to computer based mathematical tools or simulation tools it is possible to handle the whole set of S parameters for different frequencies in one shot and recalculate the effect of the matching network at these frequencies. Thus, we can analyse but not synthesize the behaviour of the amplifier at other frequencies. We can for example plot the gain as a function of frequency to ensure enough suppression of unwanted signals at frequencies other than 1GHz. We can also characterise stability properties by calculating the input and output reflection coefficients for the transistor at different frequencies. We could go even further by applying the inverse Fourier transform on the complex-valued gain transfer function to obtain the impulse response. This would provide us with

information on the behaviour of the amplifier in the time domain. A slowly decaying and oscillating impulse response indicates stability problems.

Below the calculated gain is shown as a function of frequency. The 21 points of S parameter data available from the manufacturer were used to calculate the gain. The wanted gain of 18dB is indeed achieved at 500MHz but over a broader frequency range the gain varies between -26 to +25dB.

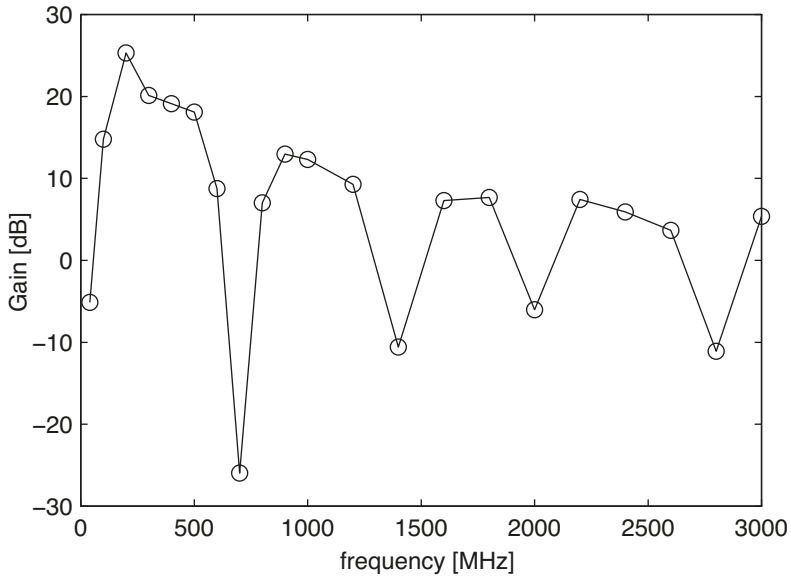


Figure 8.38 Gain as a function of frequency.

The input and output reflection coefficients have also been calculated and the result in shown in figure 8.39. Here, it is readily seen that we have a stability problem at lower frequencies. The input reflection coefficient is close to unity at lower frequencies and even worse, the output reflection coefficient larger than unity.

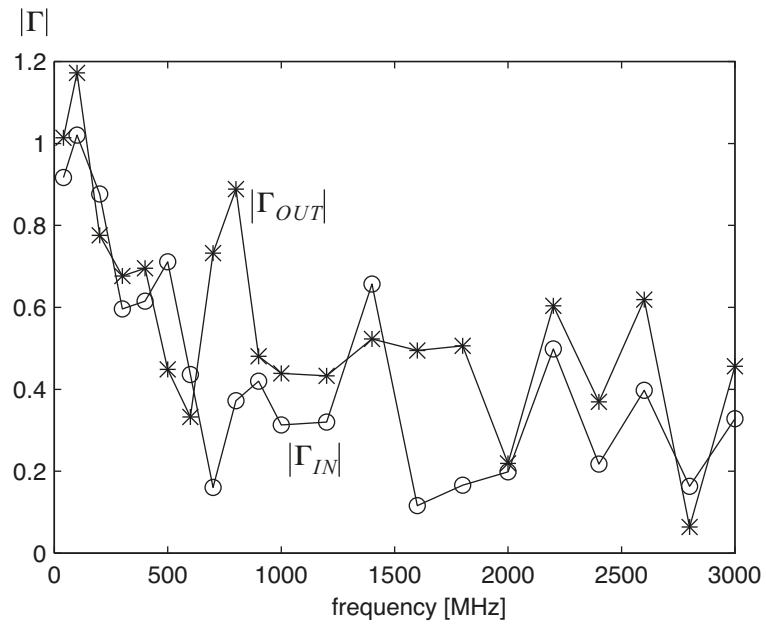


Figure 8.39 Input and output reflection coefficients.

Even though it is not a desirable situation the amplifier might still work. We should therefore investigate the loop impedance. The impedances corresponding to Γ_{OUT} and Γ_L can be calculated and added together to obtain the total loop impedance. The result is illustrated below. The criterion for oscillation is not fulfilled but a negative loop resistance is observed at lower frequencies. At these frequencies the stub at the output port will act more or less as a short-circuit to ground and the total load resistance will not be large enough to give a positive loop resistance.

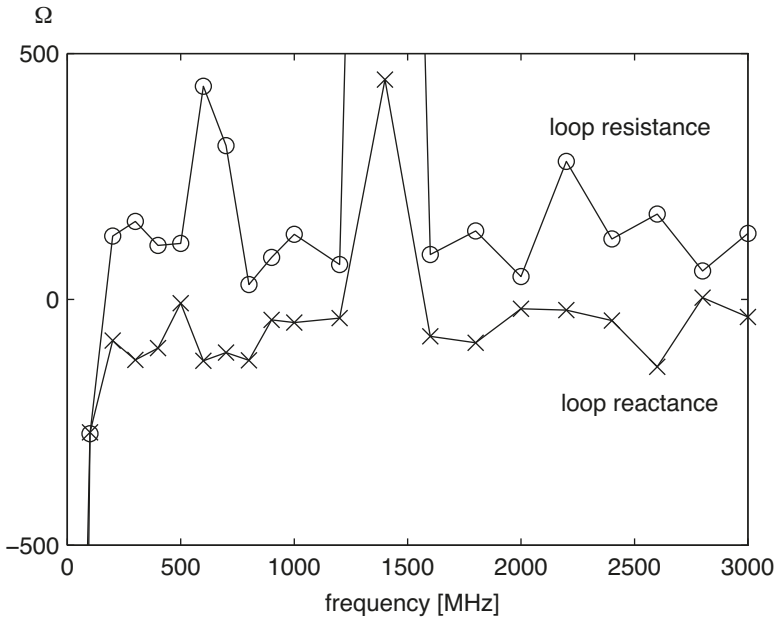


Figure 8.40 Loop resistance and reactance as a function of frequency.

8.9 References

[1] G. Gonzalez, *Microwave transistor amplifiers - analysis and design*, Prentice Hall, 1984.

Chapter 9

Transistor Biasing

The biasing circuitry forces a device into a desired operating point in terms of DC terminal voltages and currents. It is a very important part of the design because the desired properties of a device, given as S parameters for instance, can only be kept if the biasing circuitry can maintain the operating point independent of temperature and aging. However, one should remember that far from all properties of a device become constant just because the operating point is kept steady. Temperature and aging will still have some influence on the behaviour although the effects will be moderate.

In previous chapters only small-signal properties of transistors were considered either in terms of hybrid-π models (or the FET equivalent) or complex-valued parameter sets, i.e., z parameters, S parameters etc. Such a small-signal description assumes a certain operating point and temperature. Once the parameter set has been chosen matching networks can be designed to obtain the desired gain, noise figure etc. After that, the biasing circuitry is designed to obtain the desired voltages (typically V_{CE} or V_{DS}) and currents (typically I_C or I_D) associated with the parameter set chosen. However, the biasing circuitry must be designed in such a way that it does not change gain, noise figure and stability of the design. If it is anticipated that it will have some impact it must be taken into account from the very beginning when designing the matching networks.

A biasing circuitry around a transistor consists of a few resistors in its most simple form but complex designs based on transistors or operational amplifiers for active biasing might prove to be necessary in some cases. This chapter deals with both.

BJTs and FETs are very different with respect to biasing. BJTs are very sensitive to variations in temperature and FETs suffer from inaccurate parameters. Therefore, both require a more or less advanced biasing circuitry to compensate for temperature drift or inaccuracy. Most of this chapter is devoted to BJTs because they can be used to illustrate a larger set of biasing techniques compared with FETs. However, some biasing solutions for BJTs can be transferred to the FET case and a brief discussion on FET biasing is also given. Finally, means of isolating the biasing circuitry from the signal circuitry is exemplified.

9.1 Biasing of Bipolar Transistors

In principle, there are two ways to control the operating point of a bipolar device, either by applying a voltage across the base-emitter or by feeding a current into the base. These two options are discussed below.

9.1.1 Base-Emitter Voltage Drive

A bipolar transistor can approximately be modelled by the well-known exponential relationship between the base-emitter junction voltage V_{BE} and the collector current I_C that is given by [1]

$$I_C = I_S(e^{V_{BE}/V_T} - 1) \approx I_S e^{V_{BE}/V_T} \tag{9.1}$$

where $V_T = kT/q$ is the thermal voltage and I_S the saturation current. The latter one depends on the device properties and has roughly an exponential growth with the temperature [2].

As mentioned above it is the collector current and the collector-emitter voltage that constitutes the operating point that must be kept steady. For example, the collector current controls the transconductance of the device as well as the input and output impedances and the collector-emitter voltage controls the junction capacitance between the base and the collector.

Now, (9.1) indicates that we should apply a fixed voltage across the base-emitter of the device to get the desired collector current as shown in figure 9.1 - the force and the action. However, this is not such a good idea for a few reasons. To begin with, the saturation current I_S varies from device to device and even worse, it is strongly dependent on the temperature. Finally, with a fixed V_{BE} we will also suffer from the temperature dependency of V_T, see equation (9.1). An increase of one degree Celsius results in roughly 10% lower collector current due to the influence of V_T.

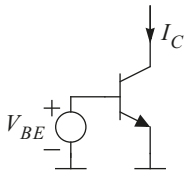


Figure 9.1 Setting the operating point with a fixed base-emitter DC voltage.

It is out of the scope in this chapter to deal with the temperature dependency of I_S in detail. However, a typical characteristic for I_S is illustrated in figure 9.2 where $I_S = 1\times10^{-16}$A at $T = 300°$K. Over the given range from $-20°$C to $80°$C the saturation current changes with a factor of more than 1 million. As a rule of thumb the saturation current increases with about 10% per degree Celsius.

As mentioned earlier it is not only the saturation current that controls the temperature dependency of the collector current. The thermal voltage found in equation (9.1) to some extent counteracts the temperature dependency of the

saturation current. If we combine the result in figure 9.2 and equation (9.1) we can plot the collector current as a function of temperature, see figure 9.3. Again we have chosen $I_S = 1{\times}10^{-16}\text{A}$ at $T = 300°\text{K}$ and also $V_{BE} = 0.7745\text{V}$ that gives a collector current equal to 1mA at $T = 300°\text{K}$ Compared with the saturation current the temperature dependency of the collector current is more relaxed but still varies with a factor of 100 or so over the given range.

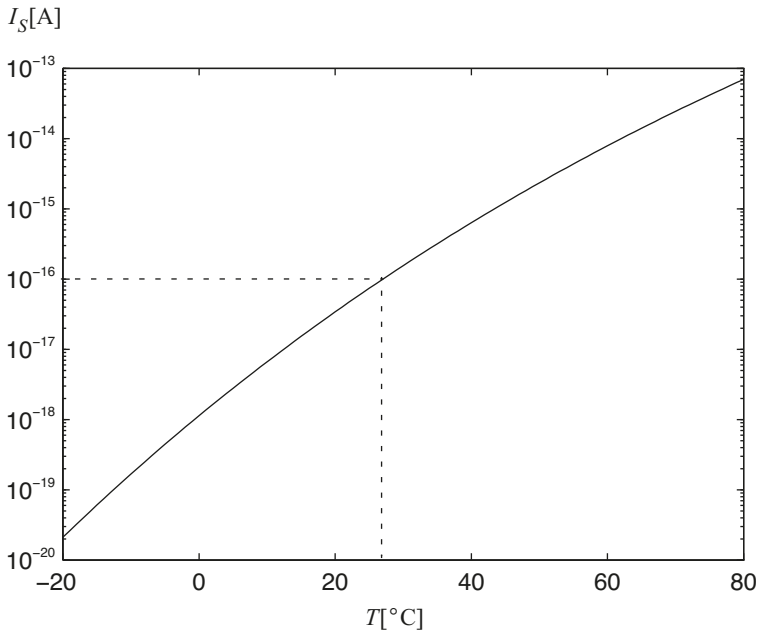


Figure 9.2 Typical characteristic of saturation current as a function of temperature.

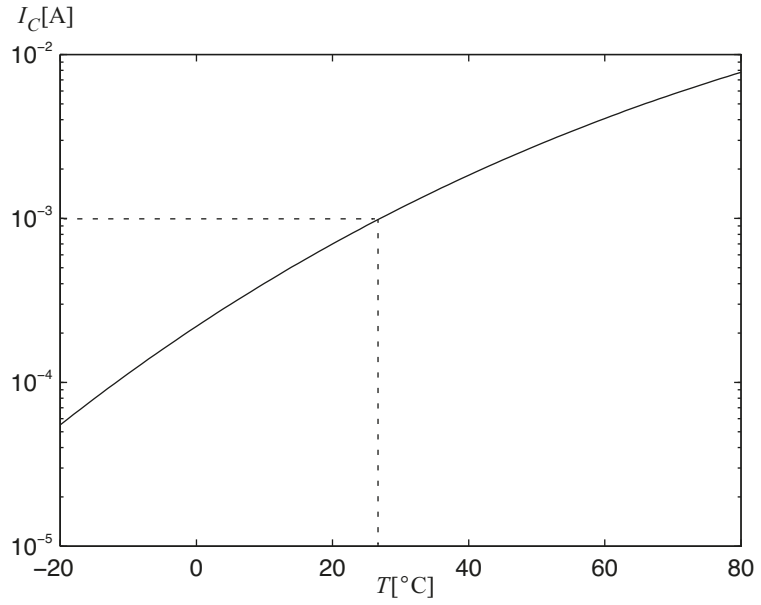


Figure 9.3 Typical characteristic of collector current as a function of temperature with a fixed base-emitter voltage.

It readily seen that it is not such a good idea to use a fixed V_{BE} to obtain a certain operating point. Instead, we should try to measure the collector current and adjust V_{BE} accordingly to obtain the desired operating point. In other words, accurate biasing is obtained by means of feedback.

9.1.2 Base Current Drive

The second approach to obtain a desired operating point is to force a current into the base as illustrated in figure 9.4. This is quite different from the voltage driven configuration presented in the previous section. Of course, equation (9.1) still holds but the base-emitter voltage is of secondary importance. Instead, the primary relation is given by the ratio between the collector DC current and the base DC current, i.e., the DC current gain β_0. As a first order approximation β_0 is typically assumed to be constant although is varies quite significantly from one device to another, say from 100 to 200 for an NPN device. A more elaborate model of the current gain shows that it is dependent not only on the temperature but also on the collector current. Again, it is out of the scope to investigate these relations in detail here but a typical temperature coefficient is +0.7% per degree Celsius.

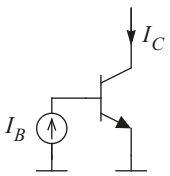


Figure 9.4 Setting the operating point with a fixed base current.

In figure 9.5 a typical characteristic of the temperature dependency of the current gain is shown. The current gain is equal to 100 at $T = 300°\mathrm{K}$. Over the given temperature range the current gain varies from about 70 to 140 which is much less than the what we saw for the saturation current and the collector current with a fixed base-emitter voltage. In the event that we would like to reduce the temperature dependency by means of feedback as discussed above we conclude that a base current control can do with more relaxed requirements on the feedback circuit. One final remark should be given on the base current control. Even though some applications could do with the moderate variations in current gain as shown in figure 9.5 it is still recommended that feedback is applied or else thermal runaway might destroy the device. This is due to the positive feedback in the bipolar transistor itself. Let us assume that we have a very small collector resistor connected to the supply voltage. This means that a change in collector current will result in negligible change in collector voltage which can be considered to be constant. Then, if the collector current increases, then the temperature will increase which will increase the current even further and so on. Thus, the power dissipation will increase accordingly and if the collector resistor is not large enough and there is no feedback to counteract this behaviour the device might be destroyed.

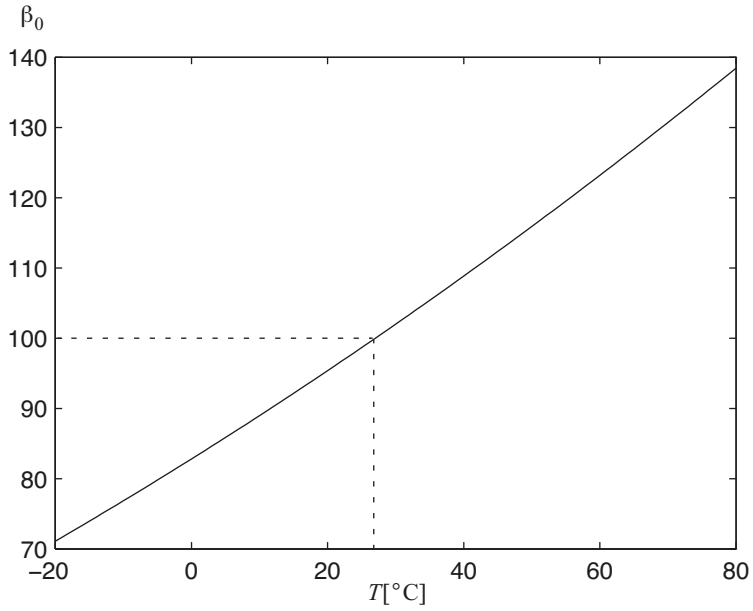


Figure 9.5 Typical characteristic of current gain as a function of temperature with a fixed base current.

9.2 Passive Biasing Circuits

A reliable biasing circuit is based on feedback where the collector current is measured to control either the base-emitter voltage or the base current. To measure the collector current we can put a resistor in series with either the collector or the emitter. By doing so the voltage over the resistor will be a measure of the current. This voltage can be fed back to the base and a few circuit solutions are shown in figure 9.6.

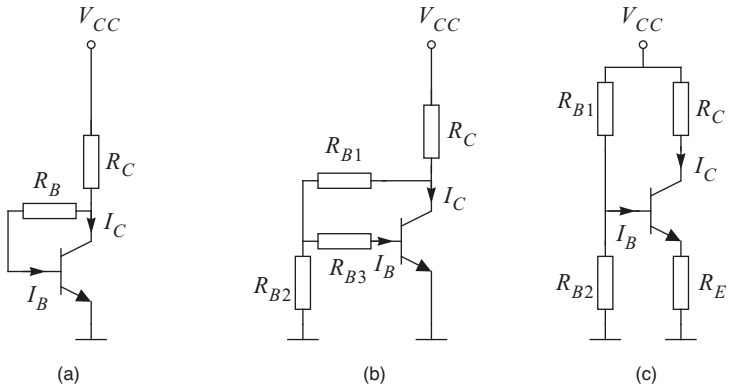


Figure 9.6 Various passive biasing circuits.

It is easy to realise how these circuits work qualitatively. Any increase in collector current will be counteracted by a reduction in base-emitter voltage or base current and consequently collector current. To realise how effective these compensation techniques are some insight into feedback systems is required. Basically, the suppression of an error, in this case the temperature dependency of the collector current, is reduced roughly by an amount equal to

the loop gain of the feedback system. This topic as well as other design considerations will be dealt with below by investigating three design examples, one for each circuit topology. The following specification is given:

- Wanted collector current: $I_C = 1\text{mA}$
- Wanted collector-emitter voltage: $V_{CE} = 3\text{V}$
- Available supply voltage: $V_{CC} = 12\text{V}$
- Transistor DC current gain: $\beta_0 = 100$ to 140

Example 9.1 Current Driven Biasing I

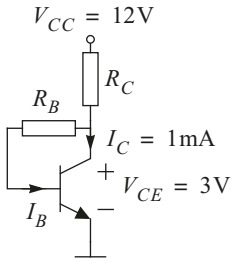


Figure 9.7 Simple passive biasing of BJT using shunt feedback.

Since both the supply voltage and the collector voltage is specified the voltage across R_C is also fixed. If the base current is negligible (which is true in this case since the current gain is high) the current through R_C is equal to the collector current. Thus, there is only one possible value for R_C, namely

$$R_C = (12-3)/1\times10^{-3} = 9\text{k}\Omega$$

The voltage across the second resistor between the collector and the base is also fairly well defined. The collector voltage is 3V and we can assume that the voltage at the base is in the neighbourhood of 0.7V or so. To be able to determine the resistor value we need an estimate of the current, the base current to be precise. The base current is given by $I_B = I_C/\beta_0$. Unfortunately, the current gain varies from device to device from 100 to 140 as specified above. If we assume a current gain of 100, R_B becomes

$$R_B = V_{CB}/I_B = (3-0.7)/(1\times10^{-3}/100) = 230\text{k}\Omega$$

If the current gain is 140 instead of 100, R_B will be 40% larger. If we design R_B with the "wrong" current gain it can be viewed as an error in collector current just like the effect of temperature. The feedback circuit will compensate for this error just like any other error in collector current.

To investigate the dependency of temperature on collector current and collector voltage we should write the large-signal relationships for these quantities. But before we do that we should investigate whether this configuration is base current driven or base-emitter voltage driven. When the collector current varies the collector voltage will vary accordingly due to R_C whereas the base-emitter voltage will be more or less constant, i.e., in the neighbourhood of 0.7V.

Thus, R_B acts as a voltage-to-current converter because the current through R_B, i.e. the base current, will vary proportionally to the collector voltage. We conclude that this is a base current driven configuration.

The collector voltage is given by $V_C = V_{CC} - I_C R_C$ and the base current is given by $I_B = I_C/\beta_0 = (V_C - 0.7)/R_B$ Combined, we get an expression for the collector voltage as a function of the current gain that is dependent on temperature,

$$V_C = V_{CC} - \beta_0 R_C (V_C - 0.7)/R_B$$

or as an explicit expression for the collector voltage,

$$V_C = \frac{V_{CC} + 0.7\frac{\beta_0 R_C}{R_B}}{1 + \frac{\beta_0 R_C}{R_B}}$$

An initiated reader will identify this as a typical expression for feedback systems where we find the loop gain to be equal to $\beta_0 R_C/R_B$. To be useful, the loop gain has to be large but at the same time we know from the discussion above that we do not really have any freedom in choosing arbitrary values on β_0, R_C and R_B. They are all tightly connected to each other once the collector voltage, collector current, supply voltage and current gain have been determined. Anyway, in this example the loop gain becomes

$$\beta_0 R_C/R_B = 100 \cdot 9\times10^3/230\times10^3 = 3.9$$

Thus, the improvement in operating point stability should be in this range. Below, the collector current is plotted as a function of temperature based on the expression for collector voltage above and the temperature dependency on the current gain in figure 9.5. Also, the collector current influenced by temperature through β_0 but without feedback is shown for comparison. It can be seen that the improvement is close to the loop gain. We also note a small discrepancy from the anticipated collector current at $T = 300°\text{K}$. Is is mainly due to the fact that we have neglected the base current through R_C and that we have assumed 0.7V base-emitter voltage.

The loop gain can be derived by inspection of the circuit schematic. Assume that we have an error in the collector current, ΔI_C. This is transformed into a voltage by R_C, $\Delta V_C = -\Delta I_C \cdot R_C$ (we assume that R_C is much smaller than R_B). Now, R_B more or less acts as a voltage-to-current converter so the collector voltage ΔV_C is transformed into a base current, $\Delta I_B = -\Delta V_C/R_B$ and finally the base current is amplified with the current gain, β_0 in the transistor. Thus, all together the loop gain becomes $\beta_0 R_C/R_B$. Note that the loop gain becomes higher for higher temperatures since β_0 increases with temperature.

This design was based on a specific current gain, $\beta_0 = 100$. However, anything between 100 and 140 could be expected according to the specification. The effect of having a current gain of 140 instead of 100 is equivalent to the effect if the temperature would alter the current gain with the same amount. Thus, we can consult figure 9.5

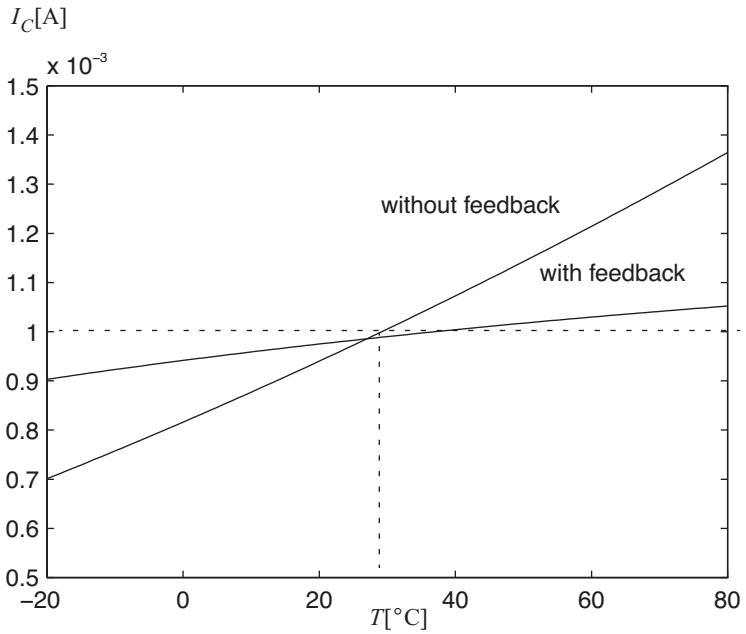


Figure 9.8 Temperature dependency of collector current with and without feedback.

where we see that at 80°C the current gain happens to be around 140. The actual collector current with feedback at this temperature is around 1.06mA. Without feedback the collector current would become 1.4mA. Thus, if the current gain would be equal to 140 at $T = 300$°K instead of 100 the collector current would be around 1.06mA. Obviously, the optimal value for R_B should be calculated from a current gain in the middle of the two extremes. In this example 120 would be a good choice.

We conclude that we do not have much freedom in choosing the resistor values for the two-resistor biasing circuit. Both R_C and R_B are fully determined by four parameters; the supply voltage, the collector voltage, the collector current and the current gain. If higher loop gain is required at least one of these quantities must be altered. For example, we could increase V_{CC} or reduce V_C

Example 9.2 Current Driven Biasing II

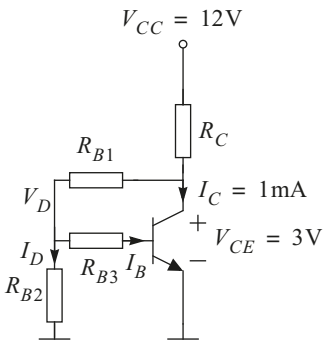


Figure 9.9 Passive biasing of BJT using shunt feedback.

The first topology relied on a given current gain and thus a base current that produced a desired voltage drop from the collector to the base. Clearly, such a solution quite strongly depends on the current gain. This problem is solved with this topology that has a voltage divider from collector to base. The current through the voltage divider must be much larger than the base current so as to not introduce any influence from a poorly defined current gain. The base resistor has the same function as in the previous circuit, i.e., to act as an voltage-to-current converter. There is an option to leave it out completely in which case we can consider the transistor to be base-emitter voltage driven rather than base current driven. This of course requires that the base does not load the voltage divider significantly.

To determine the collector resistor R_C we must consider the influence of the voltage divider, R_{B1} and R_{B2} .The voltage divider current cannot be arbitrarily small because it has to supply a base current that in turn must not load the voltage divider. This means that there is a limited headroom for choosing the current in the voltage divider. A practical approach to the problem is to partition the current gain of the transistor as $I_D = \sqrt{\beta_0} \cdot I_B$ and $I_C = \sqrt{\beta_0} \cdot I_D$. So, if β_0 is 100 then $I_D = 10 \cdot I_B$ and $I_C = 10 \cdot I_D$.

Now we can determine the value for R_C, namely

$$R_C = (V_{CC} - V_C)/(I_C + I_D + I_B) = (12 - 3)/1.11\times10^{-3} = 8.11\text{k}\Omega$$

To determine R_{B1} and R_{B2} we must know the desired value V_D. If it is chosen to be close to V_C the circuit will collapse to a circuit similar to the first biasing circuit. On the other hand if V_D is chosen to be very close to the base-emitter voltage the transistor will be voltage driven instead of current driven. Thus, to gain anything from this solution we should chose a voltage over R_{B3} that is much less than $V_C - V_B$ but not to close to V_B. It is out of the scope to investigate what is optimal in detail. Instead, as a rule of thumb it is appropriate to let V_D become 10% to 20% of V_C but not less than $2V_B$. In this design example the latter limit applies and thus we should chose $V_D = 1.4\text{V}$. So, now we know everything to determine R_{B1}, R_{B2} and R_{B3}:

$$R_{B1} = (V_C - V_D)/(I_D + I_B) = 1.6/110\times10^{-6} = 14.55\text{k}\Omega$$

$$R_{B2} = V_D/I_D = 1.4/100\times10^{-6} = 14\text{k}\Omega$$

$$R_{B3} = (V_D - V_B)/I_B = 0.7/10\times10^{-6} = 70\text{k}\Omega$$

The first biasing circuit was considered as a purely base current driven solution. The objective of the current circuit is also to have it base current driven but since the voltage over R_{B3} so small we cannot consider this design as purely current driven, i.e., we must be more careful in our assumption about the circuit. Therefore, to be accurate we should include equation (9.1) and take into account the temperature dependency of I_S, V_T and β_0. The analysis becomes quite tedious and we leave it out for the sake of clarity and only present the results, see figure 9.10.

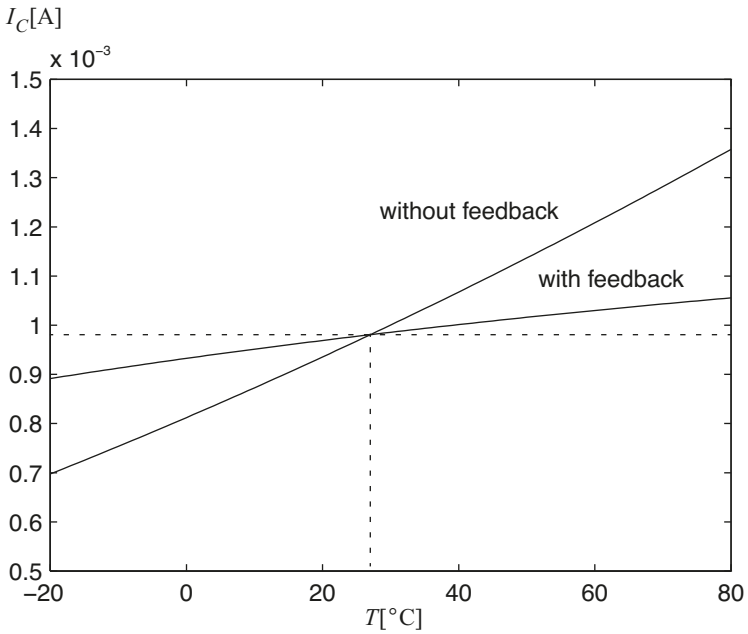


Figure 9.10 Temperature dependency of collector current with and without feedback.

By inspection we find that the loop gain is equal to

$$\frac{\beta_0 R_C}{R_{B3}} \cdot \frac{R_{B2}}{R_{B1} + R_{B2}}$$

where we have assumed that $I_C \gg I_D$ and $I_D \gg I_B$. Using the numbers in this exercise the loop gain becomes 5.84, i.e., somewhat higher than for the first biasing solution. Even though the loop gain is higher the result does not differ much from the first circuit. It can be proven that the circuit cannot be considered as purely current driven and thus there will be a major influence of I_S and V_T which will degrade the performance. Instead, the main advantage with this biasing circuit is that is gives lower resistor values which makes it suitable for thin- and thick-film implementation.

The reader is encouraged to investigate this circuit as a voltage driven configuration with R_{B3} short-circuited.

Example 9.3 Voltage Driven Biasing

The third topology, shown below, is a classical configuration that is very popular in low-frequency applications. As far as microwave applications is concerned, is it appropriate with a warning. For higher frequencies this circuit have stability problems. The emitter resistor is always bypassed for higher frequencies using a capacitor from emitter to ground to avoid waste of gain (these issues will be discussed in section 9.5). This approach introduces a parasitic inductor in series with the emitter that may have a serious impact on stability properties. Otherwise, as we will see, as a biasing circuit it is excellent.

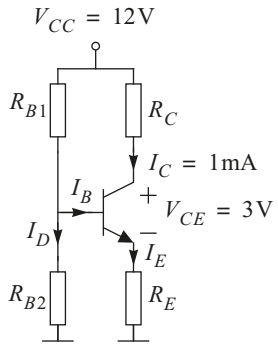


Figure 9.11 Passive biasing of BJT using series feedback.

In low-frequency applications the collector resistor R_C affects the voltage gain of the amplifier. As far as biasing is concerned it is not needed at all, it has no influence on sensitivity in the biasing circuit. The collector could just as well be connected directly to the supply voltage. This is exactly what we can do for high frequency applications because the collector can be isolated, at high frequencies, from the power supply line using for example an inductor (see section 9.5). So, from now on we choose

$R_C = 0$.

The purpose of the voltage divider. R_{B1} and R_{B2}, is to provide a well defined voltage for the transistor base. There is only one important design criteria related to the voltage divider - $I_D \gg I_B$. We conclude that the transistor is base-emitter voltage driven in this configuration.

Finally, we have the emitter resistor, R_E, which happens to be the key component in this circuit. The emitter resistor provides current-to-voltage feedback. The collector current is converted to a voltage and fed back to the base-emitter. The loop gain is simply given by $g_m R_E$.

Since $R_C = 0$ we have

$$R_E = (V_{CC} - V_{CE})/I_E = (V_{CC} - V_{CE})/(I_C(1 + 1/\beta_0)) = 8.91\text{k}\Omega$$

For simplicity we choose $I_D = 1\text{mA} \gg I_B$ and thus

$$R_{B1} = (V_{CC} - V_B)/I_D = (V_{CC} - V_{BE} - I_E R_E)/I_D =$$

$$= (12 - 0.7 - 9)/1\times10^{-3} = 2.3\text{k}\Omega$$

$$R_{B2} = V_B/I_D = (V_{BE} + I_E R_E)/I_D = 9.7\text{k}\Omega$$

The temperature sensitivity of this circuit is shown in the plot below. Again, we do not present the analysis for the sake of clarity. The transistor has been assumed to be purely voltage driven and therefore the temperature dependency of I_S, V_T and β_0 has been taken into account.

It is clear from the plot that this circuit provides the best biasing stability even though it is voltage driven which means that this result should be compared with figure 9.3. Evidently, the loop gain is very

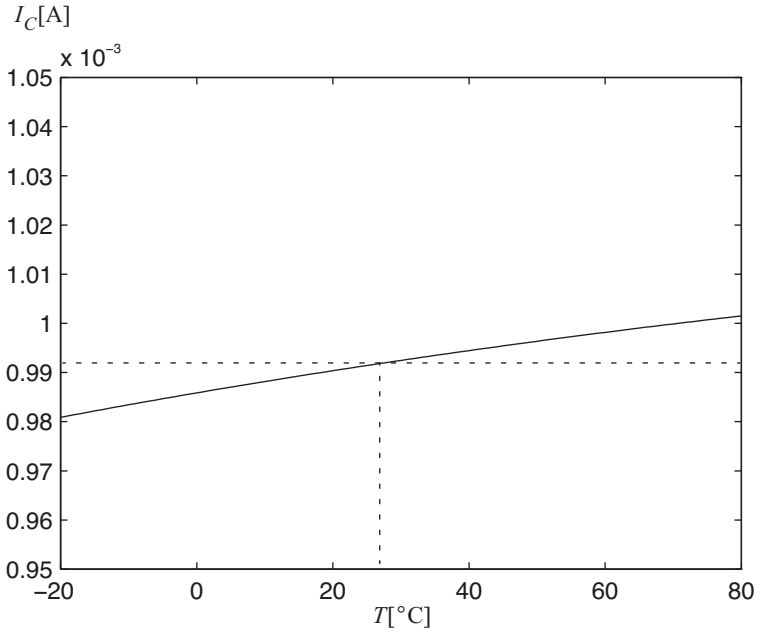
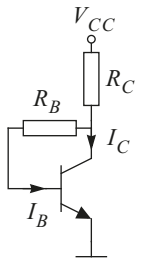


Figure 9.12 Temperature dependency of collector current.

high; $g_m R_E \approx 340$. There is an offset from the 1mA that was anticipated for $T = 300°\text{K}$. The cause is that we assumed 0.7V base-emitter voltage which is not the actual value.

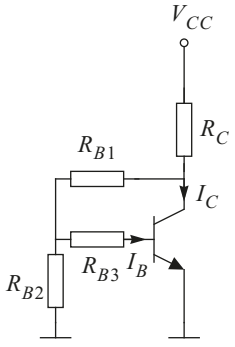
Below, we summarise the properties of the three different biasing schemes.

Properties of current driven biasing I



- Moderate bias stability
- Sensitive to variations in current gain
- Requires high resistor values, not good for thin- and thick-film implementation
- Loop gain given by $\beta_0 R_C / R_B$
- High loop gain obtained if V_{CC}/V_C large

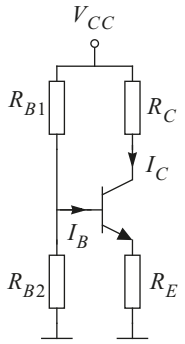
Properties of current driven biasing II



- Moderate bias stability
- Less sensitive to current gain compared with first scheme
- Not purely current driven
- Does not require high resistor values
- Loop gain given by $\dfrac{\beta_0 R_C}{R_{B3}} \cdot \dfrac{R_{B2}}{R_{B1} + R_{B2}}$
- High loop gain obtained if V_{CC}/V_C large

Properties of voltage driven biasing

- Excellent bias stability
- Not sensitive to current gain
- Loop gain given by $g_m R_E$
- R_C can be replaced with an RFC, see section 9.5
- Not suitable for higher frequencies, say in the GHz-range due to stability problems



9.3 Active Biasing Circuits

In cases where passive biasing networks does not give the desired accuracy it is possible to increase the loop gain with an amplifier in the feedback path and thereby reduce the errors even further.

There are two principles for active biasing - with direct and indirect feedback of some quantity related to the operating point. Direct feedback is equivalent to the passive solutions above where the collector current is measured by means of a component in series with the current whereas indirect feedback means estimating the operating point from other measures like the temperature of the device.

A simple design for active biasing with direct feedback is shown in figure 9.13. This circuit also has a resistor in series with the collector as was the case for the passive biasing circuits described above. However, a PNP-transistor (T2) has been added to act as a current follower (common-base stage) which provides a low input impedance at the emitter (T2), Note that the base voltage of T2 is assumed to be constant and thus the emitter voltage will be fairly constant. Consequently, since the collector resistor for the device to be biased (T1) is shared with the biasing-device (T2) an increase in collector current in T1 will result in an equal decrease in current in T2 and vice versa. The output of T2 directly drives the base of T1 with a current. Thus, a collector current error in T1 will be fed back directly to the base of T1 but with opposite sign. We could let the complete collector current of T2 drive the base of T1. However, this base current is fairly small, and if a larger current is desired through T2 a collector resistor, R_{C2} , must be inserted to sink that current, see figure 9.13. Although, this is not advantageous in terms of biasing stability since R_{C2} will have a lower resistance than the base of T1 which means that the T1 will be voltage driven rather than current driven.

We will investigate the performance of this circuit by considering an example with the same specification as for the examples based on passive biasing circuits, namely

- Wanted collector current: $I_C = 1\text{mA}$
- Wanted collector-emitter voltage: $V_{CE} = 3\text{V}$
- Available supply voltage: $V_{CC} = 12\text{V}$
- Transistor DC current gain: $\beta_0 = 100$ (to 140)

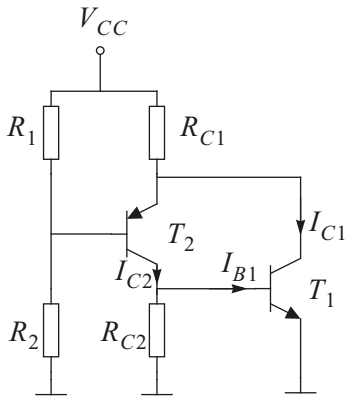


Figure 9.13 Active biasing using a PNP-transistor.

Example 9.4 Design of Active Biasing Circuit

Determine resistor values for the circuit below and investigate the performance in terms of temperature sensitivity of the collector current in T1 Consider two cases. First, without R_{C2} to have T1 purely current driven. Secondly, R_{C2} chosen such that the collector current for T2 is 10 times the base current of T1, I_{B1} .

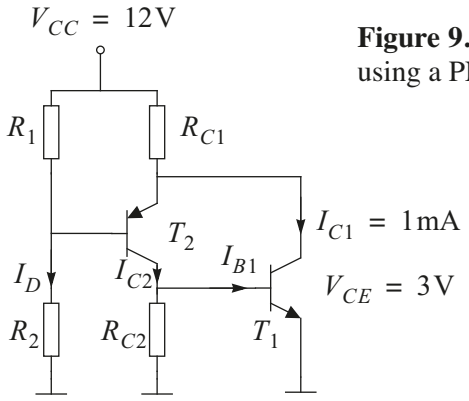


Figure 9.14 Active biasing of BJT using a PNP transistor.

Begin by considering the first case where there is no R_{C2} . Also, for simplicity assume 1mA in the voltage divider that provides current to the base of T2. The resistor values can be calculated directly.

$$\begin{aligned} R_{C1} &= (V_{CC} - V_{CE})/(I_{C1} + I_{E2}) \\ &= (V_{CC} - V_{CE})/(I_{C1} + I_{B1} \cdot (1 + 1/\beta_0)) \\ &= (V_{CC} - V_{CE})/(I_{C1} + (I_{C1}/\beta_0) \cdot (1 + 1/\beta_0)) \\ &= 9/1.0101\times10^{-3} = 8.91\text{k}\Omega \end{aligned}$$

$$R_1 = (V_{CC} - V_B)/I_D = (12 - 3 + 0.7)/1\times10^{-3} = 9.7\text{k}\Omega$$

$$R_2 = V_B/I_D = (3 - 0.7)/1\times10^{-3} = 2.3\text{k}\Omega$$

A plot with collector current in T1 versus temperature is shown below. The temperature dependency in T2 has also been taken into account.

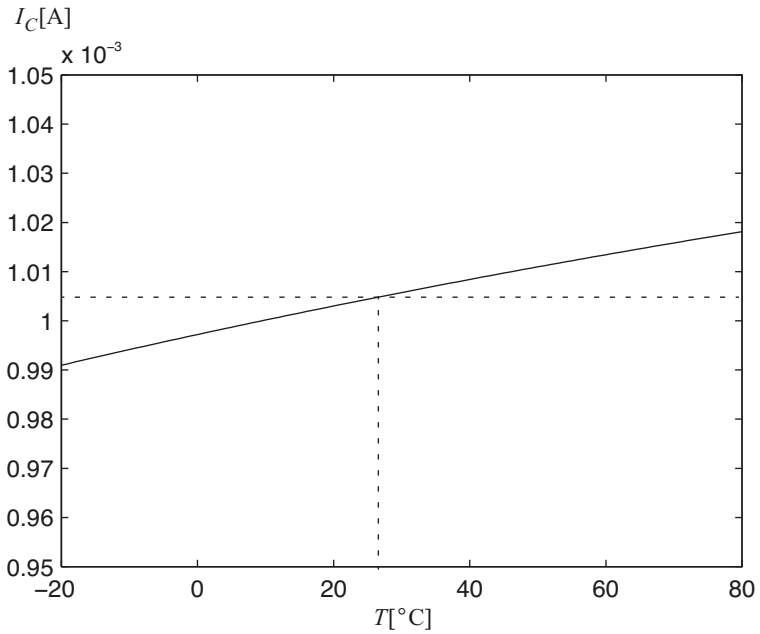


Figure 9.15 Temperature dependency of collector current using current driven active biasing.

The biasing stability is better than the passive biasing circuits that are based on a base current driven transistor. In this configuration the loop gain is roughly equal to the current gain of T1. However, it is on par with the passive biasing circuit based on a voltage driven base-emitter. The main disadvantage with the latter one is the RC circuit between the emitter and ground which may cause instability. This problem is not present in this active biasing scheme.

The second case meant that we should allow for a collector current in the feedback transistor T2 that was 10 times larger than the base current of T1. Thus, we need R_{C2} to sink this current. The reader might suspect that the performance will be degraded because a large part of the current that was previously fed to T1 directly will now go through R_{C2} instead and result in a voltage over R_{C2}. In other words, T1 will not be purely current driven any more. We will continue by calculating the resistor values and finally the collector current versus temperature will be plotted for this case.

$$\begin{aligned} R_{C1} &= (V_{CC} - V_{CE})/(I_{C1} + I_{E2}) \\ &= (V_{CC} - V_{CE})/(I_{C1} + I_{B1} \cdot (10 + 1/\beta_0)) \\ &= (V_{CC} - V_{CE})/(I_{C1} + (I_{C1}/\beta_0) \cdot (10 + 1/\beta_0)) \\ &= 9/1.1001\times10^{-3} = 8.2\text{k}\Omega \end{aligned}$$

$$R_{C2} = V_{B1}/10 \cdot I_{B1} = 0.7/100\times10^{-6} = 7\text{k}\Omega$$

R_1 and R_2 will be the same as in the first case.The plot is shown below and a degradation in performance is clearly seen.

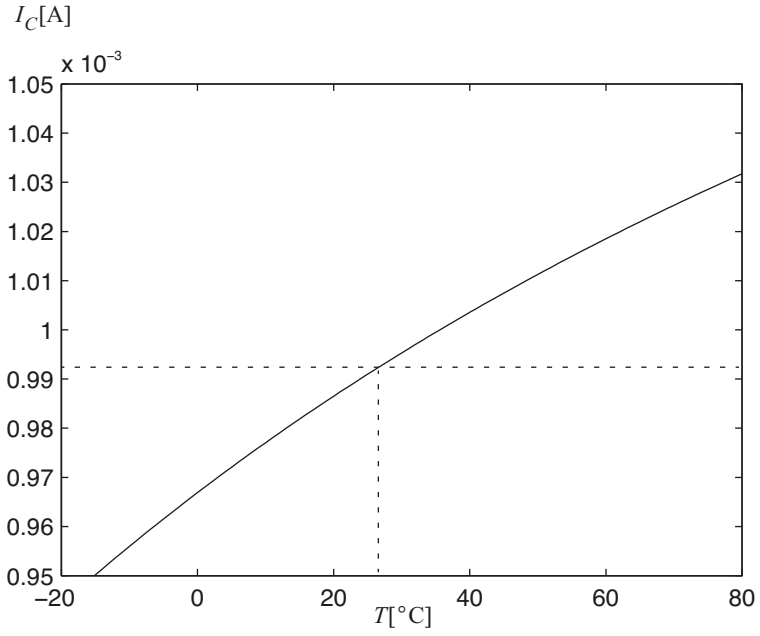


Figure 9.16 Temperature dependency of collector current using voltage driven active biasing.

In the very beginning of this section two concepts for active biasing was introduced. The second one, indirect feedback, is based on that a quantity not directly related to the operating point is measured to obtain an estimate.

The main advantage with this approach is that we do not need a resistor in series with the transistor collector or emitter. This feature is good for power amplifiers in particular where unnecessary power dissipation cannot be tolerated due to limited supply voltage, limited power consumption or due to heating problems.

One practical solution is to have a sensing device such as a diode or a transistor with good thermal coupling to the transistor to be biased. Figure 9.17 illustrates a simple solution which basically resembles a current mirror. A fixed current, I_D, is forced into the diode and thus as the temperature is changing the voltage over diode will change rather than the current (with roughly $-2\text{mV}/^\circ\text{C}$). The same voltage appears over the base-emitter of the transistor device which then "mirrors" the current I_D to the collector current I_C with some scaling factor. The scaling factor, or the ratio between these two currents is determined by individual parameters for each device but also the difference in the voltages V_D and V_C. This is also where the problems with this technique manifests themselves. If the transistor and the sensing device are not matched each amplifier must be trimmed separately to obtain a certain operating point (ratio between I_D andI_C). Moreover, they will not track over a wider temperature range because of different temperature coefficients. If both components are matched or even better, being part of the same

chip die, the sensing device will track the biased transistor for all temperatures with a good matching, i.e., the relation between the currents in the two components will be fixed and well defined. If the diode is implemented with a diode coupled transistor of the same type as the main device the I_D/I_C ratio will roughly be equal to 1. As a first order approximation the I_D/I_C ratio will equal to the ratio between the saturation currents for each device. Reference diodes are built into some RF power transistor for the purpose of biasing.

If I_D is made small compared with I_C in the circuit shown in figure 9.17 by having a scaled down version of the main device as sensing device the base current of the main device must be taken into account in the calculations or other circuit solutions should be considered to avoid the dependency on the base current.

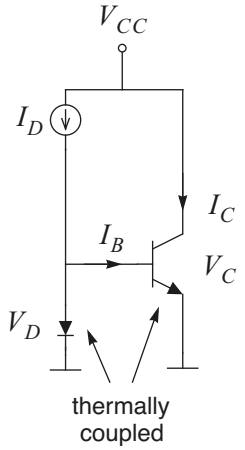


Figure 9.17 Simple circuit example with thermal feedback biasing (current mirror).

9.4 Biasing of Field Effect Transistors

The temperature sensitivity of FET devices is fairly small compared with BJTs. For that reason, FET-based amplifier may not need a feedback-based biasing to maintain the operating point because of temperature drift. Furthermore, whereas the BJT can be considered both as voltage and current driven FETs are solely voltage driven which simplifies the discussion.

For the biasing circuits that have been discussed above we rely more or less on the fact that the base-emitter voltage is almost constant for a fixed operating point. The temperature coefficient is only some $-2\text{mV}/^\circ\text{C}$ and the variation from one transistor to another is also small. The FET, however, has a very inaccurate threshold voltage. The spread in threshold voltage may very well be specified as a max-min range from, say, 1V to 3V.

The passive biasing circuits presented in figure 9.6 are not as attractive for FETs as for BJTs. Some FETs simply cannot be biased similar to BJTs since the operating point might require for example a negative gate-source voltage. Bipolar voltage supplies might be required. The large inaccuracy in threshold voltage also makes it more difficult to use the passive structures. We will illustrate that by assuming a FET that operates with a positive gate-source voltage. The first passive circuit, see figure 9.6a, is considered to be current driven for BJT devices. For a FET device the gate will have the same potential as the drain. It can be shown that if the loop gain is high the drain current

will still vary in proportion to $V_{DD} - V_D - \Delta V_{TH}$ where V_{DD} is the supply voltage connected to the drain resistor, V_D the drain voltage and ΔV_{TH} the maximum expected deviation from a nominal threshold voltage V_{TH}. Thus, to guarantee an accurate biasing point the voltage across the drain resistor must be much larger than the variation in V_{TH}. The second passive topology, shown in figure 9.6b, has a voltage divider that introduces an offsets between the drain and the gate. This adds another degree of freedom but also reduces the biasing loop gain. Finally, the last passive topology is based on series feedback with a resistor between source and ground. Again the spread in V_{TH} must be much smaller than the voltage across the resistor. One advantage with this scheme is that it can be configured for a negative gate-source voltage.

The active biasing network described for the BJT is very useful for FET devices, see figure 9.18. It is flexible and, for example, can be configured to provide a negative gate-source voltage. The collector resistor, R_C, was found to degrade the performance when biasing BJTs (see example 9.4). This is not the case for FETs that are purely voltage driven. A large R_C will result in a large loop gain.

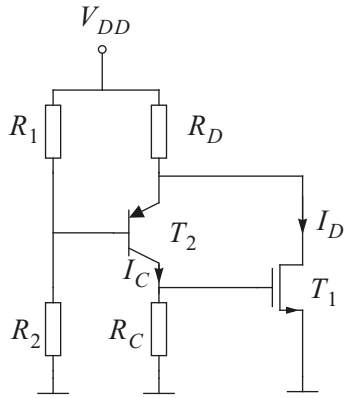


Figure 9.18 Active biasing for FET device using a PNP-transistor.

In practice it is important to take actions to prevent transient burn-out of the FET as the circuit is turned on. For example, the Schottky junction in a MESFET must not be momentarily forward-biased. This is solved either by having separate supplies for the gate and the main supply that are turned on in an appropriate sequence or by having different time constants associated with each pin.

9.5 Isolating Bias Design from Signal Design

So far we have only considered the biasing circuitry by itself. The problem is that we do not want it to affect the signal design that was dealt with in previous chapters. This section deals with various solutions to isolate the bias design from the signal design.

Obviously, we want a low-impedance DC path from the biasing circuit to the device. At the same time this path should isolate the device from the biasing circuitry at high frequencies. In this context we usually refer to a radio-fre-

quency-choke (RFC) that complies with our requirements. The RFC can be implemented in many ways. In figure 9.19 it is shown how it is used to isolate the signal design from the bias design using a passive biasing circuit. The schematic symbols indicate that the RFC is an inductive element which is one way of providing a high impedance path at high frequencies. In addition to the RFC a low impedance path to ground is sometimes required to prevent high frequencies from propagating through the biasing circuitry and into other parts in an electronic system. This is also shown in figure 9.19 where the capacitors ground the nodes between the RFCs and the biasing circuitry.

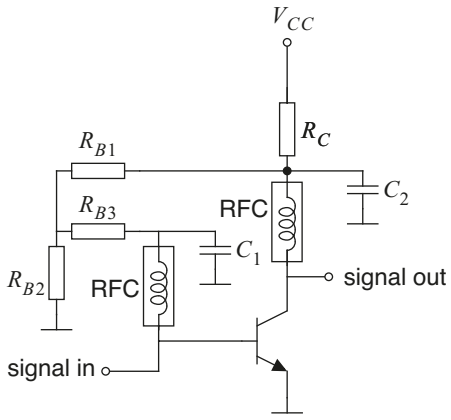


Figure 9.19 Passive biasing with RFCs.

At lower frequencies, say up to 100MHz, the RFCs can be realised with coils with ferrite cores. At higher frequencies, say to 1GHz or so, coils with air cores should be used. A coil, not matter how it is realised behaves like a parallel LC circuit due to distributed capacitance between turns. Thus, a coil exhibits a self-resonance frequency above which its impedance is capacitive. Consequently, an RFC coil must be designed such that the frequency of operation is below the self-resonance frequency.

A parallel LC circuit that is used to tune an amplifier will also be a good RFC even though the reactance of the inductive element does not provide sufficiently high impedance by itself. The impedance of the complete resonance circuit is typically very high (equal to the equivalent parallel resistance) close to the resonance frequency.

Transmission lines should be used in the GHz-range. Transmission lines can provide high impedances over a limited frequency band similar to a resonance circuit as has been demonstrated in earlier chapters. The most common method is to connect a biasing node to the desired terminal of a device through a $\lambda/4$ transmission line, see figure 9.20. In this case it is particularly important that this node is properly signal-grounded with a capacitor or the transmission line will not transform this to a high impedance. Typically, the characteristic impedance for these transmission lines are chosen much higher

than the rest of the system so as to transform the signal ground to a very high impedance. An alternative is to exploit existing stub lines that should be connected to ground anyway.

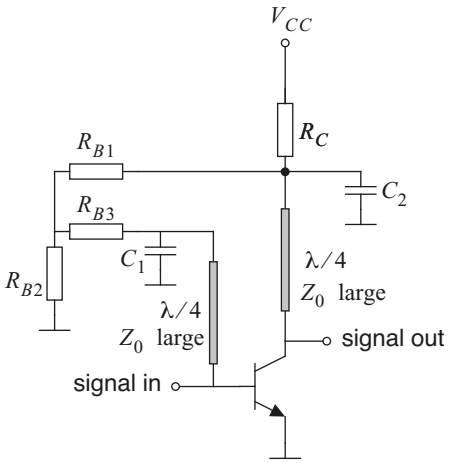


Figure 9.20 Passive biasing with transmission lines acting as RFCs (not to scale).

The grounding capacitor that was mentioned above should typically have values in the nF range to ensure stable operation over a wide range of frequencies. However, capacitors of this size may have large parasitics and a low resonance frequency. For this reason smaller capacitors in the pF-range (typically ceramic capacitors) are good and a necessary complement to the larger capacitors.

If the frequency of operation is some ten GHz or so the effect of parasitics in the grounding capacitors will have a large influence and implementing the capacitors directly on the printed circuit board might prove to be a more feasible solution. In figure 9.21, an example of such a solution is given that besides the PCB-capacitor also utilise a rat-race-like ring structure to improve the isolation even further.

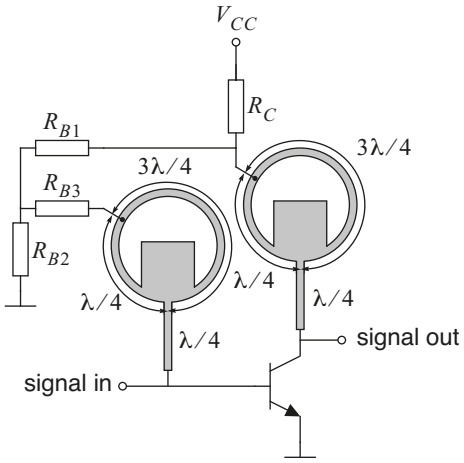


Figure 9.21 Biasing isolation without lumped circuit elements for isolation.

9.6 References

[1] P. R. Gray and R. G. Meyer, *Analysis and design of analog integrated circuits*, 3rd edition, John Wiley & Sons, 1993.

[2] G. Massobrio and P. Antognetti, *Semiconductor device modelling with Spice*, 2nd edition, McGraw-Hill, 1993.

Chapter 10

Oscillators

10.1 Black's Feedback Model

To be able to analyse the oscillating conditions Black's feedback model is used. Here the oscillator is split into two blocks. One amplifier which is considered to be wideband and one feedback network that is usually frequency selective.

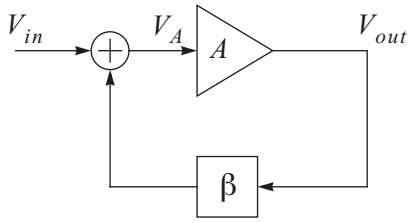


Figure 10.1 Black feedback model for oscillator.

The transfer function for the amplifier with feedback can be calculated using equation (10.1)

$$V_{out} = A \cdot V_A \tag{10.1}$$

$$V_A = V_{in} + \beta V_{out} \Rightarrow A_f = \frac{V_{out}}{V_{in}} = \frac{A}{1 - \beta A}$$

One may observe that the denominator becomes zero if the demands in equations (10.2) and (10.3) are met. The feedback gain A_f then becomes infinite.

$$|A \cdot \beta| = 1 \tag{10.2}$$

$$\angle(A \cdot \beta) = 0° \tag{10.3}$$

These equations are called the Barkhausen oscillation criteria. Assume that the loop is opened between the feedback output and the amplifier input and a voltage source is connected to the amplifier input. Each time the signal passes through the loop it is amplified with a factor A and decreased with a factor β before returning to the insertion point. If the former demands are met this implies that the output of the feedback network will be identical according to phase and amplitude. The feedback output can be connected to the amplifier input and the signal source can be removed. The system will be self-generating and we will perform continuous oscillation.

This would imply that as soon as we want to start an oscillator we need to connect a signal source? Well, literarily speaking yes and that is why a real oscillator is designed with a loop gain $A \cdot \beta$ larger than unity. The presence of noise at the amplifier input will excite the loop and cause a forever growing signal at the frequency where the oscillation demands are met. The signal will continue to grow until something in the loop is limiting. Often this is the amplifier that has a voltage or current supply that limits the magnitude of the amplifier output signal. Note that the Barkhausen criteria do not say anything about the magnitude of the oscillator signal, the equations only establish when oscillation is possible based on phase and gain.

10.2 Oscillator Analysis

The oscillator can be modelled as in figure 10.2 with the amplifier having high input and output impedance, i.e. a voltage controlled current generator.

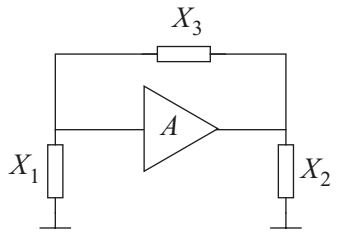


Figure 10.2 A basic model of an oscillator with reactive feedback.

From the amplifier input node to ground a reactance X_1 is connected along with X_2 from output to ground and X_3 from output to input. The Barkhausen criteria can now be translated into the following criteria derived from Appendix A

$$X_1 + X_2 + X_3 = 0 \tag{10.4}$$

$$\beta = \frac{X_1}{X_1 + X_3} = -\frac{X_1}{X_2} = \frac{1}{A} \tag{10.5}$$

Equation (10.4) can be interpreted as that there must exist reactance of opposite signs, one can not make an oscillator using solely capacitors or inductors. Equation (10.5) is the voltage division from amplifier output back to its input. Combining this with the first equation yields the final result. Here it is also established that A must have the same sign as β . If X_1 and X_2 are reactances of the same type then an inverting amplifier must be used to ensure oscillation.

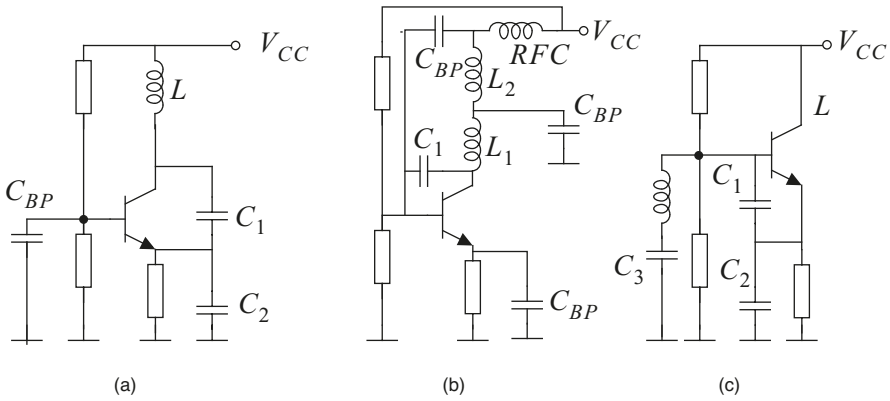
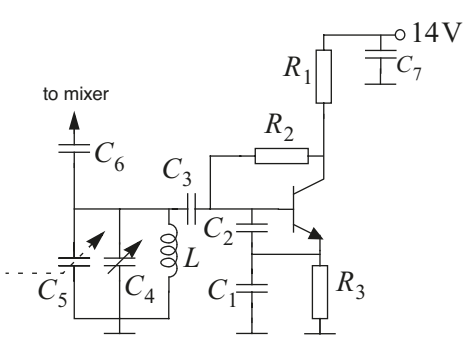


Figure 10.3 Oscillator configurations: (a) Colpitt (b) Hartley (c) Clapp.

There are some standard types of oscillators depending on how the feedback is arranged. One of the most common oscillator arrangements found is the Colpitts oscillator with its feedback path through a capacitive voltage divider, see figure 10.3.a. The feedback can also be through a inductive tap as in figure 10.3.b. This configuration is called a Hartley oscillator. The third circuit is a Clapp oscillator and looks at a first glance much like the Colpitts oscillator. The difference is that there is a small capacitor in series with the inductor. This has two benefits. First, the inductive branch can be connected to any DC voltage without affecting the quiescent point. Secondly, by making the series capacitor much smaller than C_1 and C_2 , say ten times, the influence from C_1 and C_2 in equation (10.4) will be very small. Thus it is easy to make a tuneable oscillator with a large frequency range using the Clapp configuration.

In schematics from real life we do often not find oscillators as well structured and consistent as the previous examples. In figure 10.4, parts of an oscillator found in a commercial home stereo receiver is depicted.



C_1 = 10pF
C_2 = 20pF
C_3 = 10pF
C_4, C_5 = tuneable
C_6 = 1pF
C_7 = 22nF
R_1 = 3.3kΩ
R_2 = 100kΩ
R_3 = 1kΩ

Figure 10.4 Local oscillator for home stereo FM receiver.

We start by analysing the configuration of the amplifier. Since one of the capacitors, C_7, is much larger than the others, it is likely to assume that it is an AC short-circuit and not directly crucial in determining the frequency selective feedback. The resistors $R_{1\ldots3}$ form the bias network. It would be possible to place a bypass capacitor between the collector and ground but then the AC feedback would be lost for the amplifier. This leaves us with some options on amplifier configuration.

Table 10.1 Amplifier configurations.

Type	Input	Output	$\lvert A_v \rvert$	$\angle Av$
Common Emitter (CE)	base	collector	$\lvert A_v \rvert \geq 0$	180°
Common Base (CB)	emitter	collector	$\lvert A_v \rvert \geq 0$	0°
Common Collector (CC)	base	emitter	$0 < \lvert A_v \rvert < 1$	0°

From table 10.1 a very helpful conclusion regarding oscillator amplifiers can be drawn: **the collector is never the amplifier input and the base is never the output**. There might be other odd networks configured that way but then they are not amplifiers.

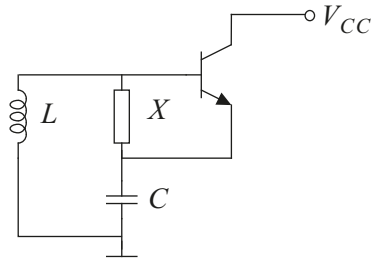
Analysing the circuit of figure 10.4, one can see that the amplifier configuration is common collector with input on the base and output on the emitter. For a common collector amplifier, A is positive and the reactances X_1 and X_2 must therefore be of opposite sign. We find that X_2 is the capacitor C_1 and X_3 is the capacitor C_2, This means that X_1 must be an inductance and in this case it is not a single component but a full network consisting of L and $C_{4\ldots6}$. Observe the two tuning capacitors where one is to set the frequency offset and the other one (dashed) is the one connected to the frequency dial on the receiver front panel. The equivalent reactance is

$$X_1 = -\frac{1}{\omega C_3} + \frac{\omega L}{1 - \omega^2 L(C_5 + C_4)} \tag{10.6}$$

By choosing C_4 and C_5 properly the second denominator becomes small and a very large tuning range of the inductive part can be achieved even though C_5 varies only a couple of pF. This fits the application well since the FM band is fairly wide (usually about 20MHz). The capacitor C_6 is small to reduce the effects from pulling, i.e. frequency variations due to the load from the following stage. Thereby it has very little impact on the frequency of the oscillator.

Example 10.1 Oscillator analysis

What kind of reactance is X and what is its value if the oscillator is supposed to run at 300 MHz? (bias network omitted)



$L = 150\text{nH}$
$C = 3\text{pF}$

Since the transistor is connected in a CC configuration we know that it is a non-inverting amplifier with voltage gain approximately 1. With the input on the base and the output on the emitter X_1 will be the coil reactance and X_2 the capacitor.

$$\beta = \frac{X_1}{X_1 + X_3} = \frac{j\omega L}{j\omega L + jX}$$

For β to be positive, X must be either an inductance, or a capacitance with $X_C < X_1$. If X is chosen to be an inductance we will have β smaller than one and oscillation is thereby impossible. Equation (10.4) gives that the sum of all X must be zero at the given frequency. The reactance and thus component value can then be calculated as

$$-\frac{1}{\omega C_x} = X_3 = -(X_1 + X_2) = -\left(\omega L - \frac{1}{\omega C}\right)$$

$$C_x = 5\text{pF}$$

10.3 Oscillator Noise

According to the previous analysis we would expect the signal from the oscillator to be a peak at a certain frequency (and perhaps harmonics of this) when viewed by a spectrum analyser. If the span is low enough combined with high resolution we will however see a peak with increased noise levels close to the peak, see figure 10.5.

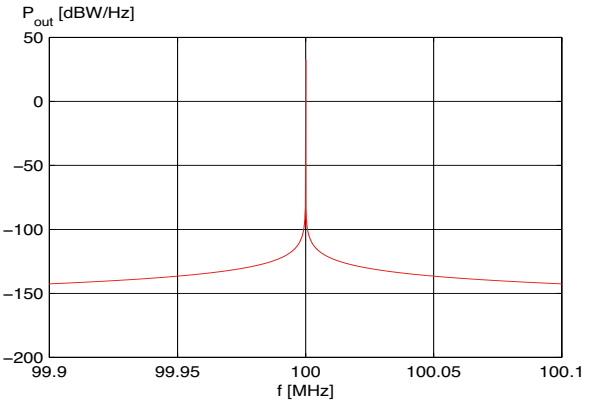


Figure 10.5 Oscillator noise density versus frequency, $f_0 = 100\text{MHz}$.

These "skirts" are called phase noise and can be regarded as the oscillator jittering somewhat in phase. Since it is mainly white noise that causes the phase jittering we will not see any peaks in the side bands, merely a degrading noise level the further we move away from the peak. Assume that we have a noisy amplifier with noise figure F as active element in the oscillator according to figure 10.6.

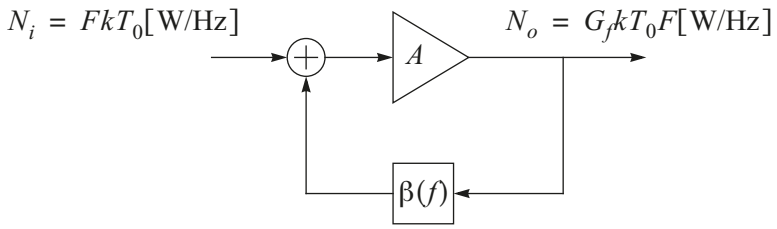


Figure 10.6 Noise model of oscillator.

The circuit can be seen as a wideband amplifier with a frequency selective feedback network providing positive feedback. The gain of the amplifier with feedback is, as derived previously but with β frequency dependent

$$A_f(f) = \frac{A}{1 + \beta(f)A} \tag{10.7}$$

$$G_f = A_f^2 \tag{10.8}$$

$$\beta(f) = \frac{\beta_0}{1 + jQ\dfrac{2|f - f_0|}{f_0}} \tag{10.9}$$

If A is chosen to be 10 and β shows bandpass characteristics with a maximum of 0.1 making the loop gain exactly unity at f_0 , we will obtain oscillation. If we observe A_f close to, but not exactly at, f_0 there is a decrease in β to for example 0.09. At this point the feedback gain will be $A/0.1 = 10A$ which is larger than the open loop gain A . This further implies that any noise that is present at the amplifier input will be heavily amplified as we approach f_0 and thus causing the behaviour observed in figure 10.5. Assuming that the feedback circuit is a first order resonator with resonant frequency f_0 and Q being the loaded Q-value of the resonator, the spectrum can be calculated [3][4] to be according to figure 10.7 where G and F are the amplifier power gain and noise figure, f_0 is the centre frequency and P_{osc} is the output power of the oscillator. The noise far away from the centre frequency is the same as for an amplifier with no feedback

$$N_{01} = FGkT_0[\text{W/Hz}] \tag{10.10}$$

At the corner frequency f_b the feedback becomes significant and an increase of noise density with 6dB per octave can be observed. The noise density is now

$$N_{02} = \frac{FGkT_0}{4Q^2}\left(\frac{f_0^2}{(\delta f^2 + \alpha^2)}\right)[\text{W/Hz}] \tag{10.11}$$

The parameter α is included to limit the output power. The integral of the noise density will thus have a power equal to P_{osc} .

$$\alpha = \frac{FGkT_0\pi f_0^2}{4P_{osc}Q^2} \tag{10.12}$$

This is the full noise spectrum consisting of both AM and PM noise. If the oscillator signal is limited, the AM noise is suppressed and only the PM noise becomes relevant. This results in a 3 dB decrease of the noise. If the oscillator signal is down converted to DC as in demodulation, we will have a folding of the noise spectrum around DC resulting in a 3dB increase. One differs between these by referring to the noise as single sideband or dual sideband noise power.

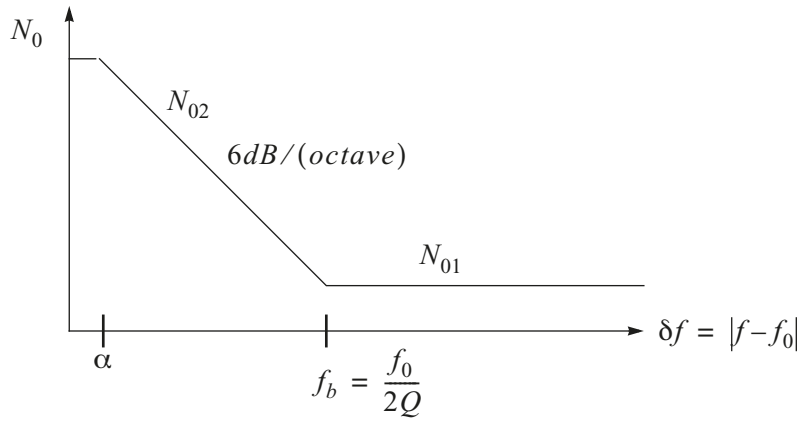


Figure 10.7 Single sideband noise spectrum of oscillator with first order bandpass feedback.

It should be carefully noted that the noise level beside the centre frequency is not determined by the oscillator power. By increasing the available power to the amplifier in the circuit, we get a higher C/N_0 for the oscillator

To minimize phase noise there are several possibilities

- Increase the Q value of the resonator.
- Use an amplifier with low noise figure.
- Have the oscillator running at high output power.
- Keep amplifier gain as low as possible.

The first and last point are somewhat incompatible since a resonator with high Q often implies high loss and thereby a need for higher gain to keep loop gain at or above unity. On the other hand, since the noise is proportional to Q^{-2}, there is still a merit in increasing Q despite the fact that gain has to be increased. Observe that increasing Q only lowers the noise close to the centre frequency. Further away the feedback is negligible and the output noise is equal to that of the amplifier itself.

Using the equations in figure 10.7, the 3dB bandwidth can be calculated. This is not the same as the feedback network bandwidth even though they are related through the equations

The 3dB break point occurs when $\alpha = \delta f$, this will yield a bandwidth equalling $B_{3dB} = 2\delta f_{3dB} = 2\alpha$. For a 1mW 900MHz oscillator with a Q of the feedback network equalling 30 and an amplifier noise figure of 4 (or 6dB) the 3dB bandwidth will be 7.3 mHz and that is milli Hertz, not Mega Hertz. This corresponds to a Q value of 12×10^{12}, a multiplication of $f_0/(2Q\alpha)$ due to the positive feedback.

Example 10.2

What is the adjacent channel selectivity (25 kHz channel separation, 16kHz bandwidth) if one has an 900 MHz oscillator with the following parameters

G=10dB, F=3dB, P_{osc}=10dBm, Q=30

Solution:

Since $\delta f = 25\text{kHz}$ the influence from α can be neglected. Using equations from figure 10.7 one gets

$$f_b = 900\times10^6/(2\cdot 30) = 15\text{MHz}$$

and the noise will be on the 6dB slope.

The oscillator noise power in the adjacent channel bandwidth will be

$$N_{osc} = 1.15\times10^{-13}\text{W/Hz}\cdot 16\text{kHz} = 1.84\text{nW} = -87.4\text{dBW}$$

Assuming multiplicity in the mixer, i.e. the wanted signal (C for carrier) is multiplied with the oscillator and the interferer (I) in the adjacent channel with the oscillator noise, the selectivity becomes

$$C\cdot P_{osc} = N_{osc}\cdot I \Rightarrow \frac{C}{I} = 67.4\text{dB}$$

if a signal-interferer ratio of 1 is accepted at the mixer output. This is not very good and can be blamed on the low Q value of the resonator. In case one has to improve the adjacent channel selectivity one should first try to replace the resonator. Observe that the adjacent channel selectivity is not dependent on how the IF filter suppresses the adjacent channel since this is mixed with the noise to the centre frequency of the IF filter. This phenomenon is called reciprocal mixing.

To make a more thorough and exact analysis of the phase noise, the $1/f$ noise of the amplifier should also be considered. This is however mostly a problem when the oscillator signal is demodulated and for very low baseband frequencies. In case such an analysis is still needed can it be found in [1]

10.4 Oscillator Design

The nonlinear behaviour of the oscillator due to compression in the amplifier makes it very hard to design for a certain output power without using simulation tools like HP-MDS by Hewlett-Packard or JOmega by EESof. Because of this it is not covered in this chapter. Designing to meet frequency and oscillation conditions is however feasible with relatively straightforward methods.

The first step towards a functioning oscillator is to see that the Barkhausen criteria are met, i.e. determine either a feasible A or β and thereby getting the other from equation (10.3).

For a fixed feedback factor β , there will always be a corresponding gain, i. e. output power that maintains a stable oscillation amplitude. In figure 10.8 the output power for an amplifier is plotted versus the input power for the fundamental, second order and third order harmonic. Observe that since the scales in the intercept diagram are in dB relating to power instead of amplitude, so are also references to gain G and feedback β .

$$G = (\Delta P_{OUT} - \Delta P_{IN}) = -\beta \ [\mathrm{dB}] \tag{10.13}$$

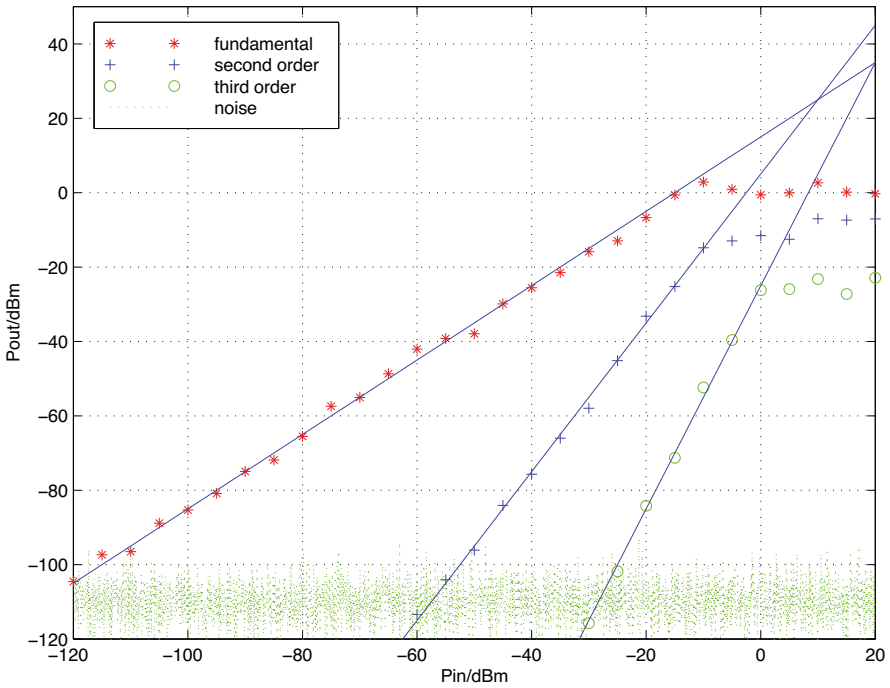


Figure 10.8 Intercept diagram of amplifier with 15 dB gain.

Depending on the feedback factor there will be a P_{IN} with a $G = P_{OUT} - P_{IN}[\mathrm{dB}]$ corresponding to $-\beta$. In case the loop gain is larger than unity the oscillator signal amplitude will increase up to a point where the relation between P_{OUT} and P_{IN} is exactly β . If, for some reason like temperature variations, the feedback factor is changed, the output signal will change as well until the Barkhausen criteria again are met. Another point that should be noted is the magnitude of the harmonics. Suppose we have designed an oscillator with a $\beta = 0.5 = -6\mathrm{dB}$ and $G = 15\mathrm{dB} = 5.6$ i.e. loop gain $G \cdot \beta$ equal 2.8. With a loop gain this high the output signal will rise until the loop gain in some way is decreased. From the intercept chart one can see that above $-10\mathrm{dBm}$ the fundamental tone deviates from the ideal line due to compression. At $P_{IN} = -5\mathrm{dBm}$ the output is not $P_{IN}+15\mathrm{dB}$ as should be the case but $P_{IN}+6\mathrm{dB}$ due to this compression of the amplifier. At this point the harmonics are 12 and 21 dB below the wanted signal. If the feedback factor is decrease to a more proper $-14\mathrm{dB} = 0.2$ the loop gain will be $5.6 \cdot 0.2 = 1.1$ and the amplifier is less into compression.

The bottom line is that the chart can be used to find where on the fundamental curve that the dynamic gain times the feedback factor must always equal one. At this point the level of the harmonics can be derived from the intercept chart.

A reasonable question is now: why not design for a loop gain of unity?

By designing for a loop gain slightly more than unity, you run the risk that component spread and variations in the surroundings (temperature, supply voltage etc.) will yield an actual loop gain of less than unity and the oscillator will never start. On the other hand, as previously stated, using too large loop gain only causes lots of harmonics and unnecessary power consumption. As an example, in a portable receiver as a mobile telephone, up to 20% of the power consumed when the telephone is in receiving mode is due to the local oscillator. Keeping the oscillator power low is therefore a crucial point in getting long battery-time in the receiver. In some TDMA systems like GSM the oscillator is shut down during the slots used by other receivers and started again in time to receive the bits of the wanted slot properly.

10.5 Negative Resistance Oscillators

In microwave frequency applications the design procedure of an oscillator is somewhat different. The task is still to calculate how the feedback should be applied to ensure oscillation, but since many of the feedback paths are either unknown or at least hard to specify at several GHz a different approach is needed. From the amplifier theory we know that from the S-parameters a stable and instable region can be calculated both for the input and output of the transistor. This implies that if we can place the source and load impedances in this unstable region, we will have an oscillator. That much is true but the problem still exist on how to determine the oscillating frequency and to ensure that the oscillation is present even when the output is loaded resistively.

If we for a short while return to the Colpitts oscillator with its feedback through a capacitive tap, we can calculate the impedance on the collector (where we will later replace the RFC for the actual coil but we leave that for the time being).

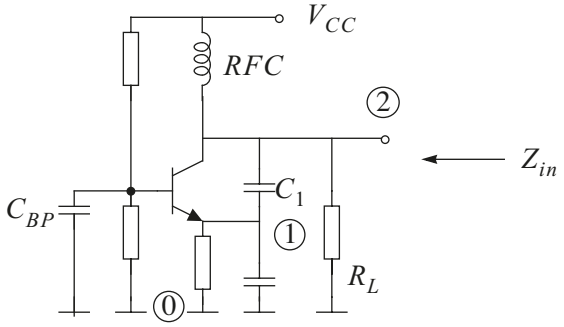


Figure 10.9 Measuring output node impedance of the Colpitts oscillator.

For calculating the input impedance in a node there are several possibilities. One can set up an equation system based on Kirchoffs voltage and current laws and solve the system by brute force. Other methods are using the indefinite admittance matrix (see Appendix A) or the superposition method [2]. The input impedance can be calculated as in equation (10.14). The term g_x is the admittance for the transistor parameters as used in chapter 6.4.2 and Z_{20} refers to the impedance between nodes 2 and 0.

$$Z_{20} = \frac{V_{20}}{I_{20}} = \frac{Y_{20}^{20}}{Y_0^0} = \frac{g_m + g_{be} + j\omega(C_1 + C_2)}{g_m g_L - \omega^2 C_1 C_2 + j\omega(g_L(C_1 + C_2) + g_{be}C_2)} \tag{10.14}$$

Multiplying with the conjugate of the denominator gives a real denominator and a real part of the numerator equalling

$$(g_m + g_{be})(g_m g_L(-\omega^2 C_1 C_2)) + \omega^2(g_L(C_1 + C_2)^2 + g_{be}C_2(C_1 + C_2)) \tag{10.15}$$

If we assume that $g_{be} \ll g_m$ and g_L is negligible (for the time being) this equation reduces to

$$-g_m\omega^2 C_1 C_2 + \omega^2 g_\pi (C_1 + C_2)C_2 = \omega^2 g_m\left(\frac{(C_1 + C_2)C_2}{\beta} - C_1 C_2\right) \tag{10.16}$$

We can have a negative real part of the output impedance of the circuit if the component values are chosen properly. One can also see from equation (10.15) that if g_L is large i.e. the load impedance is small on the output we will instead get a positive real part and oscillation will be impossible.

When calculating the imaginary part of the numerator one will find that it is negative regardless of operating point or magnitude of the resistive load.

$$\begin{aligned}\mathrm{Im(numerator)} = -j\omega \cdot [&g_{be}g_l(C_1 + C_2) \\ &+ (g_m + g_{be})g_m C_2 + \omega^2(C_1 + C_2)C_1 C_2]\end{aligned} \tag{10.17}$$

This is accurate since we what is missing to make the oscillator complete is a coil to tune out the negative imaginary part. It can be seen that the magnitude of the reactance is dependent not only on the values of the capacitors but also on the conductive terms g_m, g_L and g_{be} making the oscillator frequency load sensitive.

10.5.1 Negative Resistance vs. Reflection Coefficient

A node having a negative resistance together with a reactance of arbitrary sign can also be represented as a reflection coefficient with magnitude greater than unity. That means that each reflected wave is larger than the incoming wave and the node behaves as a generator. This representation of a system able to perform oscillation is rather favourable at microwave frequencies where each feedback element in the simplified oscillator model may be hard to determine. By adding an element that cancels out the reactive part of the negative impedance node at a certain frequency we will get oscillation at that particular frequency.

The conditions for oscillation can be expressed in three equations

$$K = \frac{1 + |\Delta|^2 - |S_{11}|^2 - |S_{22}|^2}{2|S_{12}S_{21}|} < 1 \tag{10.18}$$

$$\Delta = S_{11}S_{22} - S_{21}S_{12}$$

$$\Gamma_{IN}\Gamma_S = 1 \tag{10.19}$$

$$\Gamma_{OUT}\Gamma_L = 1 \tag{10.20}$$

Equation (10.18) says that the transistor must be potentially unstable to be able to use in an oscillator, i. e. by putting a certain passive circuit on the input and the output the transistor becomes unstable. The two equations (10.19) and (10.20) are really different representations of the same phenomenon since if the former is true so is also the latter and vice versa [5]. The equations state that the reflection factor on the load and source must be chosen so the input and output are resonated by their respective terminations and the magnitude must equal one. If $\Gamma_{IN}\Gamma_S > 1$ it will cause an infinitely growing signal that in practice will be limited by the amplifier's available power, just as in the negative impedance case. If it is smaller than unity and or the phase is not exactly zero, oscillation will not occur.

Equation (10.19) can also be written on impedance form

$$\Gamma_{IN}\Gamma_S = \frac{R_{IN} + jX_{IN} - Z_0}{R_{IN} + jX_{IN} + Z_0} \cdot \frac{R_S + jX_S - Z_0}{R_S + jX_S + Z_0} = 1$$

This is valid when the following relations apply

$$R_{IN} + R_S = 0 \tag{10.21}$$

$$X_{IN} + X_S = 0 \tag{10.22}$$

Example 10.3

The transistor NE856 has the following S parameters at an operating point of $V_{CE} = 10\text{V}$, $I_C = 10\text{mA}$, $f_0 = 1\text{GHz}$.

$$S_{11} = 0.25\angle{-168}^\circ$$
$$S_{21} = 3.59\angle{75}^\circ$$
$$S_{12} = 0.13\angle{66}^\circ$$
$$S_{22} = 0.41\angle{-41}^\circ$$

Calculating the K and Δ value of the transistor at the specified point gives

$$K = 0.9680$$
$$\Delta = 0.37$$

which renders the transistor being potentially unstable.

Plotting the output stability circle shows that due to the large K-value, only a small part of the Smith chart is in the unstable region. It will therefore be hard to transfer the load impedance, typically 50Ω, to the desired Z_L. In this case it is better to use the same transistor in a common base configuration. By transferring S parameters to Y, calculate the Y-parameters for common base and return to S parameters give

$$\begin{aligned} S_{11} &= 0.92\angle 155^\circ \\ S_{21} &= 1.89\angle -33^\circ \\ S_{12} &= 0.066\angle 145^\circ \\ S_{22} &= 1.1\angle -22^\circ \end{aligned}$$

The K factor is now -1.02 and the output stability circle is now completely inside the Smith chart and closer to the 50Ω point. This implies that even a low Q network will transfer the load to a point where oscillation is possible

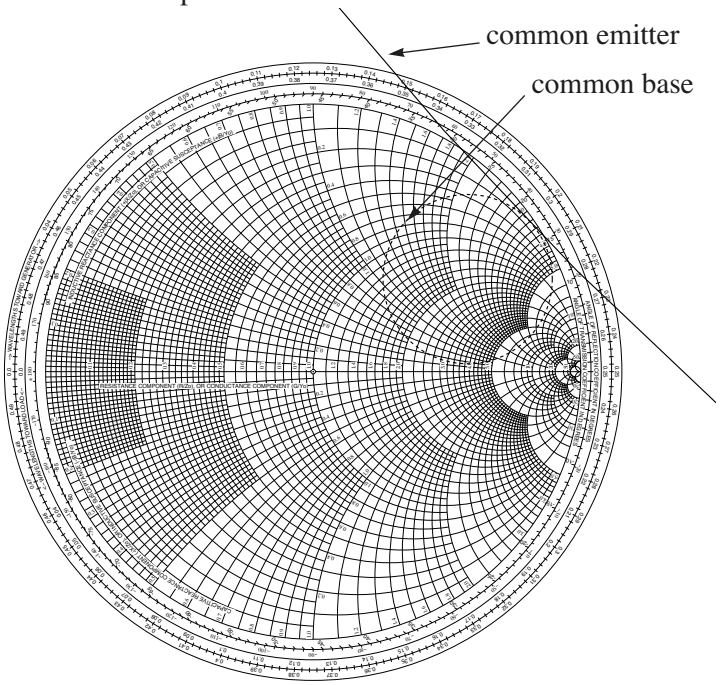


Figure 10.10 Output stability circles for common emitter and common base coupling of transistor NE856.

Start by picking a feasible point inside the unstable area, for example $\Gamma_L = 0.71\angle 45^\circ$. This is translated into an impedance of

$$Z_L = Z_0\frac{1+\Gamma_{IN}}{1-\Gamma_{IN}} = 50 + j100$$

which can be realised by a 50Ω load in series with a 16nH inductance at the chosen frequency. Connecting this to the amplifier output gives $\Gamma_{IN} = 1.08\angle 164^\circ$ which shows that oscillation is possible since the magnitude is greater than unity. Left is now to calculate the load impedance on the input, the emitter, of the amplifier. Since the oscillator demand is $\Gamma_{IN}\Gamma_S = 1$ that further implies

$\Gamma_s = 0.93\angle -164°$. It was stated earlier that the loop gain must be larger than unity to ensure oscillation even with spread in component values. To accomplish this, the input impedance must be calculated

$$Z_{IN} = -1.96 + j7.02$$

Equations (10.21) and (10.22) can now be used to determine a proper tuning network under the demands that $-1.96 + R_s < 0$ and $7.02 + X_S = 0$

There is no need of including an additional resistance in the source network. It is mentioned in literature [6] that choosing $R_s = |R_{IN}|/3$ is suitable for a high success rate. In this example the resistive part is omitted. For the reactive part

$$X_S = -7.02|_{f=1\text{GHz}} \Rightarrow C_s = 22.7\text{pF}$$

The final oscillator configuration will therefore look like this (with bias network omitted).

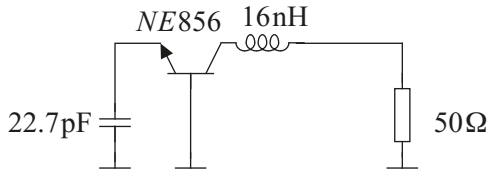


Figure 10.11 AC equivalent of 1 GHz negative resistance oscillator.

10.6 Voltage Controlled Oscillator

As seen in the previous chapters, the oscillator resonant frequency is dependent of the reactances in the surrounding network. By making either of these voltage controlled one will get an oscillator with an output frequency proportional to the control voltage, a voltage controlled oscillator (VCO).

The most common procedure is to replace one of the capacitors with a voltage controlled capacitor, a varicap or varactor. By changing its value the frequency that is changed to keep the reactance sum in equation (10.4) equal to zero. One problem that might occur is that by changing the capacitance, the demands for oscillation might not still be fulfilled and the signal disappears. It is therefore crucial to have negative impedance/$|\Gamma_{in}| > 1$ on the tuning node over the full tuning area.

A popular oscillator configuration for VCO:s is the Clapp oscillator, figure 10.12. A varicap C_{var} is added to the capacitor C_3 in series with the coil and by the relation between C_{var} and C_3 one can determine the sensitivity, the number of "MHz/volt" of the VCO.

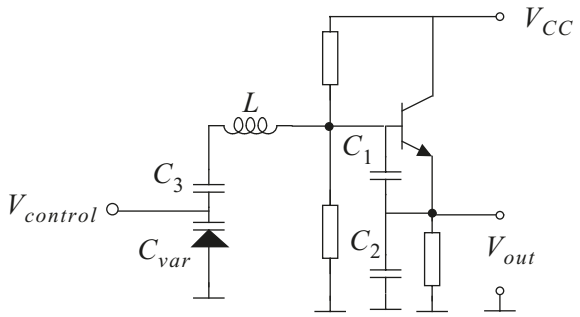


Figure 10.12 Voltage controlled Clapp oscillator.

The varicap is a diode that is optimised to behave as a standard capacitor when reverse biased. In this condition a depletion layer will occur between the p and n layers of the diode. This layer will not permit any current flow and can be regarded as a dielectric area between the two conducting p and n layers respectively. By this we will have a diode behaving like a capacitor.

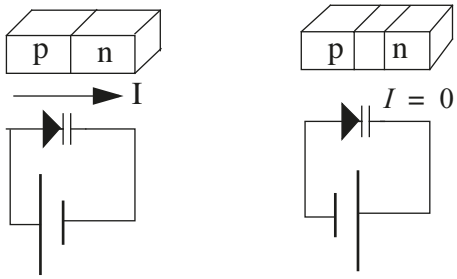


Figure 10.13 The varicap diode in forward and reversed bias.

When changing the reverse voltage on the varicap, the width of the depletion layer changes and thereby the distance between the "capacitor plates".

The capacitance of the diode is calculated using equation (10.23) where the variables are those used in the simulation software SPICE model of the diode. V_R is the reverse bias voltage. V_j is the junction potential, usually around 0.8V. $C_j(0)$ is the zero bias junction capacitance and M is the grading coefficient (0.1-1.1, typical 0.5). All these parameters are often supplied with the data sheets for each specific device.

$$C_d(V_R) = \frac{C_j(0)}{\left(1 + \frac{|V_R|}{V_j}\right)^M} \tag{10.23}$$

With the oscillator having a resonance frequency of $f_0 = 1/2\pi\sqrt{LC_{tot}}$ the sensitivity of the oscillator can be calculated as in equation (10.24).

$$\frac{df_0}{dV_R} = \frac{df_0}{dC_{tot}} \cdot \frac{dC_{tot}}{dC_d} \cdot \frac{dC_d}{dV_R} \tag{10.24}$$

If the oscillator is of Clapp type, the change in total capacitance can be regarded as the same as the change in the varicap if the varicap value is chosen much smaller than the other capacitors in the circuit. The sensitivity is then as stated in equation (10.25) provided that $M = 0.5$

$$df_0 = -\frac{1}{2}\frac{f_0}{C_d} \cdot 1 \cdot -\frac{C_d}{2V_j\left(1+\frac{V_R}{V_j}\right)}dV_R = \frac{1}{4}\frac{f_0}{(V_j+V_R)}dV_R \qquad (10.25)$$

This means that to get a sensitive oscillator the reverse voltage of the varicap should be low. On the other hand, this will cause a nonlinear relation between frequency and voltage. Again, it is the application and its demands that determines the choice.

The VCO can be regarded as a contradiction in itself. The first demand on an oscillator is that it is frequency stable and is not influenced by changes in load, temperature and supply voltage. These are wanted properties for the VCO as well but with the addition that you want to be able to tune it over a large bandwidth and have a short settling time when changing the frequency. To combine all these features one usually has to place the VCO in a phase locked loop (PLL) where a very stable reference frequency is used. The drawback of this method is that the frequency must be set in discrete steps.

10.7 Resonators

For many applications the accuracy and phase noise in a LC-oscillator is not sufficient for the demands put on a local oscillator in modern radio applications. Often the capacitors are temperature sensitive and the coils have low Q-values. The spread in component values can also be a major drawback of these otherwise simple and cheap items.

The first attempt in enhancing the oscillator accuracy performance is to replace the coil with some other high Q resonator arrangement. There are various resonators of different material that can be used depending on the frequency range and if the oscillator must be tuneable, like a VCO.

10.7.1 Microstrip Resonators

A transmission line of the electrical length $\lambda/4$ with a short circuit termination can be modelled a a parallel resonant circuit with resonant frequency $f_0 = c/\lambda$. By making the resonator somewhat shorter we get a inductance with very nice features like high Q-value where the strip inductance $X_L = jZ_o\tan(\beta\ell)$.

Microstrip resonators are especially attractive when realising small inductances with high Q value. Implementing an inductance as a microstrip strip on a Teflon circuit board value will easily achieve Q-values>100. Realizing the same inductance as a wound coil would give a Q value in the area of 10-15. At high frequencies where the wavelength is short it is definitely worth mak-

ing small inductances using a transmission line. In the chapter on transmission lines the actual design and implementation of microstrip resonators is described in detail.

Since the field pattern in and around the strip is rather complex so will the formulas be describing the electric properties of the strip. For many of the structures there are special cases where certain approximations may be used and one soon ends up with a huge amount of equations, each with their own application area. In recent years simulation programs like HP-MDS have been completed with packages that calculate for example the microstrip impedance from the physical dimensions of the board and strip. Using these software that actually solves Maxwell's equations for the strip is of course convenient in that you can use very complex structures. It is though hard to check the validity of the result by other means than implementation. Determining the parameters ε, σ and $\tan\delta$ for the conductor and dielectric is also a difficult task. The manufacturer often provides figures but the spread can be large between batches.

10.7.2 Coaxial Resonators

One major problem with the microstrip resonators is the radiation from the strip. When putting the resonator in a shielded box the metal surrounding change the electric field and thereby Z_0 and ε_{eff} for the strip. A better solution in then to use a structure where the ground plane covers the strip. It gives a more homogenous field around the strip and radiation is minimised.

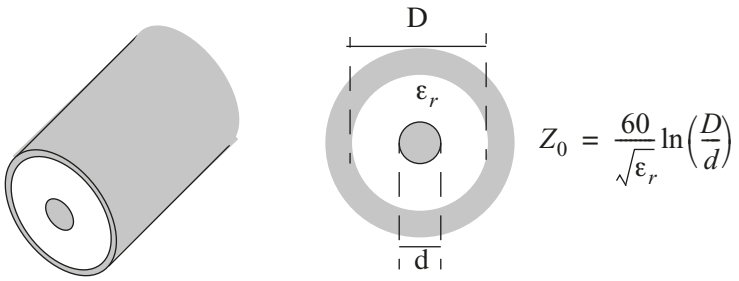


Figure 10.14 Coaxial cable with characteristic impedance.

A coaxial cable can be used in the same way as the microstrip, the difference is that the dimensions are often fixed and one has to settle for a limited selection of Z_0. Calculating the Q-value is done in a similar way as for the microstrips. One should be careful and make sure that the resistances in both the outer and inner conductor are included since these are regarded as finite structures. The ε_{eff} equals ε_r for a coaxial cable since the field runs fully in the dielectric.

The coaxial resonator can also be a ceramic filled transmission line with a reflecting termination. by using a ceramic dielectric material it is possible to achieve high ε_r and thereby a shorter physical length of the resonator.

The procedure for designing a coaxial resonator can in short be described as

- Determine the desired inductance and frequency
- Select a coaxial cable with low loss at the frequency with given Z_0 .
- Use the Smith chart or $Z_{in} = Z_0 j\tan(\beta <)$ to determine $<$. Observe that the wavelength is the wavelength in the dielectric of the coaxial cable.

10.7.3 Crystals

For many oscillator applications it is crucial to have a very exact frequency reference. For a 900 MHz radio system with 25 kHz channel separation, a 25 ppm transmitter frequency error would result in transmission on the adjacent channel. Such high precision is impossible to realise using a coil, coaxial cable or microstrip along with a capacitance as resonator. One solution to achieve better frequency accuracy as well as stability is to use a crystal as frequency detemining element. A crystal is a piece of quartz that has been ground to a certain shape that makes it resonant at a certain frequency. When applying a signal to the quartz crystal the piezo electric effect causes a mechanical movement of the crystal. The frequency response of the crystal is modelled by a serial and parallel resonant circuit as in figures 10.15 and plotted in figures 10.16.

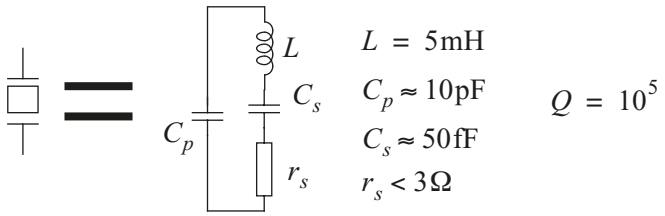


Figure 10.15 10 MHz quartz crystal and its RLC equivalent.

The crystal has one series resonant and one parallel resonant frequency that are very close together. In between is it inductive with an inductance value that increases with frequency. This close-to-resonance-phenomenon is actually a benefit of the crystal.

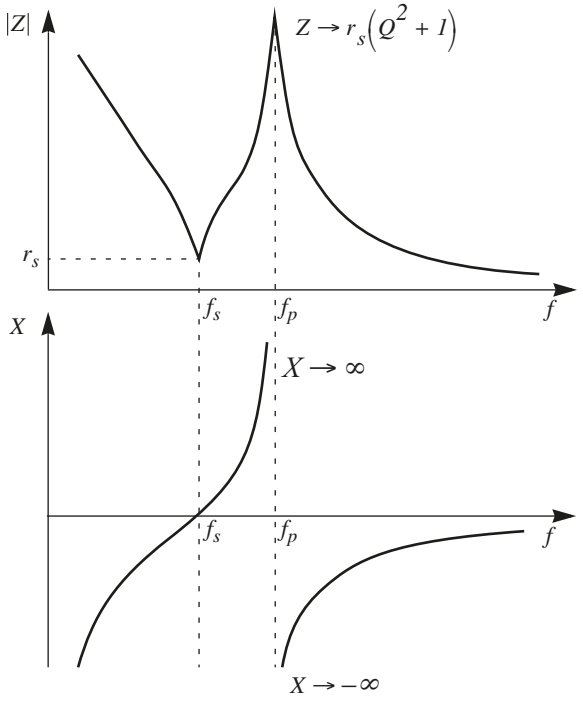


Figure 10.16 Impedance and reactance of a crystal. Series resonance is at f_s and the parallel resonance at f_p.

If we look at the Pierce crystal oscillator in figure 10.17 and compare that to the general oscillator model from figure 10.2, we can see that the crystal must behave like an inductance. If the temperature changes and along with that the values of the capacitors this would in the LC-oscillator case cause a permanent frequency drift. But here, if the frequency starts to drift upwards the inductance value of the crystal will increase and reduce the frequency drift. When the frequency change from 10MHz to 10.001 MHz the inductance increase from 30μH to 36μH . The oscillator will be "self adjusting", though a small frequency error will always be present if conditions are changed.

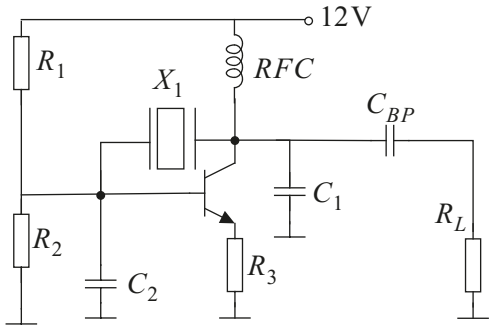


Figure 10.17 Pierce crystal oscillator.

If this frequency accuracy is not enough it can be further enhanced:

- The crystal can be placed in an $80°C$ oven to keep its operating temperature constant. The reason for using this temperature is that crystals cut in a certain way have minimum temperature sensitivity there.
- The oscillator can have a temperature dependent control system for a varicap in the crystal oscillator. The varicap voltage is designed to compensate for the crystals temperature drift.
- Crystals can be pre-aged. Over the first operating time the drift is usually rather large and by aging the crystal this can be avoided.

Even if it would be desirable, a crystal stamped 10 MHz is not exactly resonant at that frequency. This is because the designer might want to use the crystals in different oscillating modes and also be able to tune it to exactly 10MHz disregarding spread in the manufacturing.

$$f_s = \frac{1}{\sqrt{2\pi L C_s}} < 10\text{MHz} \tag{10.26}$$

$$f_p = \frac{1}{\sqrt{2\pi L \dfrac{C_s C_p}{C_s + C_p}}} > 10\text{MHz} \tag{10.27}$$

The tuning is accomplished by adding a small capacitor, either in series to increase f_s or in parallel to reduce f_p , see figure 10.18.

Figure 10.18 Tuning the crystal with an external capacitor.

This will not tune over a very large range and if the capacitors are improperly chosen the total impedance will show low dependency of the crystal. An appealing example is if one makes the shunting tunable capacitance too large, this will result in a bypass of all high frequencies and in the worst case the oscillator will not run. Even if it will run, it will not be crystal controlled if the crystal impedance is high compared to the shunting capacitance impedance at the desired frequency.

10.8 References

[1] G. D. Vendelin, *Design of Amplifiers and Oscillators by the S-parameter method*, John Wiley & Sons Inc. 1982.

[2] J. Davidse, *Analog Electronic Circuit Design*, Prentice-Hall, 1991.

[3] D.B. Leeson, "A simple model of feedback oscillator noise spectrum", *Proceedings of the IEEE*, pp. 329-330, 1966.

[4] M.J. Underhill, "Fundamentals of oscillator performance", *Electronics & Communication Engineering Journal*, pp. 185-193, Aug. 1992.

[5] S. L. Liao, *Microwave circuit analysis and Amplifier Design*, Prentice-Hall, 1987.

[6] G. Gonzales, *Microwave Transistor Amplifiers analysis and design*, 2nd edition, Prentice Hall, 1997.

Chapter 11

Directional Couplers and Ferrite Devices

The concept of travelling waves allows us to consider independent waves travelling in opposite directions along a transmission line. Equally, a port of a circuit can both receive and emit waves. This is not only useful in analysis and design of high frequency circuits. Waves propagating in opposite directions can carry different information and it is clearly seen that circuits that can couple or isolate waves between ports depending on the direction will be very useful. Directional couplers constitute one class of such circuits to which this chapter is mainly devoted. While coupling or distribution of waves is one important application the circuits may also be used for combining. The physical realisation to a large extent determines the suitability in this respect.

One basic configuration of a directional coupler is illustrated in figure 11.1 where two microstrip transmission lines are separated with a small slot. Waves entering port 1 and 2 will be coupled differently to the other ports. The length of the lines, the width of the slot, the substrate and the frequency determine the coupling properties.

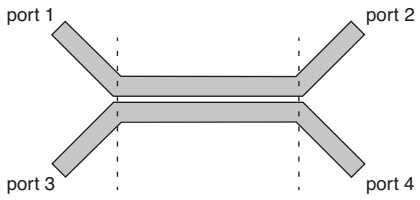


Figure 11.1 Directional coupler.

Another and important class of circuits is the non-reciprocal ferrite devices. Here a constant magnetic field is applied to the ferrite material which then exhibit different permeabilities for oppositely rotating magnetic fields perpendicular to the constant magnetic field. This technique can for example be used to implement devices that permits power flow in one direction only or a circulator that can be said to act as a commutator for power. Such devices are described briefly in the end of this chapter.

11.1 Directional Couplers

All basic configurations of directional couplers are 4-port circuits. These are reciprocal devices and they can be used both as power dividers and combiners. A generic symbol as the one shown in figure 11.2 is used to describe the properties of a 4-port coupler with power coupling and phase given for each port relative to port 1, the input port.

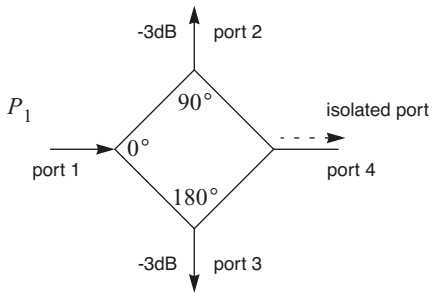


Figure 11.2 Generic symbol for a 4-port power divider. Power coupling and phase values are given as an example for an ideal quadrature hybrid (see section 11.3.1).

The ports found in figure 11.2 are commonly denoted in the following way:

- port 1 - input port
- port 2 - direct port
- port 3 - coupled port
- port 4 - isolated port.

Here, port 2 is assumed to be the main output port to which most of the power will be fed if we have unequal power division between ports 2 and 3. Theoretically, the isolated port will not emanate any power at all. Although, in practice a small amount of power will go through this port as well. This leakage is due to the non-ideal behaviour of real transmission line structures that originates from dispersive properties and structural mismatching. In connection with this discussion it is suitable to define five important quantities in which a coupler may be specified.

$$\text{Transmission } T = 10\log\left(\frac{P_2}{P_1}\right) \quad [\text{dB}] \tag{11.1}$$

$$\text{Coupling } C = 10\log\left(\frac{P_3}{P_1}\right) \quad [\text{dB}] \tag{11.2}$$

$$\text{Isolation } L = 10\log\left(\frac{P_4}{P_1}\right) \quad [\text{dB}] \tag{11.3}$$

$$\text{Directivity } D = 10\log\left(\frac{P_4}{P_3}\right) \quad [\text{dB}] \tag{11.4}$$

$$\text{Return loss } R = 10\log\left(\frac{P_1'}{P_1}\right) \quad [\text{dB}] \tag{11.5}$$

where P_1 is the power fed to port 1, P_1' is the power reflected back from the input port and P_2, P_3 and P_4 are the output power levels from ports 2,3 and 4, respectively.

For an ideal coupler with equal power division we have $T = -3\text{dB}$, $C = -3\text{dB}$, $D = 0$, $L = 0$ and $R = 0$. In practice the isolation will be in the range from -20dB to -40dB. Another common term is insertion loss that is the negated value of transmission (in dB).

Directional couplers can be divided into two groups; field-coupled couplers and branch-line couplers. The former was briefly introduced in the beginning of the chapter. The coupling effect is obtained through electromagnetic coupling only between closely spaced structures. The latter relies on direct coupling through distributed or lumped circuit elements.

Below a survey of different configurations is presented based both on distributed and lumped circuit elements. For a more detailed treatment there are several comprehensive books in this area that include both analysis and design equations. Two of these are [1] and [2] that mainly deal with distributed circuit solutions. For lumped circuit solutions refer to [3] or [4].

11.2 Field-Coupled Circuits

Two transmission lines that are brought close to each other will act as a coupling device, see figure 11.3. The coupling factor can hardly reach those of the branch-line couplers, say equal power division (-3dB), at least not without special arrangements. Typical values are, say, -5 dB or lower. In figure 11.3 a typical layout is exemplified with port numbers indicated as defined above. In this example the length of the coupled lines are given to $\lambda/4$, which corresponds to maximum coupling with all other parameters fixed.

This kind of device is useful for example in power measurements, SWR measurements or non-uniform power combining/division. Some general aspects can be outlined to introduce parameters and relations common for coupled lines without going into details about realisation which will be individual for different transmission line structures. A detailed treatment is out of the scope for this text.

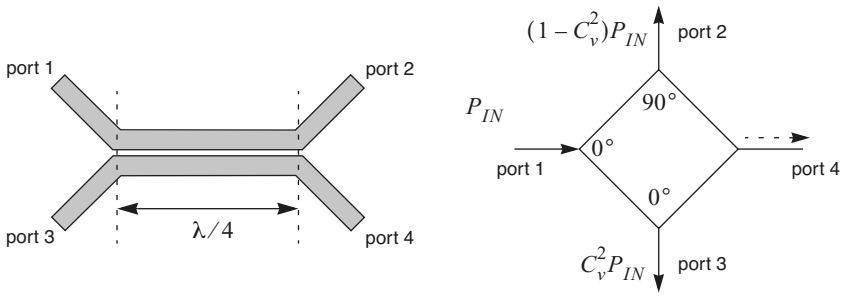


Figure 11.3 Example of layout for field-coupled circuit (directional coupler) and its phase and power relations.

11.2.1 Coupling between Parallel Lines

To appreciate the function of coupled transmission lines it is a common procedure to investigate two modes of excitation, even-mode and odd-mode excitation, see figure 11.4. The effect of coupling in the two excitation modes can be modelled as different characteristic impedances of the line sections under observation. This is readily understood if we consider the capacitive coupling between the two lines. From earlier chapters we know that the capacitance of a transmission line has a large influence on the characteristic impedance.

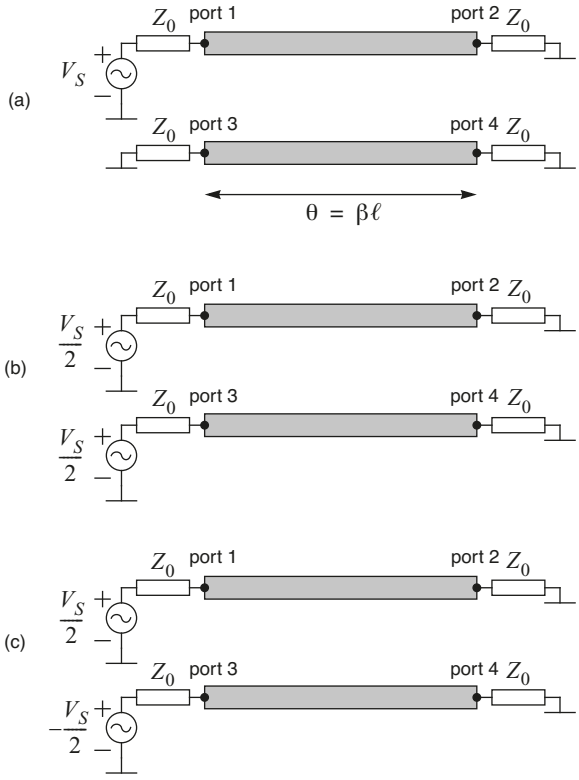


Figure 11.4 (a) Parallel-coupled transmission lines with one input signal at port 1 (b) even-mode excitation (c) odd-mode excitation.

In even-mode excitation the voltage along the lines will be exactly the same on both lines and thus there is no capacitive coupling between them. In odd-mode excitation the voltages along the lines will be equal in amplitude but with opposite signs and consequently a capacitive coupling will be observed. Thus, we can define two characteristic impedances, Z_{0e} and Z_{0o} where subscript e and o denote the even-mode and odd-mode, respectively. Once Z_{0e} and Z_{0o} are known it is possible to calculate the voltage coupling to other ports as was done for the quadrature hybrid and apply superposition to find the total contribution at the ports when a signal is fed to port 1 only. A derivation for the coupled transmission lines can be found in [1] and the result is presented below.

For perfect matching, when a signal is fed to port 1, we require that $Z_{in1} = Z_0$ and it can be shown that this condition is satisfied when

$$Z_0 = \sqrt{Z_{0o}Z_{0e}} \tag{11.6}$$

As a matter of fact this relation holds for any length of the coupler.

The voltage coupling from port 1 to the other ports are given by [5]:

$$\frac{V_2}{V_1} = \frac{\sqrt{1-C_v^2}}{\sqrt{1-C_v^2}\cos\theta + j\sin\theta} \tag{11.7}$$

$$\frac{V_3}{V_1} = \frac{jC_v\sin\theta}{\sqrt{1-C_v^2}\cos\theta + j\sin\theta} \tag{11.8}$$

$$\frac{V_4}{V_1} = 0 \tag{11.9}$$

where C_v is the maximum voltage coupling (from port 1 to port 3) which occurs for $\theta = \pi/2 + n\pi$ where n is an integer. It is worth noting that the input signal is coupled to port 3 which emits a wave in the opposite direction with reference to the input wave. For this reason it is common to refer to this structure as a backward-wave coupler. The phase and the power relations are shown in figure 11.5.

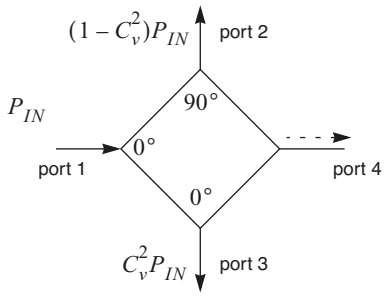


Figure 11.5 Phase and power relations for the backward-wave coupler.

It can be shown that the even-mode and the odd-mode characteristic impedances are related to C_v such that

$$C_v = \frac{Z_{oe} - Z_{oo}}{Z_{oe} + Z_{oo}} \tag{11.10}$$

In other words Z_{0e} and Z_{0o} together with the length of the coupler completely determine the electrical properties of the coupler. However, in most practical cases the length is typically chosen to be $\lambda/4$ ($\theta = \pi/2$) at midband for maximum coupling or flatness.

In summary, to design a backward-wave coupler we have to specify the midband operating frequency, f_0 , the port impedance Z_0 and the coupling factor at midband C_v ($20\log(C_v)$ dB). If the length is chosen to be $\lambda/4$ at midband the even-mode and odd-mode characteristic impedances can be calculated from

$$Z_{0e} = Z_0\sqrt{\frac{1+C_v}{1-C_v}} \tag{11.11}$$

$$Z_{0o} = Z_0\sqrt{\frac{1-C_v}{1+C_v}} \tag{11.12}$$

and these are used as input parameters when realising the actual transmission line structure either by consulting tabulated data as in [1] or by using a dedicated software CAD-tool.

11.2.2 Lange Coupler

To increase the coupling factor beyond what it possible with the basic backward-wave coupler Lange [7] proposed a method where several parallel lines (4 or more) with alternating lines tied together with bonding wires as illustrated in figure 11.6. This technique can reach a -3dB coupling factor or less. Therefore it can be used for example in balanced power amplifier designs where the signals from different amplifier stages are combined. One advantage with the Lange coupler over the branch-line couplers described below is that it has no DC-coupling between ports 2 and 3. For example, if there are different DC circuits connected to each port for, say transistor biasing, no additional components will be required for DC isolation. Details on how to design Lange couplers can be found in [8].

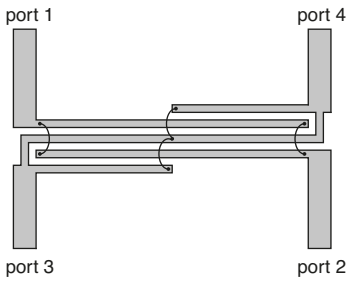


Figure 11.6 Example of Lange coupler structure (not to scale).

11.3 Branch-Line Couplers

To obtain tighter coupling and higher power handling capabilities the ports of a coupler must be connected together with branches instead of relying solely on electromagnetic coupling. A number of structures are described below and these are suitable for power combining and dividing at high power levels.

11.3.1 Quadrature Hybrid

The quadrature hybrid is a 4-port circuit and the basic structure is shown as a microstrip layout in figure 11.7. It consists of four $\lambda/4$ lines, two with the characteristic impedance Z_0 and two with $Z_0/\sqrt{2}$. All ports are typically terminated with Z_0. A functional-equivalent layout of the quadrature hybrid is shown in figure 11.8 to demonstrate that these structures can be altered to suit specific needs without significantly affecting the functionality.

The phase and power relations are shown in figure 11.9 with a signal fed to port 1. The output signals appears at ports 2 and 3 in equal portions (-3dB coupling factor). Port 4 is the isolated port. Since the circuit is reciprocal it is also possible to feed signals with equal amplitude but with $90°$ phase difference to ports 2 and 3 and obtain a combined signal at port 1 but nothing in the isolated port. On the other hand if the phase difference between the signals is $-90°$ instead the combined signal will appear in the isolated port. If two signals with equal amplitude are fed to ports 2 and 3 without any phase difference it is readily understood that due to symmetry the signals will propagate through the circuit and appear in equal amounts and with equal phase shifts at ports 1 and 4. Finally, if two uncorrelated signals are fed to ports 2 and 3 they will be combined and appear in equal portions at ports 1 and 4. In all cases the port impedances will be equal to Z_0 if all ports are terminated with Z_0.

In the case where a signal is fed to one port only, as described above, the isolated port does not have to be terminated with Z_0 because no signal appears on that port. Of course, this will no longer be true if there will be mismatch at the output ports. Such a mismatch will result in reflected waves that must be absorbed.

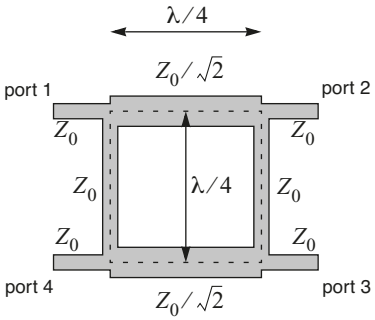


Figure 11.7 Microstrip layout of the quadrature hybrid (not to scale).

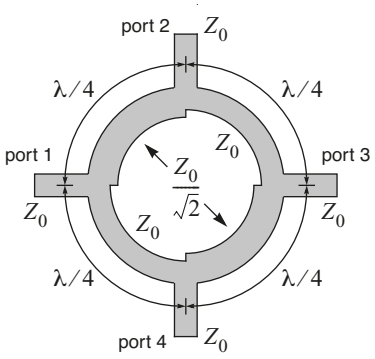


Figure 11.8 Alternative layout of the quadrature hybrid (not to scale).

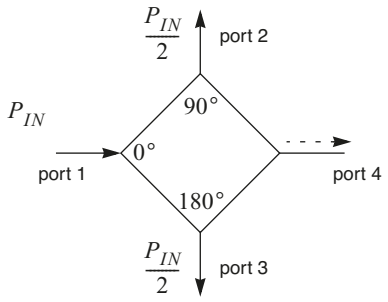


Figure 11.9 Phase and power relations for the quadrature hybrid.

The bandwidth of the quadrature hybrid is fixed but it can be extended with more advanced layouts. Furthermore, non-symmetrical versions can also be designed to obtain non-uniform power division with preserved port impedances.

In example 11.1 below it is shown how these kind of circuits can be analysed by exploiting the symmetry.

Example 11.1 Analysis of the quadrature hybrid

Consider the quadrature hybrid in figure 11.10. To analyse this circuit we will use superposition of two excitation modes that leads to the case in figure 11.10. These modes are the even-symmetry-mode where ports 1 and 4 are fed with equivalent sources and the odd-symmetry-mode where ports 1 and 4 are fed using sources in anti-phase, see figure 11.11.

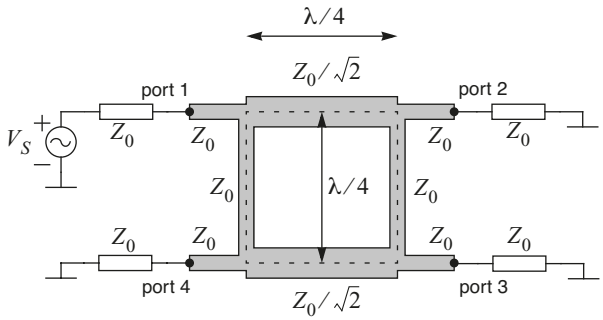


Figure 11.10 Power division using a quadrature hybrid.

We start by defining the system characteristic admittance $Y_0 = 1/Z_0$ because admittances are more convenient in this case. Because of the symmetry the quadrature hybrid can be considered as two separate circuits, one between ports 1 and 2 and one between ports 3 and 4. Furthermore, the branch lines between ports 1 and 4 and 2 and 3, respectively turn into stubs with a length of $\lambda/8$. For the even-symmetry mode the stubs are open-circuited which means that they will have a susceptance equal to $+jY_0$. Similarly, for the odd-symmetry mode the stubs are short-circuited and will have a susceptance equal to $-jY_0$.

Now, superposition applies to currents and voltages so we need to calculate the resulting voltages in each port or the current through the terminating impedances. To do that we need to investigate how currents and voltages are transformed from a source to load over a transmission line. Consider the circuit in figure 11.12.

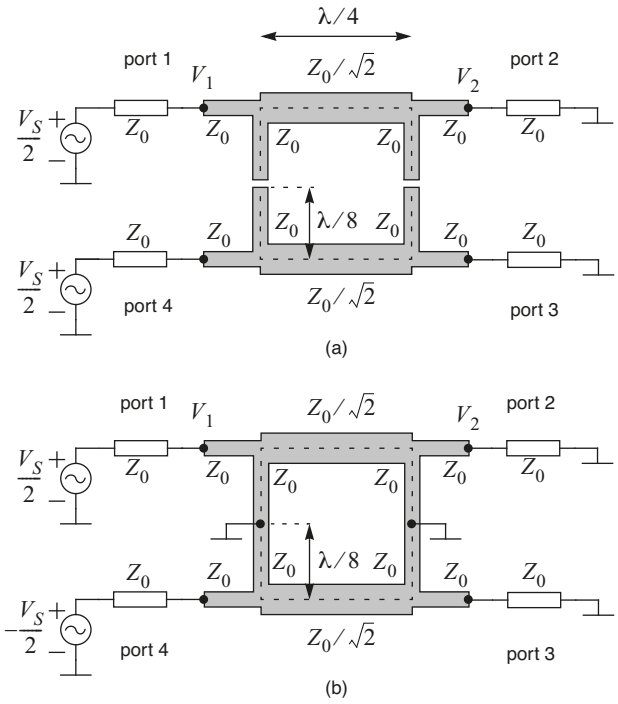


Figure 11.11 (a) even-symmetry-mode (b) odd-symmetry-mode.

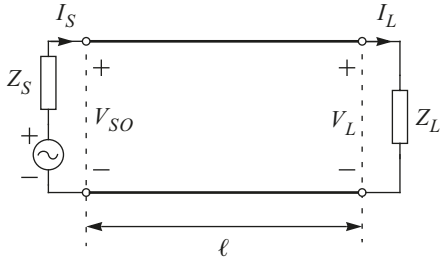


Figure 11.12 Schematic to investigate transformation of currents and voltages over a transmission line.

The source output voltage V_{SO} is the sum of the incident and the reflected wave and these waves in turn can be substituted with waves at the load end of the circuit as

$$\begin{aligned} V_{SO} &= V_{SO}^{+} + V_{SO}^{-} \\ &= V_L^{+} e^{j\beta\ell} + V_L^{-} e^{-j\beta\ell} \\ &= (V_L^{+} + V_L^{-})\cos\beta\ell + j(V_L^{+} - V_L^{-})\sin\beta\ell \end{aligned} \tag{11.13}$$

We also know that the wave currents and voltage are related through the characteristic impedance as

$$V_L^+ = I_L^+ Z_0 \tag{11.14}$$

and

$$V_L^- = I_L^- Z_0 \tag{11.15}$$

Thus we have that $(V_L^+ - V_L^-) = (I_L^+ - I_L^-)Z_0 = I_L Z_0$ and (11.13) can be rewritten as

$$V_{SO} = V_L \cos\beta\ell + jI_L Z_0 \sin\beta\ell \tag{11.16}$$

which relates the source voltage with the load voltage and current. We can deal with the source current in a similar manner:

$$\begin{aligned} I_S &= I_S^+ - I_S^- \\ &= I_L^+ e^{j\beta\ell} - I_L^- e^{-j\beta\ell} \\ &= (I_L^+ - I_L^-)\cos\beta\ell + j(I_L^+ + I_L^-)\sin\beta\ell \\ &= I_L \cos\beta\ell + j\frac{V_L}{Z_0}\sin\beta\ell \end{aligned} \tag{11.17}$$

These results will now be used to deal with the even-symmetry mode in more detail, see figure 11.11a. To equate this circuit and the one in figure 11.12 we should let the stubs in both ends be included in the source and load as illustrated in figure 11.13.

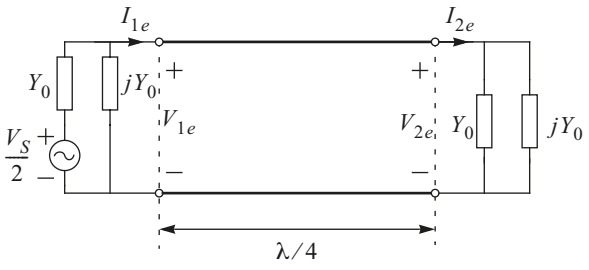


Figure 11.13 Equivalent circuit for one branch in figure 11.11a, even-symmetry mode.

In this special case the electrical length of the transmission line $\beta\ell$ equals $90°$. By applying (11.16) to this circuit and noting that the characteristic admittance of the transmission line is $\sqrt{2}Y_0$ we get

$$\begin{aligned} V_{2e} &= -j\frac{I_1}{\sqrt{2}Y_0} \\ &= -j\frac{V_1 Y_1}{\sqrt{2}Y_0} \\ &= -j\frac{V_S(1-j)}{2\sqrt{2}} \end{aligned} \tag{11.18}$$

In the same way we find the voltage at port 2 for the odd-symmetry mode to be

$$V_{2odd} = -j\frac{V_S(1+j)}{2\sqrt{2}} \tag{11.19}$$

It is easy to show either analytically or by using a Smith chart that all ports have the same input admittance, namely Y_0. Thus the voltage in the ports that are connected to the sources we simply be half of the internal voltage of the source.

We are now prepared to calculate the port voltages by adding the contributions from the even-symmetry and odd-symmetry modes. We begin with port 1:

$$\begin{aligned} V_1 &= V_{1e} + V_{1o} \\ &= \frac{1}{2}\left(\frac{1}{2}V_S\right) + \frac{1}{2}\left(\frac{1}{2}V_S\right) = \frac{V_S}{2} \end{aligned} \tag{11.20}$$

Similarly, for port 4 we get

$$\begin{aligned} V_4 &= V_{4e} + V_{4o} \\ &= \frac{1}{2}\left(\frac{1}{2}E_S\right) + \frac{1}{2}\left(-\frac{1}{2}E_S\right) = 0 \end{aligned} \tag{11.21}$$

In other words, no power is dissipated in the isolation port (4) when a signal is fed to port 1. For the two remaining ports (11.18) and (11.19) gives

$$\begin{aligned} V_2 &= V_{2e} + V_{2o} \\ &= -\frac{j(V_S/2)(1-j)}{2\sqrt{2}} + \frac{-j(V_S/2)(1+j)}{2\sqrt{2}} \\ &= \frac{-jV_S}{2\sqrt{2}} \end{aligned} \tag{11.22}$$

and

$$\begin{aligned} V_3 &= V_{3e} + V_{3o} \\ &= -\frac{j(V_S/2)(1-j)}{2\sqrt{2}} + \frac{-j(-V_S/2)(1+j)}{2\sqrt{2}} \\ &= \frac{-V_S}{2\sqrt{2}} \end{aligned} \tag{11.23}$$

Consequently, the voltage coupling factors for ports 2 and 3 with port 1 as input port become

$$C_{v12} = \frac{V_2}{V_1} = \frac{-j}{\sqrt{2}} \tag{11.24}$$

$$C_{v13} = \frac{V_3}{V_1} = -\frac{1}{\sqrt{2}} \tag{11.25}$$

which confirms the power and phase relations in figure 11.9.

11.3.2 Wilkinson Hybrid

At first sight the Wilkinson hybrid looks like a 3-port circuit, see figure 11.14. However, a 4th port is present internally in differential form between ports 2 and 3 that acts as the isolated port. Similar to the basic structure of the quadrature hybrid the Wilkinson hybrid is fully symmetrical. However, it is also symmetrical with respect to port 1 which means that a signal that is fed to port 1 is split in two equal parts that appear at ports 2 and 3 but in contrast to the quadrature hybrid there is no phase difference, see figure 11.15. Note that the resistor in the isolated part has no influence as long as the signals are the same at ports 2 and 3. The Wilkinson hybrid is reciprocal and can be used both as a power divider and power combiner.

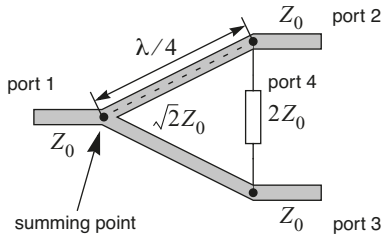


Figure 11.14 Wilkinson hybrid (not to scale).

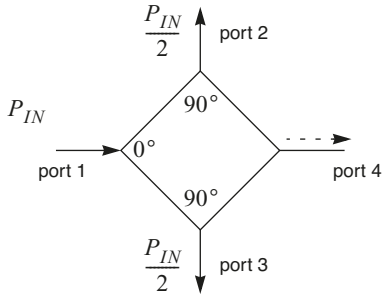


Figure 11.15 Phase and power relations for the Wilkinson hybrid.

The layout in figure 11.14 is the common way of drawing the Wilkinson hybrid for the sake of clarity. However, it is not a practical layout since the termination resistor for the isolated port is a lumped and very small component. In other words, in a practical design the connection points for this resistor must be close and the branch lines from ports 2 and 3 to port 1 must therefore either be very close to each other or bent as exemplified in figure 11.16.

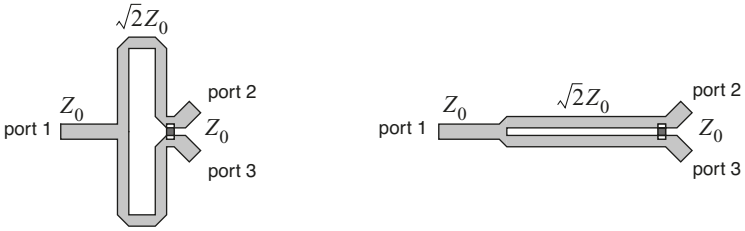


Figure 11.16 Practical layouts for the Wilkinson hybrid (not to scale).

As was the case for the quadrature hybrid there exist derivatives from the original structure that provide larger bandwidths and non-uniform power division. Furthermore, the Wilkinson hybrid can be generalised to have any number of ports [6].

11.3.3 Hybrid-Ring Coupler

Another common structure is the hybrid-ring coupler, also called rat-race coupler. For proper operation the characteristic impedance in the ring line should be $\sqrt{2}Z_0$. Similar to the quadrature hybrid and the Wilkinson hybrid this one also performs half-power coupling, for example from port 1 to ports 2 and 3. An incident travelling wave from port 1 will be split in two waves. At ports 2 and 3 the waves will be recombined in-phase whereas at port 4 they will be in anti-phase. Thus, no power appears at port 4. The power and phase relations are shown in figure 11.18. Note that the phase difference between ports 2 and 3 is $180°$.

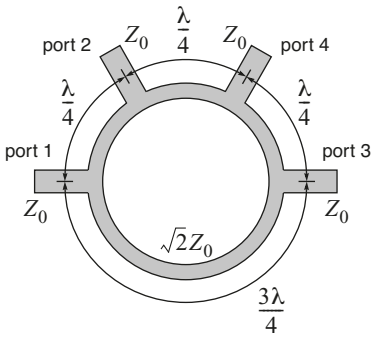


Figure 11.17 Hybrid-ring (Rat-Race) coupler (not to scale).

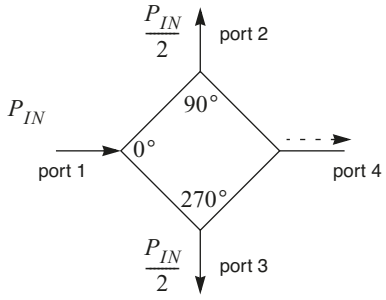


Figure 11.18 Phase and power relations for the hybrid-ring coupler.

11.4 Couplers with Lumped Components

The use of distributed circuit elements for power combiners and dividers primarily manifest itself at microwave frequencies, say above 1GHz. Below 1GHz these line structures tend to become to big for practical designs. However, in the sub-GHz range lumped circuit elements have acceptable parasitic components. This suggest that we can realise power combiners and dividers using lumped components instead. A lumped equivalent to the quadrature hybrid described in section 11.3.1 is shown in figure 11.19.

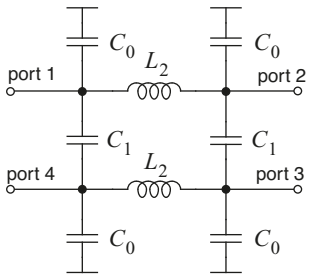


Figure 11.19 Lumped equivalent of the quadrature hybrid.

The power and phase relations shown in figure 11.9 for the quadrature hybrid holds for this circuit as well if it is designed for half-power coupling. The design equations are as follows.

$$C_1 = \frac{1}{Z_0 \omega_0 \sqrt{K}} \tag{11.26}$$

$$C_0 = -C_1 + \sqrt{C_1^2 + \frac{1}{\omega_0^2 Z_0^2}} \tag{11.27}$$

$$L = \frac{2C_0 Z_0^2}{1 + \omega_0^2 C_0^2 Z_0^2} \tag{11.28}$$

where ω_0 is the desired frequency of operation. K is the ratio of the power distribution between ports 2 and 3. In other words this configuration can be designed for non-uniform power division and still keep all port impedances equal to Z_0.

The Wilkinson coupler can be realised in a similar manner but with less flexibility with respect to the coupling, which in this case is fixed to -3dB. The circuit is shown in figure 11.20. The design equations for this structure is given by

$$L = \frac{\sqrt{2} Z_0}{\omega_0} \tag{11.29}$$

$$C = \frac{1}{\omega_0 \sqrt{2} Z_0} \tag{11.30}$$

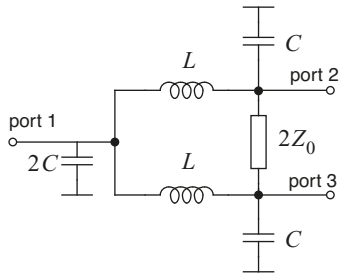


Figure 11.20 Lumped equivalent of the Wilkinson coupler.

Another example of a lumped realisation is shown in figure 11.21 that exploits a transformer with a centre tap. Here two input signals are combined and fed to the output. Signals in anti-phase will be combined and fed to the output load whereas signals in phase will be absorbed by the internal termination impedance. Since we have the freedom to choose the number of turns in the transformer we can also use this circuit as a wideband impedance transformer. If we consider it as a power combiner the input impedance for ports 1 and 2 will be

$$R_{i1,\,i2} = 2R_3\left(\frac{m}{n}\right)^2 \tag{11.31}$$

and the internal termination impedance should be chosen to be

$$R_4 = R_3\left(\frac{m}{n}\right)^2 \tag{11.32}$$

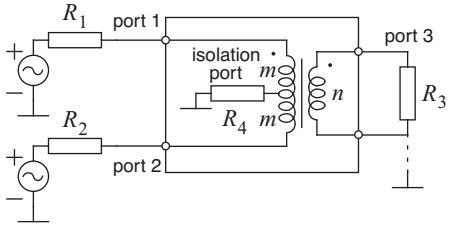


Figure 11.21 Transformer-based power combiner/divider with source and load.

11.5 Ferrite Devices

Ferrite materials are very useful in high frequency applications mainly due to their non-reciprocal properties. Ferrites are magnetic dielectrics which means that the have magnetic properties similar to ferromagnetic metals but the high resistivity of dielectrics. The latter property is important since high resistivity prevents the generation of eddy currents in contrast to ferromagnetic metals where these currents increase with increasing frequency.

The magnetic properties of ferrites stem from the electron spin generating magnetic dipole moments. When the ferrite medium is magnetised, e.g. by using either a permanent magnet or a electromagnet, the permeability will turn from a scalar quantity to a tensor. This effectively results in different permeabilities in different directions. Thus, the phase velocity given by

$$v_p = \frac{c}{\sqrt{\varepsilon_r \mu_r}} \tag{11.33}$$

will also be different for different directions (c is the free-space velocity of light, ε_r the dielectric constant and μ_r the relative permeability).

Also, if we consider the polarisation of a electromagnetic wave the permeability will be different depending on whether the wave is right-hand-circular or left-hand-circular polarised. To understand what happens when a linear polarised wave enters a magnetised ferrite (see figure 11.22) we can use the fact that a linear polarised wave can be decomposed into one right-hand-circular band and one left-hand-circular polarised wave.

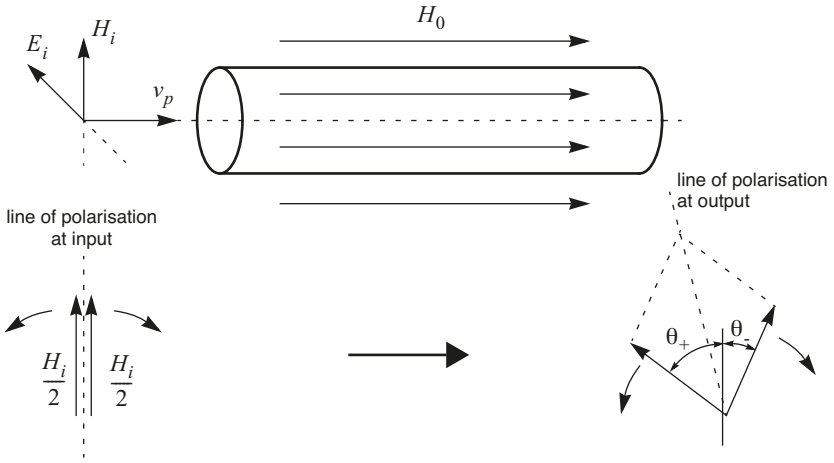


Figure 11.22 Linear polarised wave propagating through ferrite rod with the magnetic wave decomposed into two circular polarised waves.

In figure 11.22 we see that the line of polarisation (for the magnetic field) is shifted. This effect is nonreciprocal and is termed Faraday polarisation shift.

Many of the ferrite components are based on that signals propagate in closed-boundary waveguides (see e.g. [9] and [10]) where the fields are inside the waveguide rather than as for open-boundary waveguides/transmission lines such as microstrip structures where the fields are outside the boundary of the conductors. As such their use is limited since the dimensions of a closed-boundary waveguide cross section are on order of the wavelength of the signal. The circulator is one of the most useful ferrite components and it is described in the next section. Another example of ferrite components is the differential phase shifter. This is a two-port device based on Faraday rotation where the phase shift from one port to the other is different for the two directions. The gyrator is a special case of a differential phase shifter with a differential phase equal to 180 degrees. Ferrites can also be used to implement resonators with very high Q-values (up to 10^4) and substrates with low loss for microwave integrated circuits. For a more detailed description of the theory behind ferrites please refer to [9] and [10].

11.5.1 Circulators

A circulator is commutator for power such that an incident wave at one port will be transmitted only to the next port in the sequence of ports. This is illustrated by the circuit symbol for a three-port and a four-port circulator in figure 11.23. Power entering port 1 will be transmitted to port 2, power entering port 2 will transmitted to port 3 etc. Circulators with more ports can easily be obtained by combining circulators as illustrated in figure 11.24.

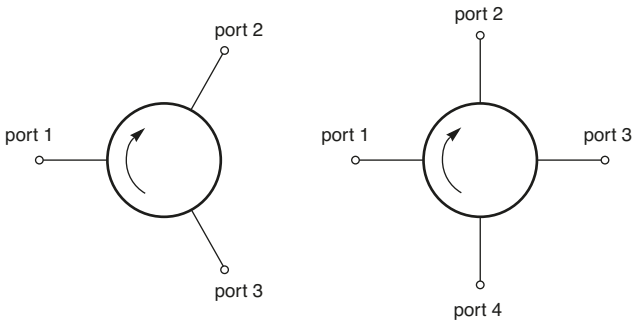


Figure 11.23 Circuit symbols for a three-port and a four-port circulator.

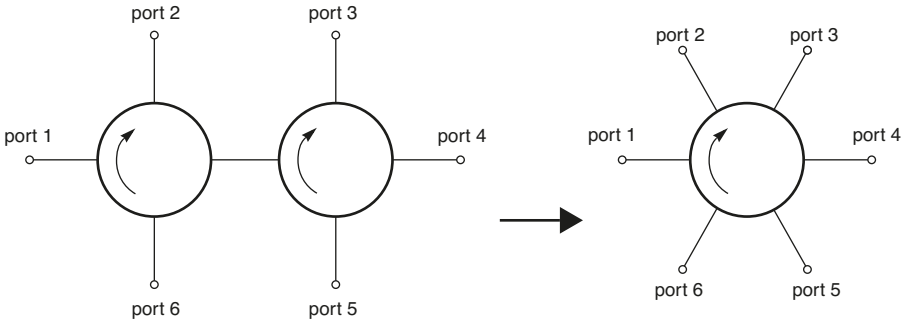


Figure 11.24 Implementation of 6-port circulator using two four-port circulators.

One important application for the circulator is found when a receiver and transmitter chain is to be connected to a common antenna, see figure 11.25. Here the high power transmitter amplifier produces a signal that should be forwarded to the antenna but not to the receiver if the both are active at the same time. Due to the circulator properties the transmitted signal does not (ideally) reach the receiver port whereas the signal received by the antenna will. For the same reason the circulator serves as an isolator for the transmitter amplifier such that if there is a mismatch at the antenna the reflected signal will not return to the amplifier but to the receiver (for which there is no protection). Thus the amplifier will see a constant impedance even if there is a mismatch at the antenna port. This is in many cases extremely important since an unknown load impedance could otherwise force the amplifier into oscillation.

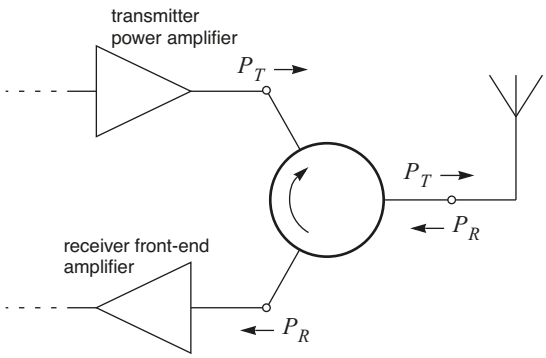


Figure 11.25 Connecting the transmitter and receiver chain to a common antenna using a circulator.

In the example above the circulator acts as a isolator for the amplifier and of course the circulator can be configured as a two-port isolator using a dummy load at one of the ports as illustrated in figure 11.26. The purpose of the dummy load is to dissipate any power reflected back to the output port.

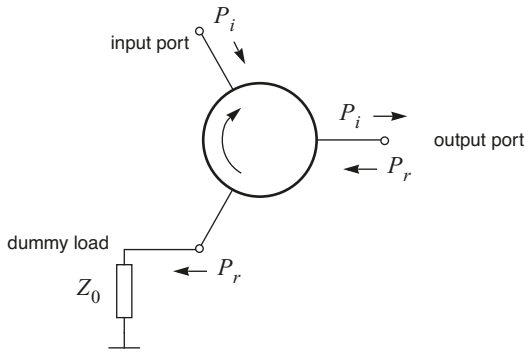


Figure 11.26 Circulator configured as an isolator.

There are several ways to implement circulators. For open-boundary transmission lines (microstrip or stripline structures) the Y-junction circulator is useful. It is a three port device where the three ports are connected together with a circular centre conductor plate with a ferrite disc on top for microstrip structures and ferrites discs on both sides for stripline structures, see figure 11.27.

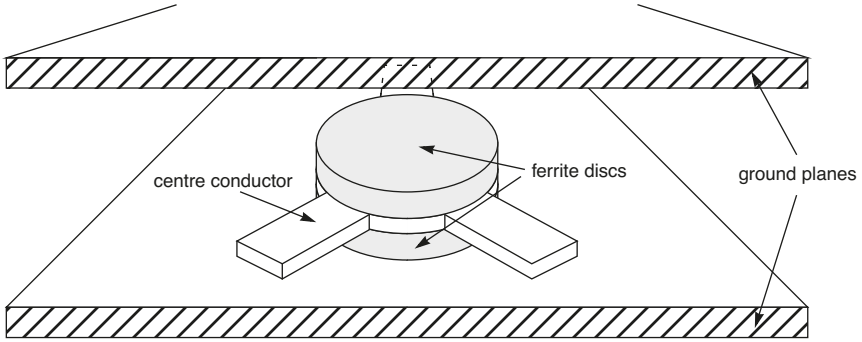


Figure 11.27 Y-junction stripline circulator.

The principle of this circulator is briefly illustrated in figure 11.28. The ferrite disc can be excited to resonate in which case a standing wave pattern will arise in the disc. This is shown in figure 11.28 without and with a constant magnetic field applied and with an incident wave at port 1. Without the magnetic field the standing wave is aligned with input port and part of the signal will be reflected back and port 2 and 3 will receive equal amounts of power. The standing wave pattern is generated by two contra-rotating waves. If a magnetic field is applied perpendicular to the disc plane the two contra-rotating waves will propagate will different phase velocities and therefore the resonant frequencies for the two rotating modes are separated. The effect will be a rotation of the standing wave ratio and with the magnetic field properly set the rotation will be 30 degrees which means that there will be no electric field and no transverse magnetic field at port 3. In practice, though, the isolation is

limited to 20-40dB. Matching networks are used at each port for optimal coupling between the ports and ferrite discs. A more detailed treatment of the Y-junction circulator is found in [11].

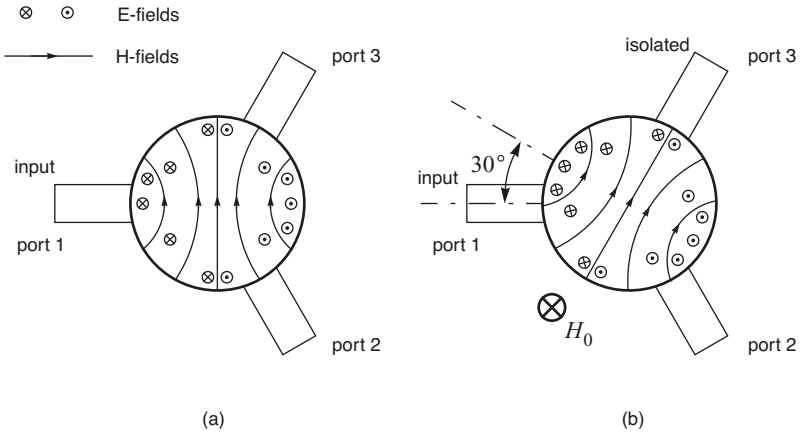


Figure 11.28 Standing wave pattern in the ferrite disc of a Y-junction circulator without (a) and with (b) a constant magnetic field applied.

11.6 References

[1] B. Bhat and S. K. Koul, *Stripline-like transmission lines for microwave integrated circuits*, John Wiley, 1989.

[2] G. Matthaei, L. Young and E. M. T. Jones, *Microwave filters, impedance-matching networks, and coupling structures*, Artech House, 1980.

[3] H. L. Krauss, C. W. Bostian and F. H. Raab, *Solid state radio engineering*, John Wiley, 1980.

[4] I. D. Robertson, *MMIC Design*, IEE, 1995.

[5] E. H. Fooks and R. A. Zakarevicius, *Microwave engineering using microstrip circuits*, Prentice Hall, 1990.

[6] E. J. Wilkinson, "An N-way hybrid power combiner", *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-8, No. 1, January 1960, pp. 116-118.

[7] J. Lange, "Interdigitated stripline quadrature hybrid", *IEEE Transactions on Microwave Theory and Techniques*, MTT-17, No. 12, December 1969, pp. 1150-1151.

[8] A. Presser, "Interdigitated microstrip coupler design", *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-26, No. 10, October 1978, pp. 801-805.

[9] D. Roddy, *Microwave Technology*, Prentice Hall, 1986.

[10] R. E. Collin, *Foundations for microwave engineering*, McGraw-Hill, 1966.

[11] C. E. Fay and R. L. Comstock, "Operation of the ferrite junction circulator", *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-13, No. 1, January 1965, pp. 15-27.

Chapter 12

Filters

Signal shaping and signal selection in the frequency domain are performed by means of filters. Filters are one of the basic building blocks in electronic systems. They can be used to alter spectral properties of signals or to simply select some portions of the frequency band and exclude others. In radio communication applications filters are commonly used for removal of strong signals that otherwise would interfere with wanted (and typically weak) signals.

There are many ways of specifying and implementing filters. This chapter will serve as a tutorial in classification of filters, filter modelling and filter implementation for high frequency applications.

12.1 Modelling

This chapter is restricted to filters that can be modelled electrically by a two-port network terminated with purely resistive source and load impedances, see figure 12.1. The filter transfers the signal source (port 1) to the load (port 2) with a frequency dependent transfer function both in terms of amplitude and phase. Theoretically, the filter is lossless, i.e., the filter consists of reactive components only.

The simple matching networks that was described in earlier chapters were based on achieving conjugate matching for one frequency only and from the topology chosen it was only possible to conclude its qualitative behaviour, i.e., if it had a lowpass, highpass or bandpass characteristic (see below), but the exact properties for other frequencies were not considered at all. In this context, a filter can be seen as a matching network whose spectral properties are known and selected beforehand.

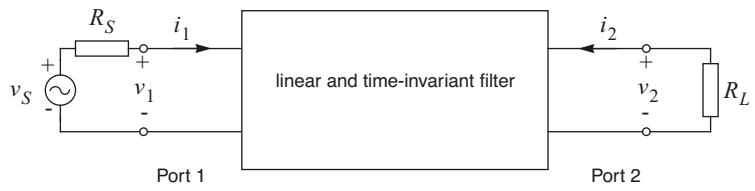


Figure 12.1 Filter modelled as a two-port network.

The filters that are considered here are time-invariant, linear and of finite order. Such a filter can be described by a transfer function, the quotient of the output signal and the input signal, given by a rational polynomial in the s-domain,

$$H(s) = K\frac{(s-z_1)\cdot(s-z_2)\ldots(s-z_n)}{(s-p_1)\cdot(s-p_2)\ldots(s-p_m)} \tag{12.1}$$

where $s = \sigma + j\omega$. Assuming a sinusoidal input with frequency f Hz, the transfer function is given by (12.1) with $s = j\omega = j2\pi f$. The transfer function has a constant gain factor, K , n zeros and m poles. The number of poles and zeros and their respective locations in the complex plane determine the characteristics of the filter. We can classify a filter by investigating the amplitude of the transfer function as a function of frequency, $A(f) = |H(j2\pi f)|$, and the phase function, $\phi(f) = \arg(H(j2\pi f))$.

In general, poles should be located in the left-half of the s-plane where the amplitude function, $A(f)$, should be large and zeros should be located on the imaginary axis where $A(f)$ should be small. Below, the various filter types are exemplified together with their respective pole-zero placement.

12.1.1 Lowpass Filter

A lowpass (LP) filter passes low frequencies to the load and rejects high frequencies. An example of pole placement and amplitude function for a lowpass filter is given in figure 12.2.

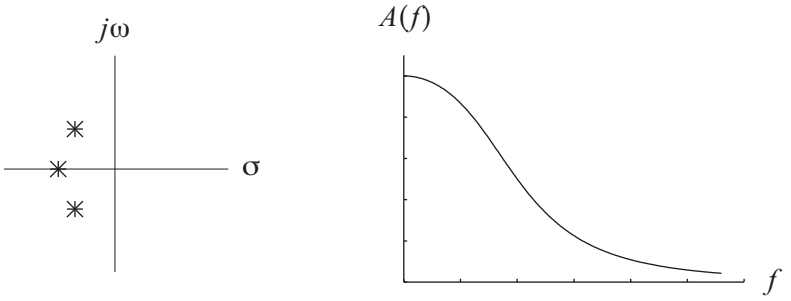


Figure 12.2 Pole placement and amplitude function for lowpass filter.

The transfer function for the lowpass filter in figure 12.2 is given by

$$H(s) = \frac{1}{(s+1)\cdot\left(s+\frac{1-i}{\sqrt{2}}\right)\cdot\left(s+\frac{1+i}{\sqrt{2}}\right)}$$

Of course, a LP filter can have zeros as well but since we want $H(j\omega) \to 0$ when $\omega \to \infty$ (high frequencies) we conclude that the number of poles must exceed the number of zeros.

12.1.2 Highpass Filter

A highpass filter (HP) is the very opposite of a lowpass filter, i.e., low frequencies are rejected and high frequencies passes with minor loss. An example of pole-zero placement and amplitude function for a HP filter is given in figure 12.3.

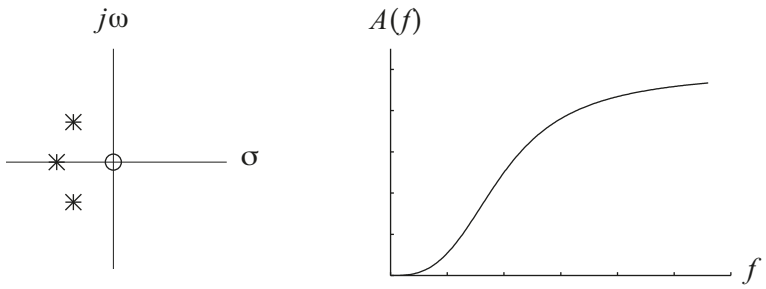


Figure 12.3 Pole placement and amplitude function for highpass filter.

The transfer function for the high pass filter in figure 12.3 is given by

$$H(s) = \frac{s^3}{(s+1)\cdot\left(s+\frac{1-i}{\sqrt{2}}\right)\cdot\left(s+\frac{1+i}{\sqrt{2}}\right)}$$

In a HP filter the number of poles is equal to the number of zeros and there is at least one zero in the origin.

12.1.3 Bandpass Filter

A bandpass filter (BP) transfers a limited band of frequencies and rejects low and high frequencies. An example of pole-zero placement and amplitude function for a BP filter is given in figure 12.4

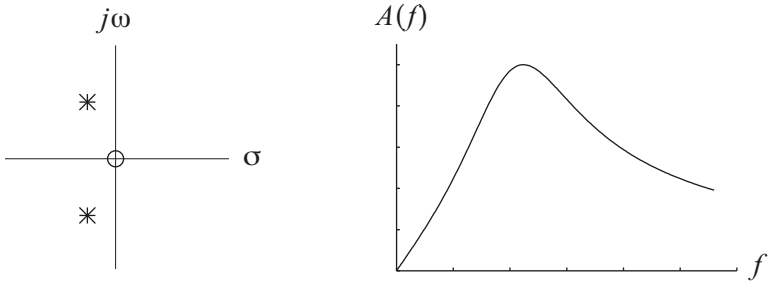


Figure 12.4 Pole placement and amplitude function for bandpass filter.

The transfer function for the bandpass filter in figure 12.4 is given by

$$H(s) = \frac{s}{(s + 0.5 - i) \cdot (s + 0.5 + i)}$$

The number of poles exceeds the number of zeros in a BP filter and at least one zero is located in the origin.

12.1.4 Bandstop Filter

A bandstop filter (BS) rejects a limited band of frequencies and passes low and high frequencies. An example of pole-zero placement and amplitude function for a BS filter is given in figure 12.5

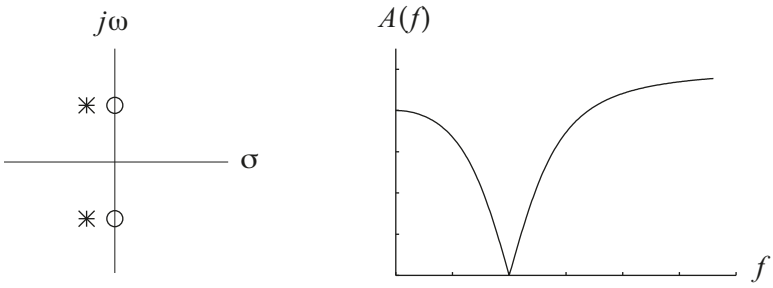


Figure 12.5 Pole placement and amplitude function for bandstop filter.

The transfer function for the bandstop filter in figure 12.5 is given by

$$H(s) = \frac{(s - i) \cdot (s + i)}{(s + 0.5 + i) \cdot (s + 0.5 - i)}$$

In a BS filter the number of poles equals the number of zeros. The "sharpness" of the filter is increased by reducing the distance between the zeros and the poles.

12.1.5 Allpass Filter

An allpass filter (AP) have a constant amplitude function. An example of pole-zero placement and amplitude function for a BS filter is given in figure 12.6

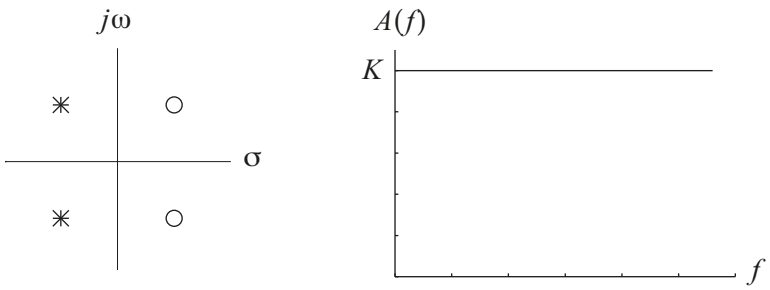


Figure 12.6 Pole placement and amplitude function for allpass filter.

The transfer function for the allpass filter in figure 12.6 is given by

$$H(s) = \frac{(s-1+i)\cdot(s-1-i)}{(s+1+i)\cdot(s+1-i)}$$

Note that the poles and the zeros are mirrored in the imaginary axis. The purpose of AP filters is to alter the phase function only. The phase function of the AP filter, $\phi(f)$, is illustrated in figure 12.7

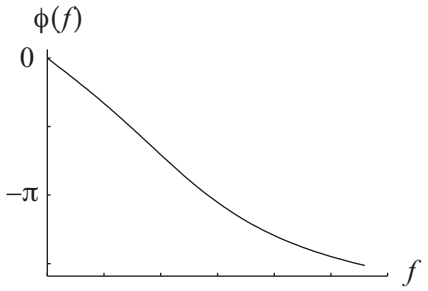


Figure 12.7 Phase function for allpass filter.

The allpass filter is a rather special filter that is mentioned here for the sake of completeness but will not be addressed any further in this chapter.

12.2 Filter Specification and Approximations

When a signal is to be removed by using a filter we can specify this filter with one or several cut-off frequencies depending on where the unwanted signal is located relative to the wanted signal. For example if signals are to be removed above a certain cut-off frequency f_c we could specify a LP filter as in figure 12.8. Here we have a well-defined passband below f_c with no attenuation and a stopband above f_c with infinite attenuation.

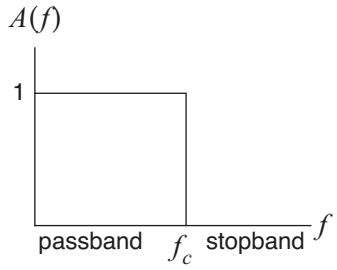


Figure 12.8 Ideal LP filter specification.

It is readily understood that we cannot implement such a filter, not even close. Thus, a more relaxed specification, an approximation of the desired behaviour, must be allowed given as boundaries in the amplitude-versus-frequency plot, see figure 12.9. Here the passband may be defined by the range from 0 to f_p . Similarly, the stopband is specified as ranging from f_r and upward. In addition to these we also have a transition band between the two latter bands, i.e., between f_p and f_r . In most applications a specification in this form is available or can be calculated. Sometimes, there is even a specification in the phase-versus-frequency plot which of course complicates the design procedure.

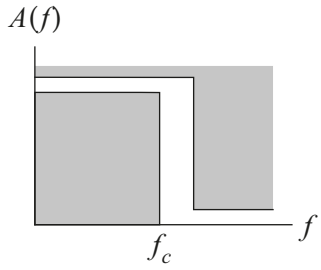


Figure 12.9 Example of filter specification for a LP filter.

To comply with the specification, a filter has to be found in a systematic manner. Of course, we could start by iterating various combinations of pole-zero patterns but this method is time-consuming because it involves to many variables to play with. Furthermore, we might end up with pole-zero pattern that is impractical to realise.

Instead, the most common procedure is to adopt a predefined filter approximation (described below) with well-defined characteristics and a limited number of variables than can be altered.

12.2.1 Filter Approximations

There are many predefined filter approximations. These are defined as LP filters. This is not a restriction since the design of other filters like HP, BP etc. are based on LP equivalent filters. More on this later. In this section three different filter approximations will be described; Butterworth, Chebyshev and

Bessel. Many more can be found in books dedicated to filter theory and design. All filters presented are of all-pole types which means that they have no zeros in the LP configuration. Furthermore, they can be designed with an arbitrary order n (number of poles). A higher order means steeper transition region between the pass band and the rejection band.

12.2.1.1 Butterworth

The most common filter is the Butterworth filter. Its amplitude function fulfils

$$A(f) = \frac{K}{\sqrt{1 + \left(\frac{\omega}{\omega_0}\right)^{2n}}} = \frac{K}{\sqrt{1 + \left(\frac{f}{f_0}\right)^{2n}}} \tag{12.2}$$

where n is the order of the filter and f_0 is a normalisation frequency. The amplitude function is shown in figure 12.10 for various orders n and unity normalisation frequency, $f_0 = 1$.

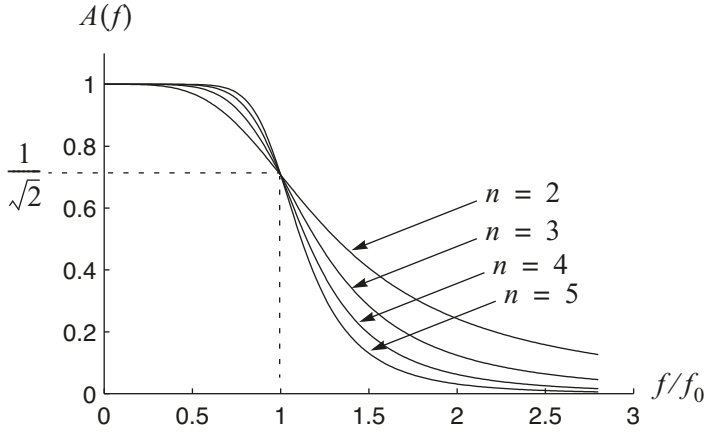


Figure 12.10 Butterworth filter amplitude function using (12.2) with $K = 1$.

The Butterworth filter has the following properties:

- *Maximally flat* around $f = 0$, i.e., the first $2n - 1$ derivatives of $A(f)$ equal to zero at $f = 0$
- $A(0) = K$, independent of n
- $A(f)$ is monotonic
- $A(f_0) = 1/\sqrt{2}$ independent of n
- $A(f)$ decreases with $6 \cdot n$ dB per octave for large f

The complete transfer function can be developed from the amplitude function and it is straightforward to calculate the poles but usually tables are used instead either given as expanded denominator polynomials or as poles for various orders, assuming a unity normalisation frequency, f_0 .

12.2.1.2 Chebyshev

Another LP filter approximation is the Chebyshev filter. Its amplitude function fulfils

$$A(f) = \frac{K}{\sqrt{1 + \varepsilon^2 C_n^2\left(\frac{\omega}{\omega_0}\right)}} = \frac{K}{\sqrt{1 + \varepsilon^2 C_n^2\left(\frac{f}{f_0}\right)}} \tag{12.3}$$

or with normalised frequency, $\omega_0 = 1$, and $K = 1$:

$$A(f) = \frac{1}{\sqrt{1 + \varepsilon^2 C_n^2(\omega)}} \tag{12.4}$$

where ε is a ripple factor (see below) that defines the amount of ripple in the passband. $C_n(\omega)$ is the Chebyshev polynomial for order n and is given by

$$C_n(\omega) = \begin{cases} \cos(n \cdot \arccos(\omega)) & |\omega| < 1 \\ \cosh(n \cdot \operatorname{arccosh}(\omega)) & |\omega| > 1 \end{cases} \tag{12.5}$$

The amplitude function has been plotted in figure 12.10 using (12.4) for various orders n, ripple factor $\varepsilon = 1$ and unity normalisation frequency, $f_0 = 1$.

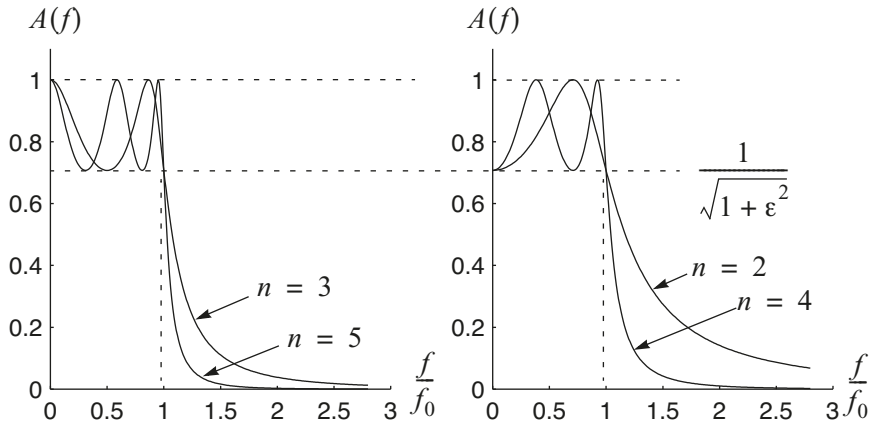


Figure 12.11 Chebyshev filter amplitude function using (12.4).

The Chebyshev filter has the following properties (assuming $f_0 = 1$):

- Ripple in the pass band ($f < 1$) between K and $K/\sqrt{1+\varepsilon^2}$
- $A(1) = K/\sqrt{1+\varepsilon^2}$, independent of n
- $A(0) = K/\sqrt{1+\varepsilon^2}$, even n
- $A(0) = K$, odd n
- For $f > 1$ $A(f)$ is monotonic
- $A(f)$ decreases with $6 \cdot n$ dB per octave for large f

It is possible to calculate the poles but it is more common and convenient to consult tables either given as expanded denominator polynomials or as poles for various orders and ripple levels, with a unity normalisation frequency, f_0 .

Compared with the Butterworth filter we have gained one degree of freedom with the ripple factor ε . Ripple in the pass band is not something that we want but the fact that we can trade ripple with steepness of the transition region makes this filter approximation very useful in some applications. That is, for a given filter order the steepness of the transition region will increase as we allow more ripple in the pass band. On the other hand, the phase function is not as good as for the Butterworth filter.

12.2.1.3 Bessel

The characteristics of the Butterworth and the Chebyshev filters originate from a desired amplitude function and the phase function is of little concern. In some applications, however, it is the phase function that is the most important property or the group delay to be precise. The group delay is related to the phase function as

$$\tau_g = -\frac{d\phi(\omega)}{d\omega} \tag{12.6}$$

A linear phase function - that is a constant group delay - will not change the shape of the signal in the time domain if the amplitude function is constant in the frequency range occupied by the signal. A group delay that varies with the frequency typically results in a step-response with overshoot and damped oscillation. These effects increase with higher order and higher ripple in the pass band.

The Bessel filter has been developed with constant group delay in mind. It has a maximally flat group delay in the pass band. Note that this does not mean a perfectly constant group delay. Since the filter has a limited order it is only an approximation of the desired property, a true delay e^{-sT_d} . In addition, the amplitude function will not be as steep in the transition region as for Butterworth or Chebyshev.

The amplitude function has been plotted in figure 12.12 for various orders n unity normalisation frequency, $f_0 = 1$. Note the gentle slope after the normalisation frequency and the fact that it does not change much with increasing order.

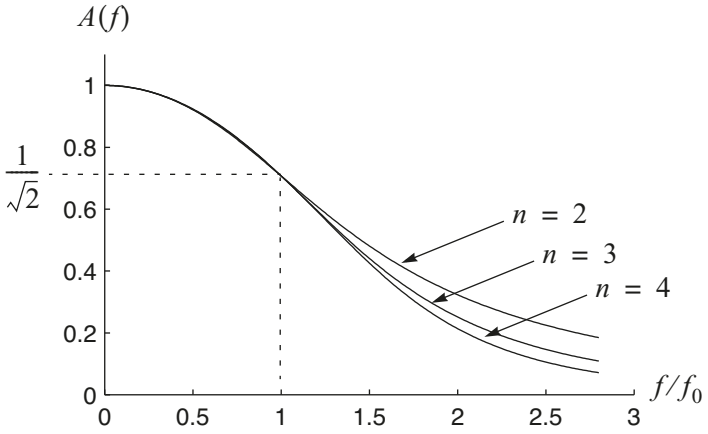


Figure 12.12 Bessel filter amplitude function.

There is no general function for the Bessel amplitude function for different orders as was the case for Butterworth and Chebyshev filters. Tables are commonly used to find frequency-normalised poles or polynomial coefficients from which the amplitude function can be plotted as in figure 12.12. In the design procedure for the Bessel filters both the amplitude characteristics (frequency ranges for pass band and rejection band) and the deviation from a constant group delay are considered. For example, a maximum relative delay error at a given frequency alongside with a maximum amplitude error at another frequency can be specified. From this information it is possible to use filter tables to find the minimum filter order. The final filter order is chosen based on the rejection band specification. Since there is no simple amplitude function that includes the order of the filter, one filter order at the time must be investigated to see whether it fulfils the rejection band specification or not.

For comparison, the amplitude, phase and group delay for all three filter approximations are plotted in figures 12.13 to 12.15 for 4th order configurations with unity 3dB cut-off frequency and 3dB Chebyshev ripple.

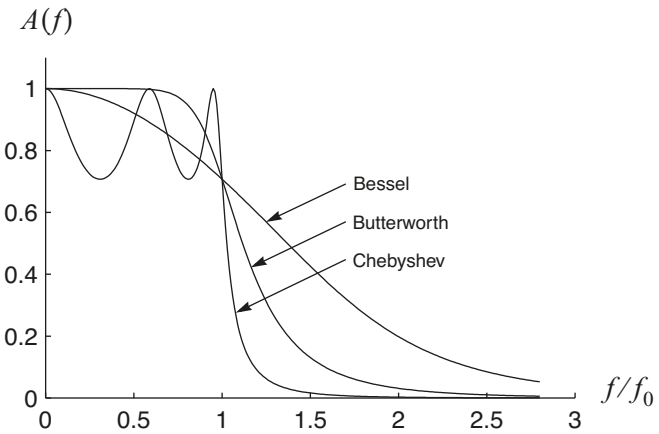


Figure 12.13 Amplitude function for 4th order Butterworth, Chebyshev (3dB ripple) and Bessel filter.

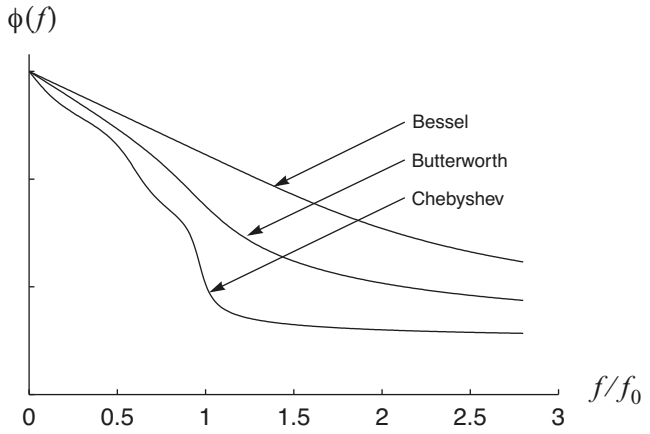


Figure 12.14 Phase function for 4th order Butterworth, Chebyshev (3dB ripple) and Bessel filter.

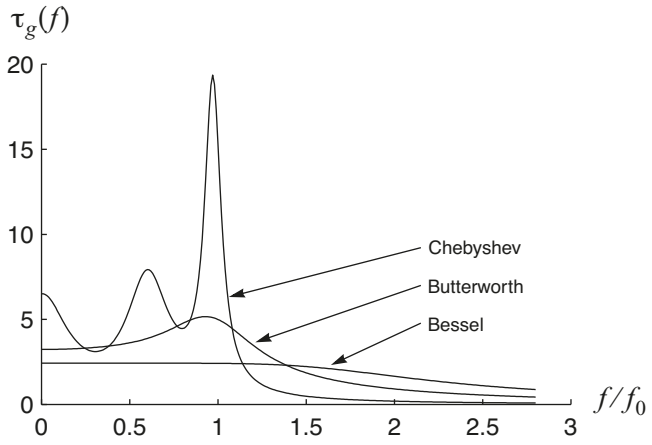


Figure 12.15 Group delay for 4th order Butterworth, Chebyshev (3dB ripple) and Bessel filter.

12.3 Designing from Specification

In section 12.2 a typical filter specification was shown for a lowpass filter consisting of three regions, the passband, the transition region and finally the rejection region. Once the filter approximation (Butterworth, Chebyshev, Bessel etc.) has been chosen it is possible to decide the order, cut-off (normalisation) frequency and ripple that results in a filter that complies with the specification. This will be demonstrated below. Also, the procedure for transforming other filter types (HP, BP, BR) to a lowpass equivalent and design any of the filter types in the lowpass domain is also demonstrated.

Note that finding the order, cut-off frequency and ripple for a given filter approximation is only the first step in the design procedure. The next step involves the actual implementation and this topic is treated in section 12.4 and onward.

12.3.1 Lowpass Filter Design

If the amplitude function, $A(f)$, is given as was the case for Butterworth and Chebyshev it is straightforward to find a solution that complies with a filter specification as the one shown in figure 12.16.

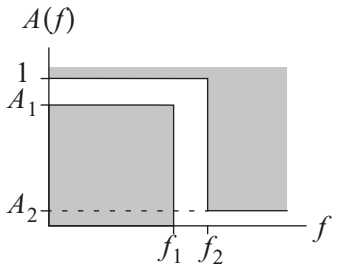


Figure 12.16 Simple lowpass filter specification.

A system of equations can be developed based on the amplitude function and the specification. For example we know that the amplitude at $f = f_1$ must equal or higher than A_1 and at $f = f_2$ the amplitude must be equal or lower than A_2. Thus we get two equations where the order n and the normalisation frequency are the unknowns to be found. For the Chebyshev filter we can choose a ripple level that is equal to or less than the allowed gap in the pass band $(1 - A_1)$. If the ripple is chosen to be equal to the gap the normalisation frequency becomes equal to f_1.

Example 12.1 Lowpass Filter Design

Determine the order and the normalisation frequency for a Butterworth filter and a Chebyshev filter that fulfil the specification in figure 12.17 below. Chose Chebyshev ripple with the same magnitude as the gap in the passband. Also, plot the amplitude function for the filters chosen and compare with the specification.

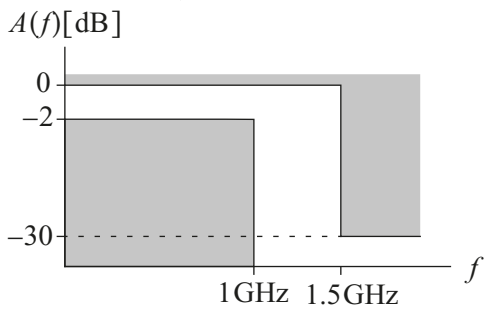


Figure 12.17 Lowpass filter specification for example example 12.1. (not to scale)

Butterworth:

The amplitude function is given by

$$A(f) = \frac{K}{\sqrt{1 + \left(\frac{f}{f_0}\right)^{2n}}}$$

Since the gain at $f = 0$ is unity (0dB) K must be equal to 1. In the pass band the amplitude should be equal to or higher than -2dB up to 1GHz, i.e.,

$$A(1\times10^9) = -2\text{dB} = 0.7943 = \frac{1}{\sqrt{1 + \left(\frac{1\times10^9}{f_0}\right)^{2n}}} \Rightarrow$$

$$\left(\frac{1\times10^9}{f_0}\right)^{2n} = 0.5849 \Rightarrow f_0 = \frac{1\times10^9}{0.5849^{1/2n}}$$

The rejection band is defined by the -30dB corner at 1.5GHz, i.e.,

$$A(1.5\times10^9) = -30\text{dB} = 0.0316 = \frac{1}{\sqrt{1 + \left(\frac{1.5\times10^9}{f_0}\right)^{2n}}} \Rightarrow$$

$$\left(\frac{1.5\times10^9}{f_0}\right)^{2n} = 999 \Rightarrow f_0 = \frac{1.5\times10^9}{999^{1/2n}}$$

Combining the results from the two corners we get $n = 9.2$ and since we cannot allow a fractional order it must be rounded to the nearest higher integer number, i.e., $n = 10$. Now when we know the order of the filter we should go back to the first corner defining the pass band (-2dB at 1GHz) and calculate the normalisation frequency using $n = 10$:

$$A(1\times10^9) = 0.7943 = \frac{1}{\sqrt{1 + \left(\frac{1\times10^9}{f_0}\right)^{2\cdot10}}} \Rightarrow f_0 = 1.03\times10^9$$

From this result it is now possible to write the transfer function or the individual poles for the transfer function using tabulated data.

Chebyshev:

The ripple should have the same magnitude as the gap in the pass band, i.e., 2dB. Referring to figure 12.11 we have

$$-2\text{dB} = 0.7943 = \frac{1}{\sqrt{1 + \varepsilon^2}} \Rightarrow \varepsilon^2 = 0.585$$

In this special case when the ripple is equal to the gap in the pass band we know that the normalisation frequency is equal to the first corner, i.e., $f_0 = 1\text{GHz}$. Thus, the second corner defining the rejection region remains to be investigated:

$$A(1.5\times10^9) = -30\text{dB} = 0.0316 = \frac{1}{\sqrt{1+\varepsilon^2 C_n^2\left(\frac{1.5\times10^9}{f_0}\right)}} \Rightarrow$$

$$C_n^2\left(\frac{1.5\times10^9}{1\times10^9}\right) = 2919 \Rightarrow \cosh(n\cdot\operatorname{arccosh}(1.5)) = 54.0 \Rightarrow$$

$$n = 4.86$$

Thus we select the nearest higher order, $n = 5$. Note that this is half of the order of the Butterworth filter.

In this case we do not need to go back to the first corner and calculate a new normalisation frequency as was the case for the Butterworth filter. In this case the ripple has the same magnitude as the gap and thus we have a fixed normalisation frequency given by the first corner, 1GHz. Since the order of the filter chosen is higher than the calculated filter order we automatically know that this filter will comply with the specification in thew second corner at 1.5GHz. This holds because a higher the transition region becomes steeper for increasing order. The amplitude functions for the two filters chosen are shown in figure 12.18 below.

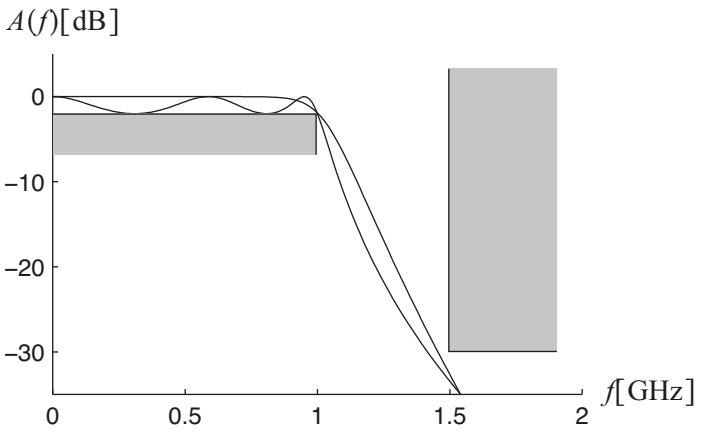


Figure 12.18 Final result of filter design, example 12.1.

For the Bessel filter and other filters whose amplitude function cannot be written as a function of the order some iterative work will obviously be necessary to find a solution that complies with a specification. Also, if the filter specification is more complicated, i.e., more corners in the amplitude function, group delay specification etc. an iterative approach might also prove to be necessary whereas in other cases it is possible to utilise tabulated filter data. However, it is out of the scope of this chapter to deal with all these cases. Instead we refer to literature dedicated to filter design and theory.

12.3.2 Highpass Filter Design

The design procedure for a highpass, bandpass or band-rejection filter is performed as a lowpass filter design. The procedure begins by transforming the original filter specification to a lowpass equivalent filter specification. Then we design a lowpass filter that complies with our lowpass specification and finally the result is transformed back to the desired filter type.

Thus, the basic idea is to find transforms for the frequency axis that makes a lowpass filter transform into the desired filter type. Moreover, these transforms must be as simple as possible and are restricted to rational polynomials so that the new transfer function will become a rational polynomial as well. Finally, the transforms should be simple with respect to realisation so that a transformed circuit becomes realisable with minimum effort and complexity. This will be discussed in more detail in section 12.4.

In the case of highpass filters the transform is quite simple, see figure 12.19. We note that by simply inverting the frequency axis we get the desired properties. ω_I is an optional frequency scaling that is usually set to unity. Since the design procedure includes a transformation from HP to LP and back to HP again ω_I will not affect the design at all. However, it is used here for consistency with other transformation formulas. Also, note the corresponding transformation in the s-domain ($s_{LP} \rightarrow \omega_I^2 / s_{HP}$) that is used to transform a transfer function.

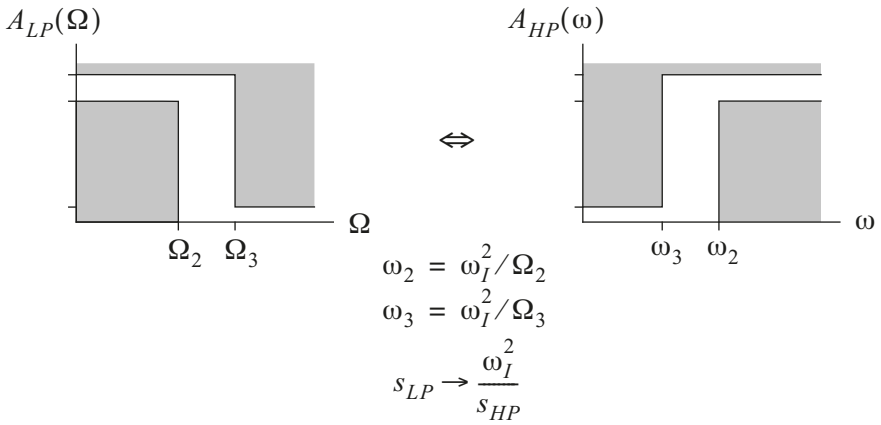


Figure 12.19 Transformation of a lowpass to a highpass filter specification.

12.3.3 Bandpass Filter Design

In the case of bandpass filters the transform involves more parameters due to the two-sided passband and rejection-band regions, see figure 12.20. Also, the transformation becomes somewhat more complicated than for the highpass filter. Note the equality between the frequency products, $\omega_1\omega_2 = \omega_3\omega_4$. A filter cannot be specified arbitrarily due to this restriction, which is induced by the specific transformation chosen.

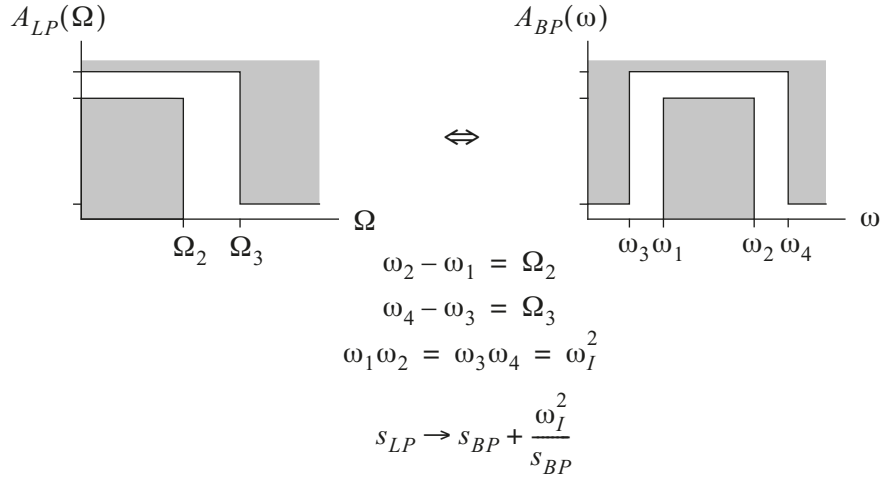


Figure 12.20 Transformation of a lowpass to a bandpass filter specification.

12.3.4 Bandstop Filter Design

The bandstop filter is similar to the bandpass filter in that it needs four frequency variables that will be transformed to two lowpass filter frequency variables, only the s-domain transformation is different, see figure 12.21. As was the case for bandpass filter the bandstop to lowpass transform also restricts the possible combinations of frequencies in the band-rejection domain.

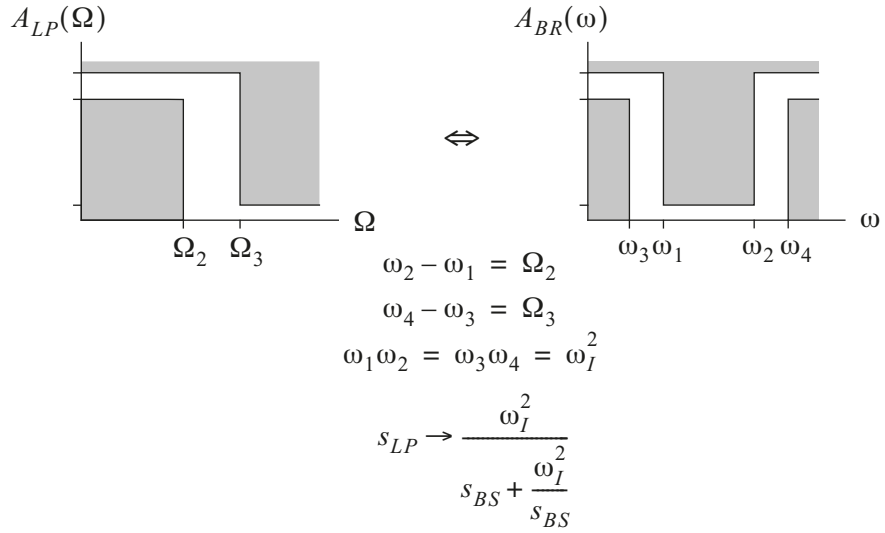


Figure 12.21 Transformation of a lowpass to a band-rejection filter specification.

12.4 Realisation

So far, we have defined different types of filters, i.e., lowpass, highpass etc. and their typical placement of poles and zeros. We have also described various filter approximations and their properties and we know that we can calculate the poles for these filters or simply look them up in filter tables. The next and the last step involves realisation. Depending on the frequency band different techniques can be used. In low frequency applications, typically below 1MHz, we can choose from a number of different techniques ranging from passive filters through active filters using operational amplifiers, switched-capacitor filters to digital filters. As the frequency increases techniques based on active circuit elements becomes gradually less attractive for several reasons, mainly due to increased parasitic effects, less linearity, more noise and more power. However, today there exists techniques for implementing active filters with fairly high cut-off frequencies (100MHz or so) in silicon without external components and the frequency limit increases with never-ending advances in process technology. In radio-frequency applications passive filter structures are widely used, typically realised with lumped and ideally lossless circuit elements, i.e., inductors and capacitors. As the frequencies of interest passes, lets say, 1GHz the parasitic effects in the lumped components, especially the inductors, becomes more dominant and techniques based on distributed circuit elements (transmission lines) are ultimately the only choice for the designer. Below, the realisation of passive filters using lumped and distributed circuit elements will be discussed in more detail.

12.4.1 Realisation Using Lumped and Passive Circuit Elements

This section deals with the implementation of passive filters using reactive circuit elements (L and C). The ladder topology illustrated in figure 12.22 is the most common passive filter structure. Its regular structure allows generalisation of methods for calculating circuit element values or specifying these values in tabular form for different filter approximations. Tabulated data will be used here. Each series and shunt element may consist of one single element or several elements depending on the filter that will be realised. Also, depending on the filter order, type etc. the elements at the ends can be either shunt or series elements.

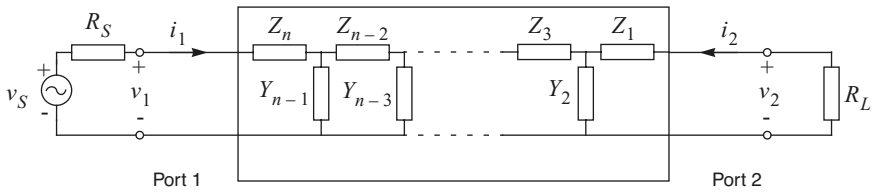


Figure 12.22 Ladder topology.

12.4.1.1 Lowpass Ladder Filter

In a lowpass configuration all series elements are inductors and all parallel elements are capacitors as illustrated in figure 12.23. Moreover, depending on whether the order of the filter is odd or even the last element is either an inductor or a capacitor.

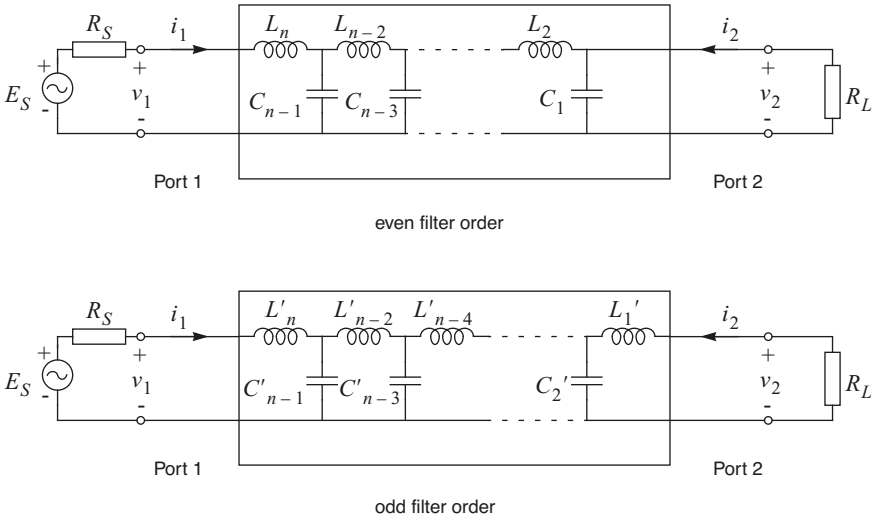


Figure 12.23 Lowpass filter topologies for odd and even filters orders.

Once the filter order, the normalisation frequency and the ripple has been determined according to the methods in section 12.3, circuit element values are looked up in tables. Different tables exist for a few different R_L/R_S ratios, typically 0.1, 0.25, 0.5 and 1. If the ratio is larger than unity the tables can still be used because of the reciprocity of the filter, i.e., the filter topology can be flipped between the source and the load. The component values obtained from the tables are normalised values that should be denormalised as follows:

- Butterworth and Chebyshev filters
  All inductor values should be multiplied by R_S/ω_0 and all capacitor values should be multiplied by $1/(\omega_0 R_S)$ where ω_0 is the angular normalisation frequency.
- Bessel filters
  All inductor values should be multiplied by $R_S\tau_0$ and all capacitor values should be multiplied by τ_0/R_S where τ_0 is the delay of the filter and is equal to $1/\omega_0$, the inverse of the normalisation frequency. Thus a normalised Bessel filter always has a group delay equal to unity.

The gain of the filter at $f = 0$ is simply determined by the voltage division between R_S and R_L. From this we conclude that power matching is only obtained for $R_L/R_S = 1$.

Given a lowpass filter specification, the design procedure is:

1. Find suitable filter approximation and order.
2. Calculate or look up component values for LP ladder filter.
3. Denormalise component values.

Example 12.2 Design of 5th order Butterworth Filter

Design a fifth order Butterworth ladder filter with a cut-off frequency of 1GHz with $R_S = R_L = 50\Omega$.

We have an R_L/R_S ratio equal to one. Tabulated data can be found in appendix B. Since it is an odd order filter the values in the table applies to the odd order structures in figure 12.23. Thus we get

$L'_1 = 0.6180$

$C'_2 = 1.6180$

$L'_3 = 2.0000$

$C'_4 = 1.6180$

$L'_5 = 0.6180$

and denormalisation gives

$(R_S/\omega_0) \cdot L'_1 = (50/2\pi 10^9) \cdot 0.6180 = 4.91\text{nH}$

$(1/\omega_0 R_S) \cdot C'_2 = (1/100\pi 10^9) \cdot 1.6180 = 5.15\text{pF}$

$(R_S/\omega_0) \cdot L'_3 = (50/2\pi 10^9) \cdot 2.0000 = 15.9\text{nH}$

$(1/\omega_0 R_S) \cdot C'_4 = (1/100\pi 10^9) \cdot 1.6180 = 5.15\text{pF}$

$(R_S/\omega_0) \cdot L'_5 = (50/2\pi 10^9) \cdot 0.6180 = 4.91\text{nH}$

The filter is illustrated in figure 12.24 below together with denormalised component values.

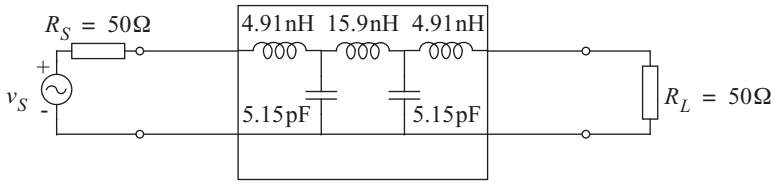


Figure 12.24 Realisation of 1GHz 5th order Butterworth filter.

The transfer function is shown below. It is exactly the same as the ideal Butterworth transfer function given as a rational polynomial, see equation (12.1), i.e., no approximations or deviations are introduced when realising the filter with lumped components.

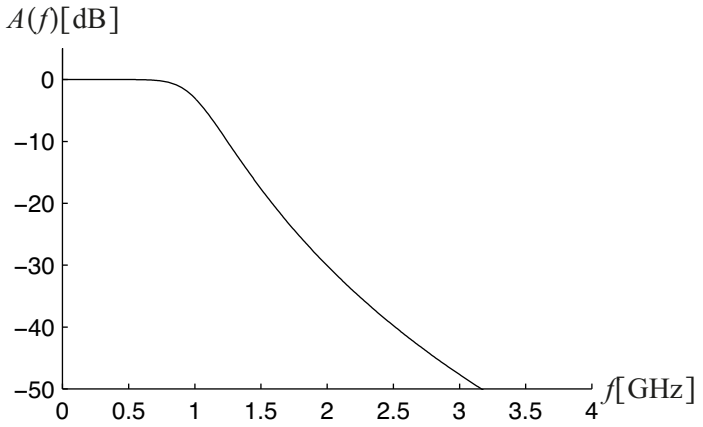


Figure 12.25 Transfer function for fifth order Butterworth filter.

12.4.1.2 HP, BP and BR Ladder Filters

Earlier in this chapter we presented frequency transformations to transform a filter specification for any filter type to a lowpass equivalent. In this lowpass domain an appropriate filter approximation could be found. The same transformation could be used to transform that filter approximation back to the original filter type. This transformation technique can be applied directly to the ladder filter where each component can be transformed to another component or several components depending on whether it is a HP, BR or BR filter that is to be designed. A summary of all frequency transformations are given in table 12.1.

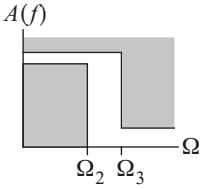
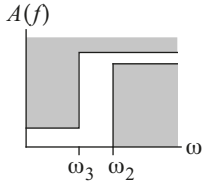
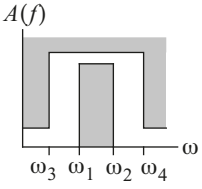
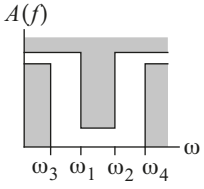
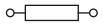
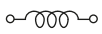
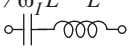
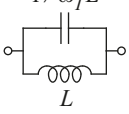
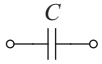
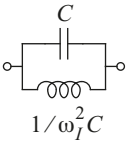
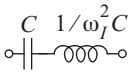
In summary, the design procedure for a non-LP ladder filter becomes:

1. Transform filter specification to LP filter specification.
2. Find suitable filter approximation and order.
3. Calculate or look up component values for LP ladder filter.
4. Denormalise component values.
5. Transform component values to the desired filter type.

12.4.2 Realisation Using Distributed Circuit Elements

The ladder topology with lumped components that were described above realises a given filter transfer function exactly, assuming ideal components. This cannot be done with distributed circuit elements, i.e., transmission lines. Therefore, the number of possible techniques in the latter case outnumbers those using lumped components since different techniques will give different compromises. It is far beyond the scope of this textbook to deal with more than just a few of these techniques.

Table 12.1 Frequency transformations

LP	HP	BP	BR
$A(f)$ 	$A(f)$  $\omega_2 = \omega_I^2/\Omega_2$ $\omega_3 = \omega_I^2/\Omega_3$	$A(f)$  $\omega_2 - \omega_1 = \Omega_2$ $\omega_4 - \omega_3 = \Omega_3$ $\omega_1\omega_2 = \omega_3\omega_4 = \omega_I^2$	$A(f)$  $\omega_2 - \omega_1 = \omega_I^2/\Omega_2$ $\omega_4 - \omega_3 = \omega_I^2/\Omega_3$ $\omega_1\omega_2 = \omega_3\omega_4 = \omega_I^2$
s	$\frac{\omega_I^2}{s}$	$s + \frac{\omega_I^2}{s}$	$\frac{\omega_I^2}{s + \omega_I^2/s}$
R 	R 	R 	R 
L 	$1/\omega_I^2 L$ 	$1/\omega_I^2 L \quad L$ 	$1/\omega_I^2 L$ 
C 	$1/\omega_I^2 C$ 	C 	$C \quad 1/\omega_I^2 C$ 

12.4.2.1 LP Filters

The most intuitive way of realising a lowpass filter with transmission lines is to start from the ladder topology, see figure 12.22. With some limitations it is possible to implement series inductors and shunt capacitors using transmission lines. Consider the Smith chart in figure 12.26 below. Here it is shown how the impedance and the admittance are altered by a series inductor and a shunt capacitor, respectively. It is also seen that a transmission line in series with the load will approximately emulate a series inductor if the characteristic impedance of the transmission line is high compared with the load resistance and a capacitor if the characteristic impedance is low. It is readily seen that the transmission line must not be to long (less than $\lambda/8$) or the deviation between the curves becomes too large. Thus, a sequence of series-connected alternating high and low Z_0 will resemble a filter that will be a approximation of a LP ladder filter. We will now present how the length of a transmission line together with the characteristic impedance, Z_0 , and the propagation coefficient, γ , relate to L and C .

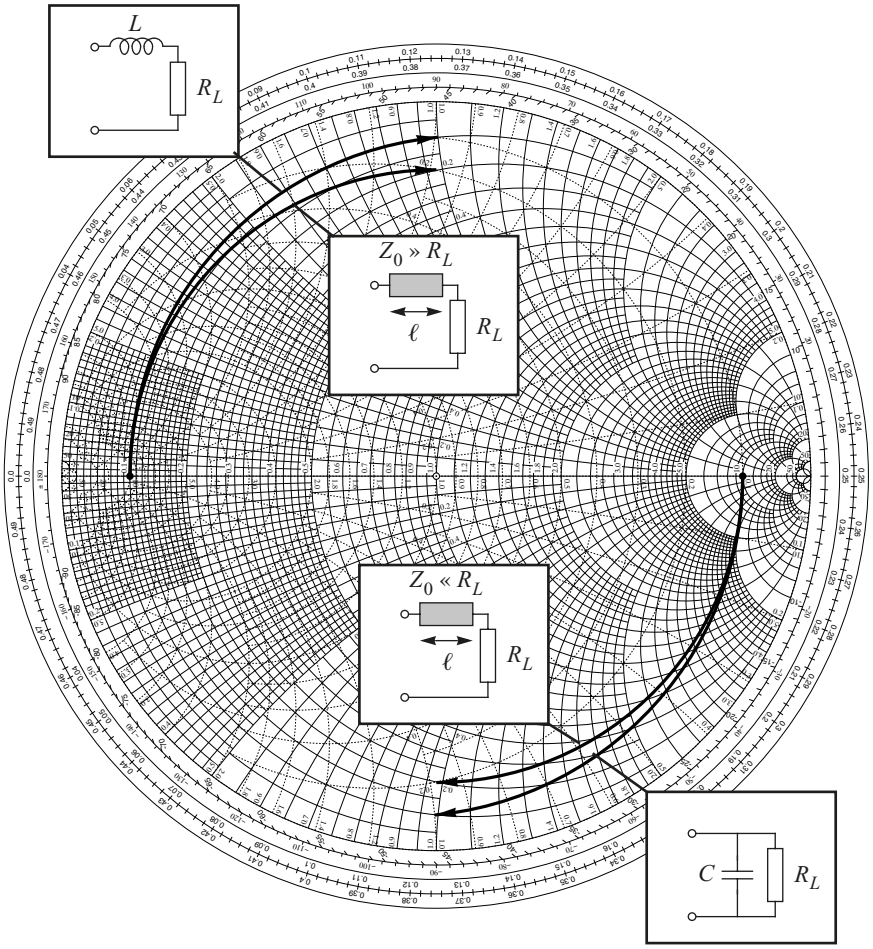


Figure 12.26 Transmission line acting as a series inductor and series capacitor.

It is possible to model a transmission line with a given length in a manner similar to what was done for an infinitesimal segment of a transmission line in an earlier chapter. That is, we can relate the properties of the transmission line to a TEE or PI equivalent circuit as illustrated in figure 12.27. Derivation of the expressions are given in [1].

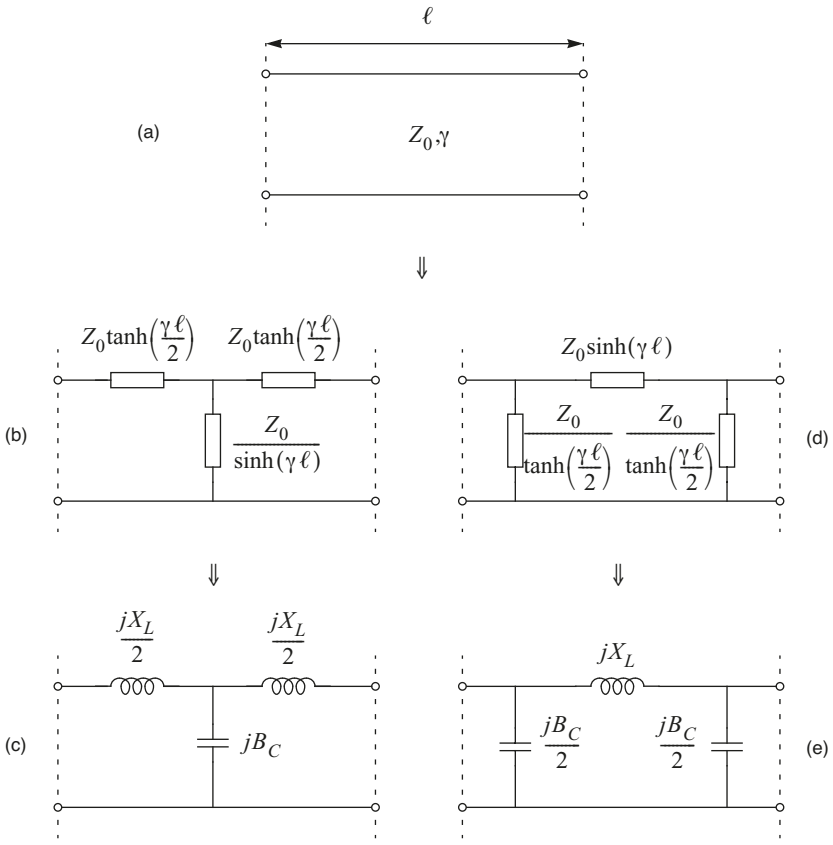


Figure 12.27 Lumped line representation of a short line section (a) transmission line (b) TEE representation (c) lumped TEE representation for lossless transmission line (d) PI representation (e) lumped PI representation for lossless transmission line.

From now on we will assume lossless lines (line attenuation $\alpha = 0$) which means that the model in figure 12.27c applies. From the expressions given in figure 12.27 we get

$$X_L = 2Z_0 \tan\left(\frac{\beta\ell}{2}\right) \tag{12.7}$$

and

$$X_C = \frac{1}{B} = \frac{Z_0}{\sin(\beta\ell)} \tag{12.8}$$

where β is the line phase coefficient. Similarly, If a PI equivalent is used we get

$$X_L = Z_0 \sin(\beta\ell) \tag{12.9}$$

and

$$X_C = \frac{1}{B} = \frac{Z_0}{2\tan(\beta\ell/2)} \tag{12.10}$$

We can now substitute $\beta = \omega/v_p$, where v_p is the phase velocity, into these equations and use the fact that $\tan\phi \approx \sin\phi \approx \phi$. Then for the TEE equivalent we get

$$X_L = \omega L = 2Z_0\tan\left(\frac{\omega\ell}{2v_p}\right) \approx \left.\frac{Z_0\omega\ell}{v_p}\right|_{\ell<\lambda/8} \tag{12.11}$$

$$B_C = \omega C = Y_0\sin\left(\frac{\omega\ell}{v_p}\right) \approx \left.\frac{Y_0\omega\ell}{v_p}\right|_{\ell<\lambda/8} \tag{12.12}$$

and for the PI equivalent

$$X_L = \omega L = Z_0\sin\left(\frac{\omega\ell}{v_p}\right) \approx \left.\frac{Z_0\omega\ell}{v_p}\right|_{\ell<\lambda/8} \tag{12.13}$$

$$B_C = \omega C = 2Y_0\tan\left(\frac{\omega\ell}{2v_p}\right) \approx \left.\frac{Y_0\omega\ell}{v_p}\right|_{\ell<\lambda/8} \tag{12.14}$$

Note that the equations for the TEE and the PI circuits are approximately the same. To summarise equations (12.11) to (12.14) we have

$$L = \frac{Z_0\ell}{v_p} = \frac{Z_0\ell}{\lambda_\ell f} \tag{12.15}$$

$$C = \frac{\ell}{Z_0 v_p} = \frac{\ell}{Z_0\lambda_\ell f} \tag{12.16}$$

What these two last equations say are that a short transmission line will have a large series inductor and a small shunt capacitor if the characteristic impedance is high and vice versa if the characteristic impedance is low. This is in accordance with the Smith-chart in figure 12.26. Thus, to realise a series inductor Z_0 must be chosen high enough so that the parasitic shunt capacitance given by (12.16) will be negligible. Accordingly, a shunt capacitor must be realised using a sufficiently low Z_0 such that the parasitic inductance given by (12.15) can be considered negligible.

At this stage two things remain unanswered. How short is a short transmission line and what is a low and a high characteristic impedance? As for the length of the transmission line it should not be any longer than $\lambda/8$ at the highest frequency of interest, that is the highest frequency where the filter should operate properly. $\lambda/8$ is not a strict limit merely a rule of thumb. As

the length approaches $\lambda/8$ the filter will deviate more and more from the ideal behaviour. Figure 12.26 gives the answer to what is low and high characteristic impedance. The impedances loading a transmission line (both ends) should have a much lower or higher impedance with respect to the characteristic impedance of the transmission line. Thus, a LP filter realised with this technique should have $Z_{0low} \ll R_S,R_L$ and $Z_{0high} \gg R_S,R_L$.

Example 12.3 Design of LP transmission line filter

Implement the filter in example 12.2, a 5th order Butterworth filter with 1GHz cut-off frequency, using microstrip technology. The filter should operate accurately up to 2GHz. The microstrip carrier is an epoxy-laminate with $\varepsilon_r = 4.5$ and h=1.56mm.

There are two limiting factors that we must deal with. First, the length of each line section should not exceed $\lambda/8$ for the highest frequency of interest, in this case 2GHz. However, from (12.15) and (12.16) we know that for a given inductance and capacitance a decrease in line length must be compensated by and increase of Z_{0high} and a decrease of Z_{0low}. Furthermore, practical values for Z_{0high} and Z_{0low} is fairly limited and ranges from say a few Ω to a few hundred Ω depending on what kind of geometry (i.e. microstrip, stripline, coaxial etc.) that is used to realise a transmission line.

The largest values of the inductors and capacitors results in the longest line sections. We can use (12.15) and (12.16) to determine the required Z_{0low} and Z_{0high} if the length should not exceed $\lambda/8$. We have

$$C = \frac{\ell}{Z_{0low}\lambda f} \rightarrow Z_{0low} = \frac{\ell}{C\lambda f}$$

$$L = \frac{Z_{0high}\ell}{\lambda f} \rightarrow Z_{0high} = \frac{L\lambda f}{\ell}$$

With $\ell/\lambda = 1/8$ at 2GHz, $C_{max} = 5.15\text{pF}$ and $L_{max} = 15.9\text{nH}$ we get $Z_{0high} = 256\Omega$. and $Z_{0low} = 12\Omega$. With the given laminate we can calculate the widths of the microstrip structures; $w_{low} = 19.4\text{mm}$ and $w_{high} = 0.0088\text{mm}$. Obviously, w_{high} is much to small. A practical limit for the given laminate is a few tenths of a millimetre depending on the manufacturing tolerances. As far as w_{low} is concerned it should not be any wider than $\lambda/4$ at the highest frequency of interest. In conclusion it appears difficult to comply with the length condition, $\ell < \lambda/8$, at 2GHz.

We will now take on a pragmatic approach and begin by choosing some practical values for the characteristic impedances;
$Z_{0high} = 125\Omega$ and $Z_{0low} = 20\Omega$

Now the widths become $w_{low} = 10.7\text{mm}$ and $w_{high} = 0.333\text{mm}$.

In example 12.2 the component values were found to be $L'_1 = 4.91\text{nH}$, $C'_2 = 5.15\text{pF}$. $L'_3 = 15.9\text{nH}$, $C'_4 = 5.15\text{pF}$ and $L'_5 = 4.91\text{nH}$. Using (12.15) and (12.16) we get $\ell_1 = 0.0786\lambda$, $\ell_2 = 0.206\lambda$, $\ell_3 = 0.254\lambda$, $\ell_4 = 0.206\lambda$ and ,$_5 = 0.0786\lambda$ at 2GHz. We note that the line section in the middle is the longest and is about two times longer than the limit.

The wavelength, λ, is a function of the characteristic impedance. We get $\lambda_{low} = 75.9\text{mm}$ and $\lambda_{high} = 86.1\text{mm}$. The actual line lengths become $\ell_1 = 6.77\text{mm}$, $\ell_2 = 15.6\text{mm}$, $\ell_3 = 21.9\text{mm}$, $\ell_4 = 15.6\text{mm}$ and $\ell_5 = 6.77\text{mm}$. The complete layout is shown in actual size in figure 12.28 below.

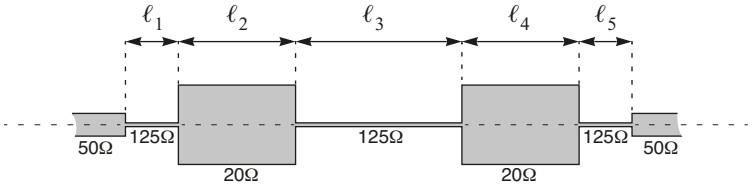


Figure 12.28 5th order Butterworth filter realised with microstrip lines.

In this design we have neglected all parasitic effects such as the shunt capacitance in the inductive line, the series inductance in the capacitive line and the fact that the laminate is far from lossless. Also, the discontinuity between thin and wide lines introduce parasitic effects. All these effects are seldom calculated by hand. Instead, the designer relies on software tools for design and simulation of distributed circuits that fairly accurately model these effects. In figure 12.29 the transfer functions are shown for the lumped ladder filter, distributed filter without discontinuity parasitics and losses and finally distributed filter with all parasitics.

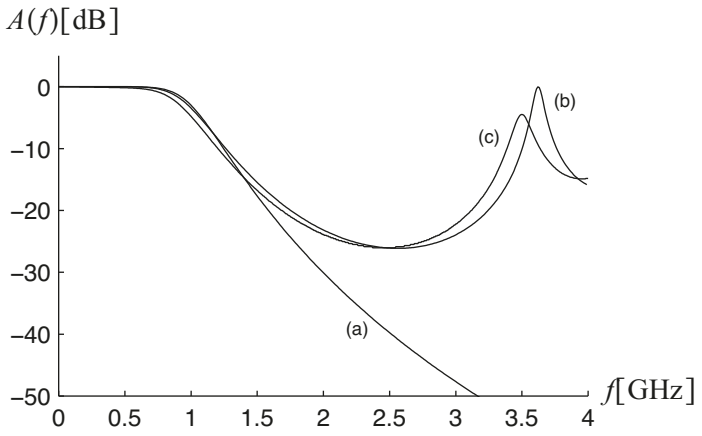


Figure 12.29 Comparison of transfer function for (a) lumped ladder filter (b) equivalent distributed filter without discontinuity parasitics and line losses (c) equivalent distributed filter with all parasitics and line losses.

12.4.2.2 Other Filter Types

Before continuing with the HP, BP and BS filters we will summarise some possible techniques to implement inductive and capacitive elements as well as combinations of these using transmission lines, see figure 12.30. Here, the geometry of the transmission line equivalents correspond to the layout of microstrip or stripline structures. In figure 12.30a and b we have the series inductance and shunt capacitance as described above. A shunt inductor can also be realised with a short-circuited stub or the same stub extended with an open-circuited $\lambda/4$ transmission line, see figure 12.30c. Combinations lead to shunt connected series and parallel resonance circuits, see figure 12.30d and e. Finally, a series $\lambda/2$ transmission line with low characteristic impedance will act as a parallel resonance circuit as well. In all cases but the last the lengths of the structures are limited such that $\ell < \lambda/8$ where λ denotes the wavelength for which the transmission line equivalent circuit should model the lumped circuit accurately.

These structures does not present a complete set of structures to realise other filter types besides the LP filter. In table 12.1 we see that a series capacitor is required to implement a HP filter, a series-connected series resonance circuit to implement a BP filter and a series-connected parallel resonance circuit to implement a BS filter. That is, all remaining circuits requires a series capacitor.

Series capacitors are far from trivial to implement using transmission line techniques. For microstrip and stripline structures one possibility is to have a physical series break of the structure, a gap. Typically, the gap is so small that it is difficult to obtain good accuracy. Another problem is the fact that the gap will not act as a pure capacitor. Therefore, if this technique is to be used accurate modelling will be required to be able to take into account the non-ideal behaviour of the gap at the design stage. Another option is to have a hybrid with lumped capacitors and distributed inductive structures. This is a viable technique for frequencies up to several GHz.

More advanced techniques exist to realise bandpass filters that should be considered when the frequency of operation does not allow the use of hybrid solutions. One intuitive scheme is based on that a parallel resonance circuit can be made to behave as a series resonance circuit with impedance inversion and vice versa, see figure 12.31. The Smith chart reveals that this impedance inversion can be implemented with a $\lambda/4$ transmission line. Otherwise, one of the most popular schemes for bandpass filter realisation is based on field-coupled lines, so called interdigital filters where the individual characteristic impedances of the transmission lines as well as the distance between the lines controls the behaviour of the filter. These techniques requires a more comprehensive treatment which is out of the scope in this chapter. Reference 2 is an excellent design book that deals with most kinds of distributed filters.

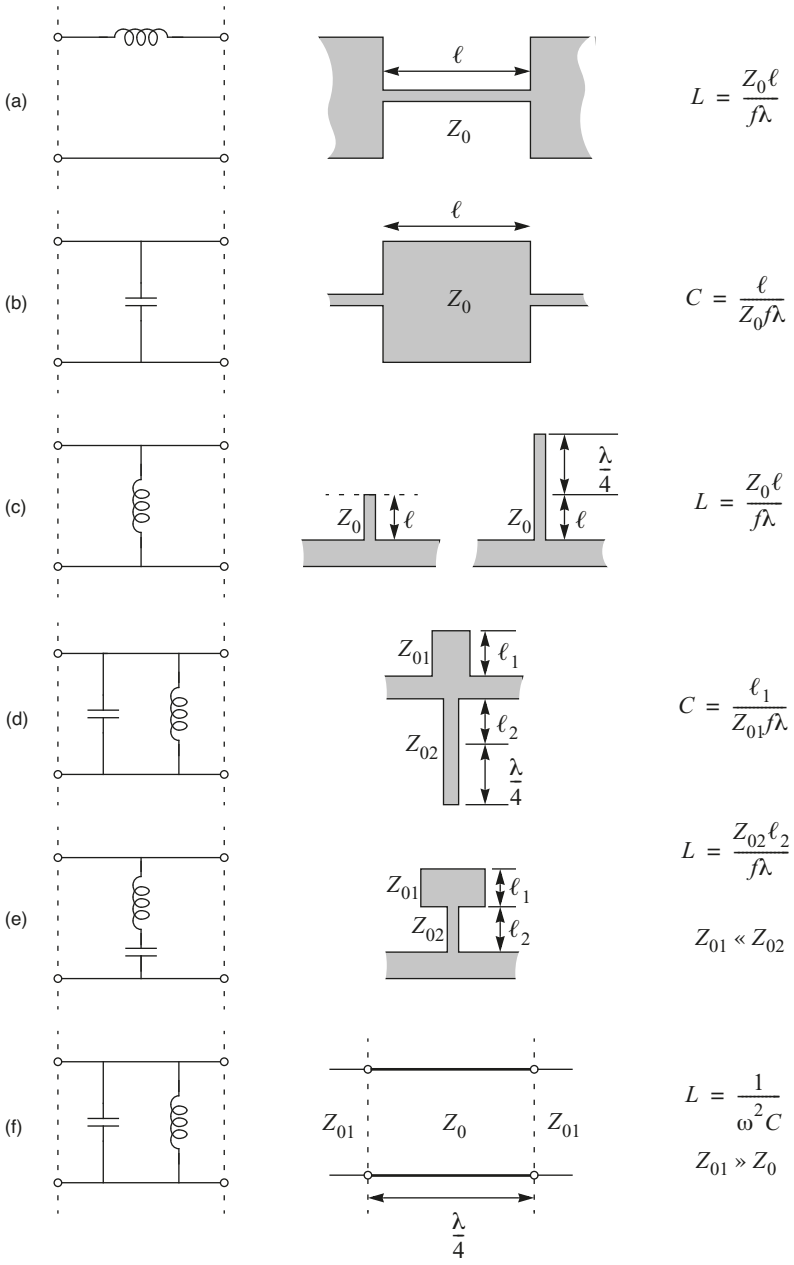


Figure 12.30 Transmission line (microstrip or stripline type) representation of different lumped elements. Dashed terminations of stubs represent ground.

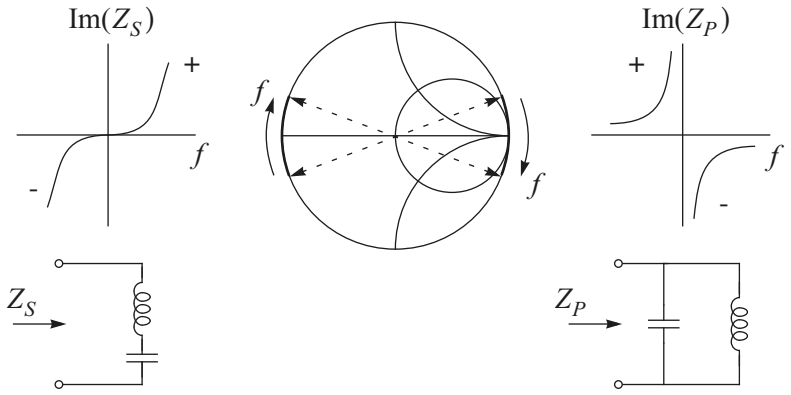


Figure 12.31 The effect of impedance inversion for a resonance circuit.

12.5 References

[1] V. F. Fusco, *Microwave Circuits: Analysis and Computer-aided Design*, Prentice-Hall International, 1987.

[2] G. Matthaei, L. Young and E. M. T. Jones, *Microwave filters, impedance-matching networks, and coupling structures*, Artech House, 1985.

Chapter 13

Mixers

So far we have assumed the active components to be linear. In fact, the small-signal models used in the previous chapters are only valid for weak signal amplitudes. In this chapter we will explore the nonlinear behaviour of devices in order to produce new frequencies not present at the input. Such circuits are termed mixers. The ability to "mix" signals arises either from a nonlinear characteristic of components or a nonlinear transfer function of a circuit. Thus mixers may utilise diodes, BJT's or FET's. The design choices depends on considerations of gain, noise figure, stability, dynamic range and possible generation of undesired frequency components

Mixers are used in a wide range of applications where there is a need to transpose frequencies or frequency bands, as well as creating new frequencies:

- **Modulator.** In amplitude-modulated systems a mixer is often used to modulate the carrier frequency.
- **Transmitter Converter.** If the modulation is carried out at an intermediate frequency, a mixer can be used to convert the signal to the desired carrier frequency.
- **Receiver Mixer.** In the superheterodyne receiver a mixer is used to convert the received signal to an intermediate frequency.
- **Demodulator.** Almost every kind of modulation can be demodulated by circuits that utilise mixers.
- **Phase Detector.** In phase-locked loops, operating at high frequencies, mixers are used as phase detectors.
- **Frequency Synthesizers.** Mixers are essential parts of the circuitry used for signal generation where they are used for up or down conversion of frequencies.

This chapter emphasises mixers for receiving circuits, but it should be kept in mind that mixers also are used for frequency conversion in transmitters and in instrumentation equipment. The mixer theory also applies to some of the modulator and demodulator circuits presented in later chapters.

13.1 Basic Mixer Theory

When a signal is applied to a nonlinear device, new frequencies will show up in the output due to the *nonlinear transfer function*. Only if the device is perfectly linear the output signal will contain exactly the same frequencies as the input signal. However, if the device has a *multiplying function*, the output will contain the difference and sum of the input frequencies due to the trigonometrical identities 13.2) and 13.3). In general the nature of nonlinearity will dictate what other frequencies that will be generated [1].

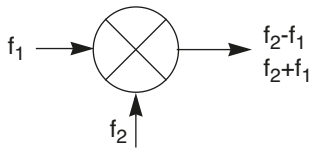


Figure 13.1 The schematic symbol of a mixer.

In figure 13.2 the nonlinear device will produce a set of new frequencies.

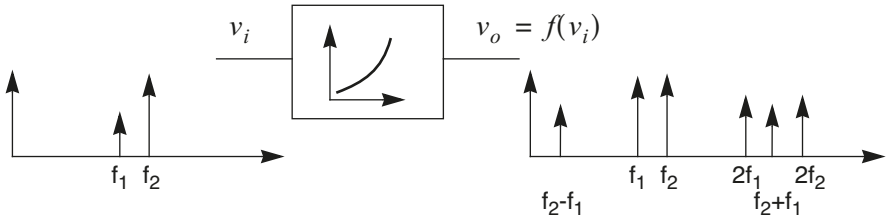


Figure 13.2 A nonlinear device used as a mixer.

One way to analyse the behaviour is to express the input-output relationship in the time domain by a Taylor series:

$$v_o = f(v_i) = a_0 + a_1(v_i) + a_2(v_i)^2 + a_3(v_i)^3 + \ldots \tag{13.1}$$

DC term — linear term — higher order of terms

By use of the trigonometrical identities

$$(\cos\omega t)^2 = \frac{1}{2}(1 + \cos 2\omega t) \tag{13.2}$$

$$\cos\omega_1 t \cdot \cos\omega_2 t = \frac{1}{2}[\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t] \tag{13.3}$$

it is clearly seen that the higher order terms will produce new frequencies that are multiples or combinations of the input frequencies. The sum and difference frequencies generated by the squared term in equation (13.1) are called *second-order intermodulation products* and those generated by the cubed term are *third-order* products.

Usually only the difference-frequency *or* the sum-frequency output component is desired. The DC term, the input frequencies and their harmonics are of no interest for mixer applications. In practical circuits it may be necessary to filter them out.

Considering that the second-order products are desirable, a square-law device is ideal for mixer applications, since the least number of undesired frequencies will be produced. This is one reason why FET's are commonly used in simple mixer circuits.

Example 13.1 Deduction of the second-order intermodulation products

Assume that a device has the transfer characteristic

$$v_o = a_0 + a_1(v_i) + a_2(v_i)^2$$

If the input signal consists of two frequencies

$$v_i = V_1 \cos\omega_1 t + V_2 \cos\omega_2 t$$

the output voltage is

$$v_o = a_0 + a_1(V_1 \cos\omega_1 t + V_2 \cos\omega_2 t) + a_2(V_1 \cos\omega_1 t + V_2 \cos\omega_2 t)^2$$

The three first terms represents a DC component and the input frequencies. These are of no interest in the mixer operation.

By expanding the last term the second-order intermodulation products are given by

$$a_2(V_1 \cos\omega_1 t + V_2 \cos\omega_2 t)^2$$

$$= a_2[(V_1 \cos\omega_1 t)^2 + (V_2 \cos\omega_2 t)^2 + 2V_1V_2(\cos\omega_1 t \cdot \cos\omega_2 t)]$$

$$= \frac{a_2V_1^2}{2}(1 + \cos 2\omega_1 t) + \frac{a_2V_2^2}{2}(1 + \cos 2\omega_2 t) \quad \text{(DC and second harmonics)}$$

$$+ a_2V_1V_2[\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t] \quad \text{(second-order products)}$$

The first and second term are seen to add some DC voltage and produce second harmonics of the input signals. The final term is called the *product term* and yields the desired output. Note that the sum- and difference-frequency components are proportional to the product $V_1 V_2$ of the input-signal amplitudes.

13.2 Mixer Terminology

Single-ended mixer. Any nonlinear device can serve as a mixer. A transistor amplifier stage is often used as a mixer in simple consumer receivers, assembled with discrete components. To obtain frequency conversion the magnitude of the local oscillator signal should be large enough to drive the stage into compression.

Single-Balanced Mixer. By the use of two (or more) nonlinear devices and by a symmetrical connection of one of the signals, this frequency component and its odd harmonics will be suppressed at the IF output.

Double-Balanced Mixer. At the expense of more complicated circuitry, this approach has both the LO and RF inputs applied in a differential mode. The main advantage is that neither signal appears at the other two ports. That is, the LO signal is suppressed at both the RF and the IF port and so forth. These circuits generally require well-balanced transformers as well as accurate matching of the nonlinear-device characteristics.

Conversion Gain is the ratio of the (IF) output signal power to the (RF) input signal power.

Noise Figure is the signal-to-noise ratio (SNR) at the RF input port divided by the SNR at the IF output port, assuming an input source noise temperature of $T_0 = 290$ K.

Isolation represents the undesired amount of leakage between the mixer ports. For example, the "f_{LO} at RF port isolation" is the amount the f_{LO} signal is attenuated when it is measured at the RF port.

Conversion Compression relates to the level beyond where the curve of IF output power versus RF input power deviates from a linear behaviour. Exceeding this level, additional increases in RF input level does not result in proportional rises in output level. Quantitatively, the conversion compression is the output level reduction in dB below the linear characteristic. Typically, the level at which the compression is 1 or 3 dB is given in mixer specifications.

Dynamic Range is the amplitude range over which the mixer can operate without loss of performance. The lower limit is set by the noise figure and the bandwidth, and the upper limit is set by the conversion compression point of the mixer.

Spurious-Free Dynamic Range is the amplitude range over which the mixer can operate without detectable spurious frequencies. The lower limit is set by the noise figure and the bandwidth, and the upper limit is set by the intercept point of the mixer.

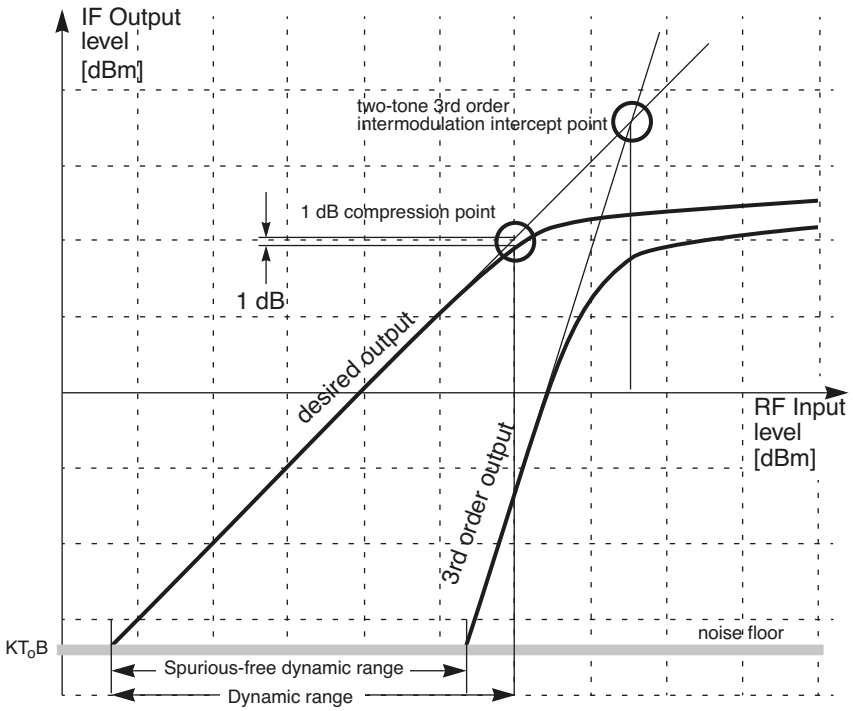


Figure 13.3 The diagram shows the compression point and the theoretical third-order intercept point of a hypothetical mixer.

Two-Tone, Third-Order Intermodulation Distortion is the amount of third-order distortion caused by the presence of a second received signal at the RF port. Mathematically, the third-order distortion is defined in terms of the frequency component at $2f_2 - f_1 \pm f_{LO}$, where f_1 is the desired input signal, f_2 is the second input signal and f_{LO} is the local oscillator signal. Usually, the higher the conversion compression or intercept point of a mixer is, the greater will the suppression of this product be.

Intercept Point is the intersection between the extrapolated fundamental response and the third-order spurious response curves. The higher the intercept point is, the better will the third-order suppression be.

Desensitisation is the compression at the desired signal frequency caused by a strong interfering signal at an adjacent frequency.

Harmonic Intermodulation Distortion results from the mixing of mixer-generated harmonics of the input signals. These distortion products have frequencies $mf_{LO} \pm nf_{RF}$ where *m* and *n* represent the harmonic order.

Cross-Modulation Distortion is the amount of modulation transferred from a modulated carrier to an unmodulated carrier when both signals are applied to the RF input port. The higher the conversion compression or intercept point of a mixer is, the greater will the suppression of the cross-modulation product be.

13.3 General Behaviour of Mixers

Active mixers that have power gain can cause instable operation, as with amplifiers. However, if the LO port is separated from the RF port and all three ports are tuned to different frequencies, stability is not hard to achieve. If the load impedance at each port approximates a short circuit at the other frequencies, the Stern stability criterion indicates that stability will be assured. If the y parameters are known at all three frequencies, the Stern factor test can be applied at each frequency to check the possibility of instability.

The mixer is typically the noisiest stage in the receiver front end. In the case of passive mixers, such as diode mixers, the stage also submits a conversion *loss* and the noise generated in the first IF stage may contribute to the overall noise figure as well. Consequently active mixers that have a conversion *gain* are more attractive in this perspective. Another advantage to the active mixer is lower LO power requirement.

A problem common to all mixer circuits is the spurious response, i.e. the appearance of outputs at the intermediate frequency f_{IF} due to signals at frequencies other than the desired received frequency f_{RF}. These other frequency components may be

- coming directly from the antenna, if no RF preamplifier stage is used
- produced by nonlinear behaviour of the RF amplifier
- produced in the mixer itself
- arising from LO harmonics

If any of these frequencies are present in the mixer stage, the result will be interference with the desired IF signal. Figure 13.4 shows some of these spurious response frequencies.

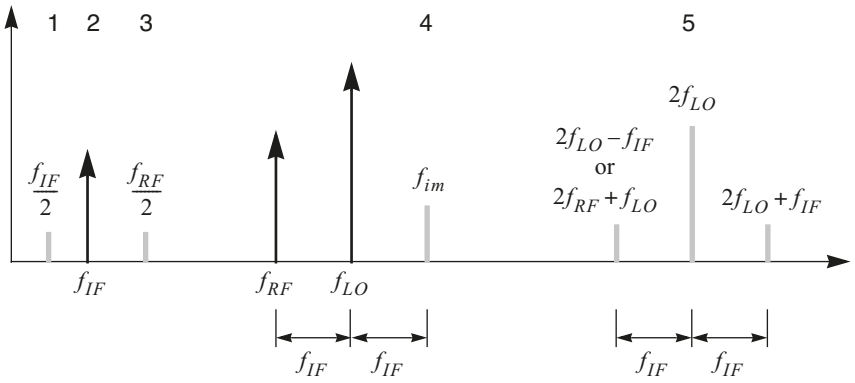


Figure 13.4 A series of spurious signals may produce mixer output at the intermediate frequency.

The desired frequencies are the LO frequency f_{LO}, the RF received frequency f_{RF} and the intermediate frequency $f_{IF} = f_{LO} - f_{RF}$ at the mixer output.

The major sources of unwanted interference are (numbering as depicted in figure 13.4)

1. An input frequency equal to $f_{IF}/2$ may be doubled by the mixer and appear in the output.
2. An input signal at f_{IF} will appear in the output due to normal amplifier behaviour and poor isolation.
3. An input frequency equal to $f_{RF}/2$ may be doubled to f_{RF} by the square-law mixer term and the be mixed with f_{LO} to produce output at f_{IF}.
4. The image frequency $f_{im} = f_{LO} + f_{RF}$. If a signal of this frequency reaches the mixer input, it will be mixed with f_{LO} and a difference-frequency component equal to f_{IF} is produced in the output.
5. If the LO output includes a second harmonic at $2f_{LO}$, or if the mixer generates $2f_{LO}$, this component can mix with the received inputs at $2f_{LO} \pm f_{IF}$ to produce output at f_{IF}.

All of these possibilities illustrate the need for carefully prefiltering in front of the mixer stage as well as good linearity in the RF stage to avoid generation of spurious frequencies at that point.

13.4 Implementation of Mixers

13.4.1 Simple Diode Mixer

A very simple single-ended mixer can be set up as shown in figure 13.5 where both the RF and local oscillator signals are applied to a single diode. Since the performance of this kind of mixer is very poor, the circuit has only the value to illustrate the basic principles of mixers. However, the circuit is frequently used as a diode envelope detector in AM systems which will be discussed in the next chapter.

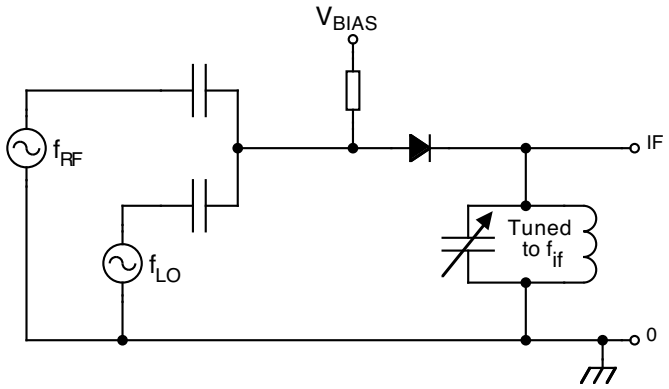


Figure 13.5 A simple diode mixer circuit.

Among several disadvantages the following may be mentioned:

- a significant conversion loss. That is, the IF signal power output is less than the RF signal input power.

- a poor isolation between the input ports, thus increasing the hazard that LO signal may be fed into the receiving antenna.
- a relatively large feed-through of LO frequency to the IF port, which may tend to overload the following IF stage.
- high-order of nonlinearities due to the exponential characteristic of the diode.
- a relatively high noise figure.

13.4.2 Single-Balanced Diode Mixer

As the diode circuits to be described produce both sum and difference of the two input frequencies, they are used as amplitude modulators and demodulators as well as mixers. Therefore the terms "balanced modulator" and "balanced mixer" are synonymous.

A typical single-balanced mixer is shown in figure 13.6. The characteristic of such a mixer is that there is good isolation between the local oscillator and the RF ports relatable to the inherent circuit balance between the LO and RF. However, there is no balance between the RF and IF ports. The output spectrum is shown in figure 13.8.

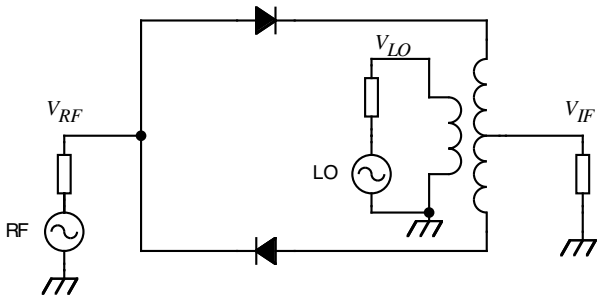


Figure 13.6 A single-balanced mixer offers good isolation between the LO and RF ports, but poor isolation between the RF and IF ports.

The local oscillator voltage V_{LO} is assumed to be large enough to turn the diodes completely on during one of the half-cycles and completely off during the other. It is also assumed that V_{LO} is much greater than V_{RF} so that the local oscillator controls the diode states all the time. The diodes therefore act like switches and causing V_{IF} to be zero when they are open.

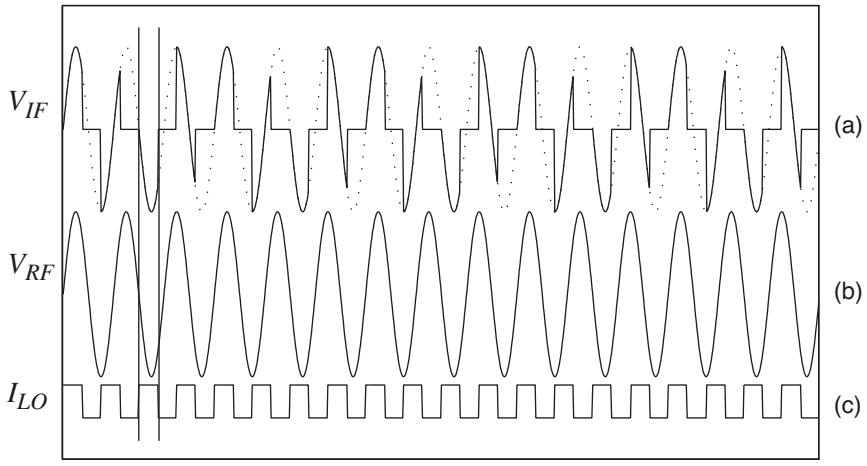


Figure 13.7 Waveforms showing the operation of a single-balanced diode mixer: (a) Output voltage across the resistive load, (b) Input voltage applied at the RF port, (c) switching function caused by the diodes and the local oscillator signal.

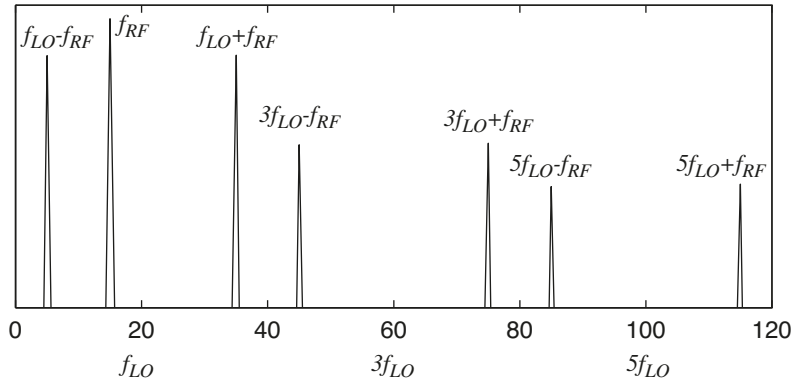


Figure 13.8 The IF output spectrum shows that there is good isolation between the local oscillator and the RF ports. f_{RF} = 15 MHz and f_{LO} = 20 MHz

Due to the square shape of the LO signal, all even harmonics of the oscillator frequency will disappear. Therefore the output spectrum contains only the odd product terms of f_{LO} and f_{RF}. Note that the input signal f_{RF} will also appear in the output with reduced amplitude. All the components except the desired frequency at $f_{IF} = f_{LO} - f_{RF}$ has to be removed by filtering.

At microwave frequencies the parasitic components, especially in the transformer, are hard to avoid. An alternative is to use a microwave coupler as illustrated in figure 13.9. It is assumed that the RF and LO are close enough in frequency to allow the quadrature hybrid to function reasonably well for both signals. The use of open $\lambda/4$ stubs at the RF and LO frequencies at the output is an additional measure to ensure that these signals are minimised at the IF port. An exhaustive description of microwave mixers is found in [2].

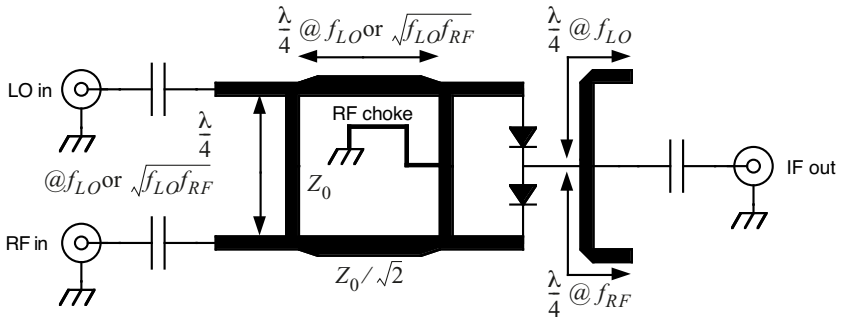


Figure 13.9 Single-balanced microstrip hybrid mixer.

13.4.3 Double-Balanced Diode Mixer

The double-balanced mixer in figure 13.11 provides isolation between all the three ports. The isolation is determined by the balance of the transformer windings and careful matching of the diode characteristics. This is a common type of mixer because of its simplicity and because it will work over a wide

Figure 13.10 Data sheet from Mini-Circuits showing a selection of double-balanced mixers. (www.minicircuits.com)

frequency range. Primarily the frequency range is limited by the transformers. However, if toroidal-cored, transmission-line transformers are used, bandwidths of 1000:1 can be achieved. These mixers typically have a conversion loss of about 6 dB, and a noise figure on the order of 6 to 8 dB. The isolation of the LO from the RF port is around 60 dB, decreasing at higher frequencies because of unbalance due to stray capacitances, and so forth. The two-tone, third order intermodulation products are typically 50 to 60 dB down from the desired IF components.

Let us examine how balance is achieved. The LO voltage at point A and are the same as the centre-tap of the transformer or ground. Therefore, there is no voltage between A and B, and no voltage across the RF or IF ports. Similarly the isolation between the RF and the IF or LO ports is implied by examining the RF voltage at point C and D.

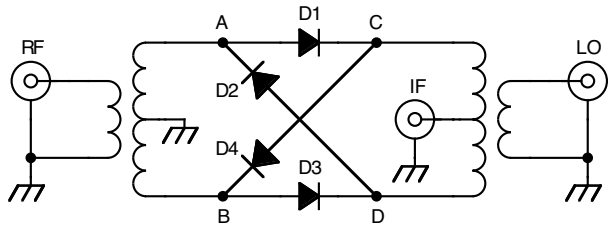


Figure 13.11
The symmetry of a double-balanced diode mixer yields excellent isolation be-

As in the single-balanced mixer, the local oscillator voltage is assumed to be large enough to control the on-off cycle of the diodes. That is, the currents due to the RF signal is small compared with those due to the LO signal. The waveforms are shown in figure 13.12.

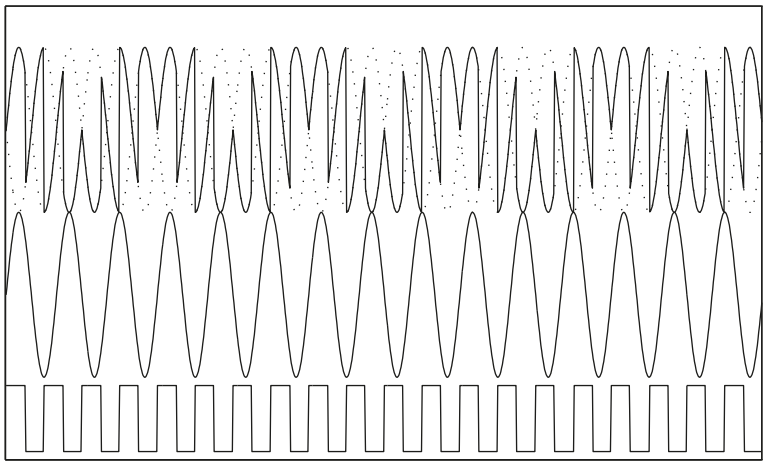


Figure 13.12 The waveforms at the tree ports at a double-balanced diode mixer when $f_{RF} = 15$ MHz and $f_{LO} = 20$ MHz .

The IF output spectrum will contain only the frequencies $nf_{LO} \pm f_{RF}$, with n odd. Neither f_{LO} nor f_{RF} appears in the output as shown in figure 13.13

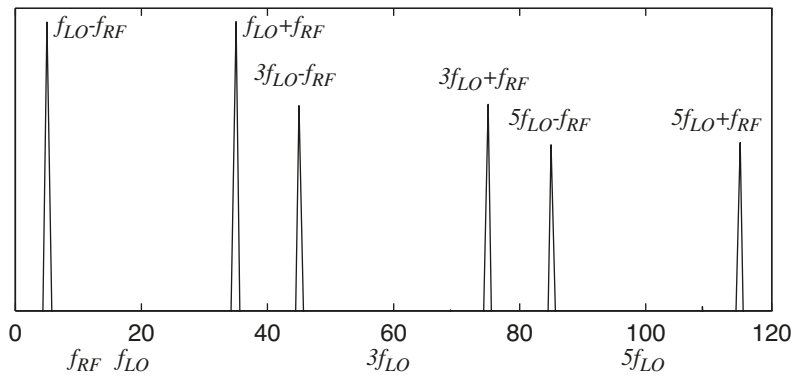


Figure 13.13 IF output spectrum from a double-balanced diode mixer when $f_{RF} = 15$ MHz and $f_{LO} = 20$ MHz

13.4.4 Single Transistor Mixer

The single stage BJT mixer may produce a conversion gain of 30 dB or more. However, as the BJT has an exponential form of transfer function, the third-order intermodulation distortion (IMD) tends to be significant. Another drawback is the limited dynamic range. An ideal mixer should be able to accept a wide range of input amplitudes without producing intermodulation and cross-modulation distortion. This can to a certain extent be carried out by an automatic gain control (AGC) in the RF amplifier. Figure 13.14 shows a typical single transistor mixer. The LO amplitude should be large enough to ensure nonlinear operation in the transistor and the collector tank circuit should be tuned to f_{IF} .

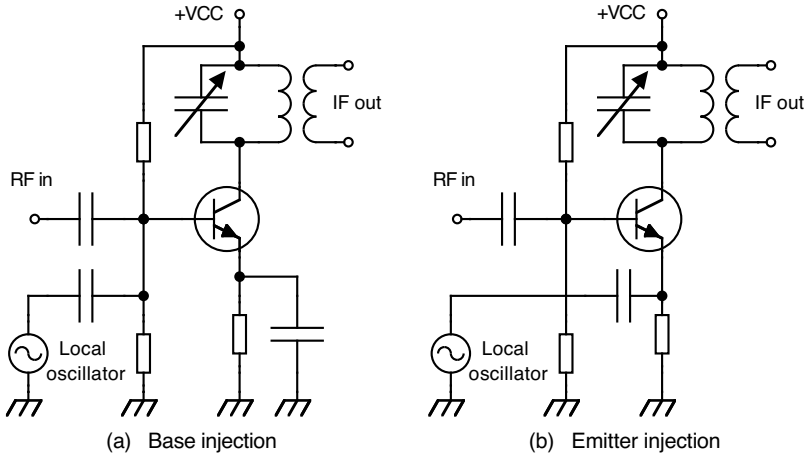


Figure 13.14 Single-ended transistor mixer

Obviously the isolation between the ports is poor. Especially care has to be taken to prevent LO leakage to the receiving antenna by proper filtering in the RF stage. The local oscillator signal can be injected either at the base as in figure 13.14a, or at the emitter as in figure 13.14b. Emitter injection yields slightly lower conversion gain because of the impedance inserted between the emitter and ground but gives a better isolation between the LO and RF ports.

Field-effect transistors are preferred over BJT's for high-frequency mixer applications. Although FET mixers have a lower conversion gain, they produce less intermodulation and cross-modulation distortion because of the square-law transfer characteristic. In addition their lower feedback capacitance provides better circuit stability. Both JFET's and MOSFET's are used, with the latter generally showing higher values of transadmittance and more power gain. With dual-gate MOSFET's the RF and LO inputs can be connected to separate gates, thus improving the isolation between the input ports.

A dual-gate MOSFET is similar in the structure to a single-gate device, except that it includes a second gate between the first gate and the drain. This gate has several effects on the operation. Its primary use is to control the small-signal transconductance of the first gate. Second, because the second gate is usually grounded at the RF frequency, it acts as a shield between the drain and the first gate, reducing the feedback capacitance C_{gd} to a very low value. The low value of C_{gd} , typically less than 0.1 pF, ensures good stability and high maximum available gain.

Dual-gate FET's are usually modelled as two single-gate devices in cascode as shown in figure 13.15.

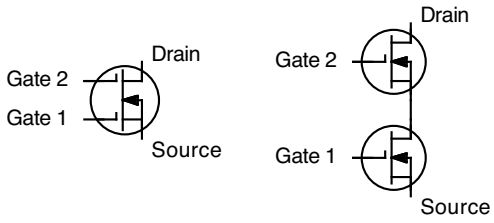


Figure 13.15 (a) Schematic diagram for the dual-gate MOSFET, (b) an equivalent representation showing two MOSFETS in cascode connection.

As would be expected for a cascode stage, larger gain is provided for signals at gate 1 than at gate 2. Accordingly the RF signal is connected to gate 1 and the LO signal to gate 2. Approximately 20 dB of LO-to-RF isolation is thereby achieved. This kind of circuit can often allow the use of a single-ended mixer instead of a balanced mixer, or at least simplify the RF input filter. An example of a single-ended active mixer using dual-gate MOSFET is illustrated in figure 13.16.

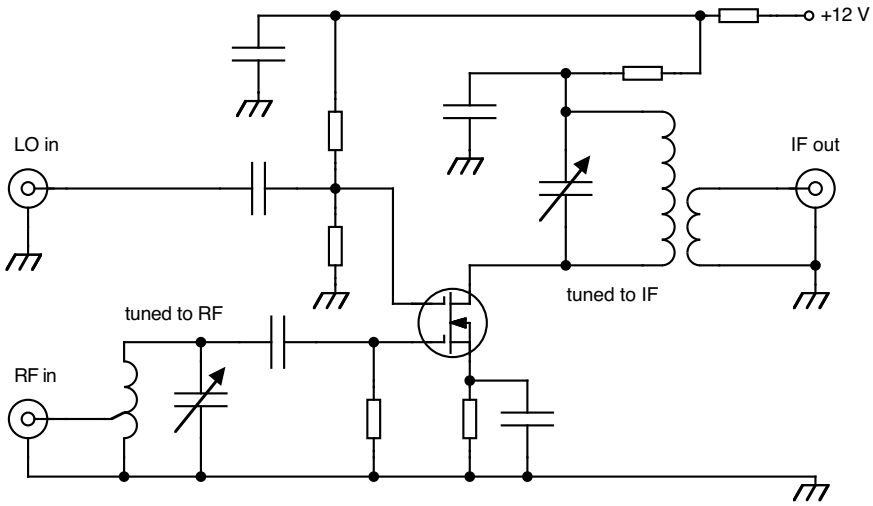


Figure 13.16 Single-ended active mixer using dual-gate MOSFET.

13.4.5 Integrated Mixer

An *analog multiplier* is a circuit that takes two analog inputs and produces an output proportional to their product. The emitter-coupled pair [3] meets this property and can be used as an integrated single-balanced mixer. Integrated circuits such as CA (or LM) 3028 operates to 120 MHz. As seen schematically in figure 13.17 it has an emitter-coupled pair with a (normally) constant

current source which is controlled by the LO signal. Thus, sum and difference frequencies are produced. This IC has numerous applications, including AM modulator and cascode amplifier with AGC.

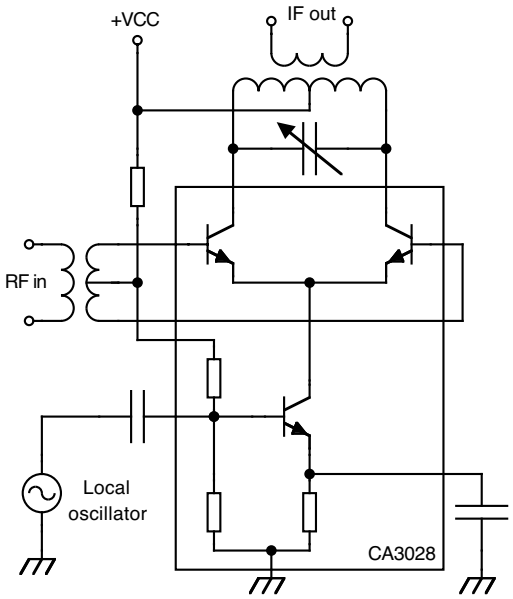


Figure 13.17 Emitter-coupled pair used as mixer.

The basis for most integrated mixers is the Gilbert multiplier cell shown in figure 13.18. The series connection of an emitter-coupled pair with two cross-coupled emitter-coupled pairs produces a particularly useful transfer characteristic. [3]

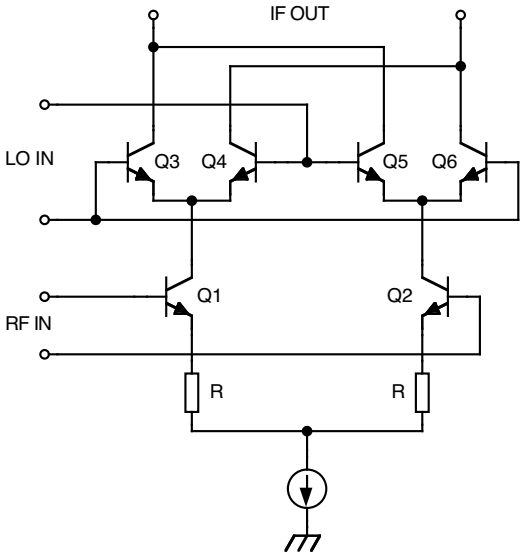


Figure 13.18 Gilbert multiplier circuit.

The lower emitter-coupled pair Q_1 - Q_2 is operated in linear mode for most applications. With no emitter degeneration $(R = 0)$, the maximum input voltage for linear operation is approximately V_T. However, by inserting proper values of R, emitter degeneration can be utilised to increase the linear RF sig-

nal range. Unfortunately, this can not be done with the cross-coupled pairs Q_3 - Q_6 because the degeneration resistors destroy the required nonlinear relation between I_c and V_{be} in those devices.

The upper emitter-coupled pairs may be operated in either a linear or a saturated mode:

- For *low-level operation*, $V_{RF} \ll V_T$, the IF output will contain sum and difference frequency components and have an amplitude which is a function of the product of the input signal amplitudes.
- For *high-level operation*, $V_{RF} > 2V_T$, and linear operation at the RF port, the IF output will contain sum and difference frequency components and the fundamental and odd harmonics of the LO frequency. The output amplitude will be a constant times the RF amplitude. Thus, any amplitude variations in the LO signal will not appear in the output.

13.5 References

[1] H. L. Krauss, C. W. Bostian and F. H. Raab, *Solid state radio engineering*, John Wiley & Sons Inc., 1980.

[2] Stephen A. Maas, “Microwave Mixers” second edition, Artec House, 1993.

[3] Paul Gray/Robert G. Meyer “Analysis and Design of Analog Integrated Circuits” third edition, John Wiley & Sons Inc., 1993

Chapter 14

Power Amplifiers

Power amplifiers (PAs) are used when power efficiency and output power are the main design parameters. Power amplifiers can be designed in many ways depending on the requirements on efficiency, output power and linearity. There are nine commonly accepted classes of operation (A, B, C, D, E, F, G, H and S) that can be found in the literature [1]. What distinguish one class from another is primarily based on how the transistor is biased but also the circuit topology. For example, the small signal amplifiers that were considered in earlier chapters are by definition class A amplifiers. They are biased such that the operation is always confined to the active region of the transistor. It is readily understood that such an amplifier can be used as a power amplifier as well even though it might not always be the optimal choice. The other classes differs from the class A amplifier such that they have biasing or a circuit topology that has been optimised for power efficiency and output power.

Previous chapters have presented amplifier design methodologies based on linear models. At low frequencies it is appropriate to use a more or less simplified hybrid-π model whereas at higher frequencies it is more convenient to use for example S parameters. As the signal power that is to be dealt with increases or the requirements on power efficiency increases these models will gradually become more and more inaccurate. There are two reasons for that. Firstly, a transistor is never perfectly linear by itself and the deviation from a linear behaviour increases with increasing signal levels. In addition to this inherently nonlinear circuit solutions might be needed to obtain a high power efficiency. Secondly, the supply voltage and the circuit topology changes the conditions to optimise gain and output power.

A transistor in nonlinear operation generates distortion. This can be a serious problem by itself but unfortunately it is not the only one. Another problem arises already in the design procedure. There is no point in trying to measure for example S parameters. The values obtained will only be valid for the source and load impedances and the input signal power that was used in the measurement. As a matter of fact, the complexity of a transistor in nonlinear operation makes it impossible for the designer to apply purely analytical models of the transistor and from there try to calculate source and load impedances that result in a certain gain, output power, power efficiency or whatever the specification might contain. Simulation and measurements are used to a large extent over analytical methods.

This chapter begins by discussing two important properties of power amplifiers, namely power efficiency and linearity. This is followed by a survey of some of the most important classes of transistor operation. Design considerations are also brought up. After that measurement techniques are discussed and finally amplifier topologies on block level are described.

14.1 Properties of Power Amplifiers

14.1.1 Power Efficiency

The total power consumption increases with increasing output power. High power consumption is not desired either because of limited battery capacity in the equipment or problems with generation of heat. Therefore, an important parameter for power amplifiers is power efficiency. To be exact, *power efficiency* is the ratio of the RF output power to the DC input power, i.e.,

$$\eta = \frac{\text{RF power delivered to load}}{\text{DC power delivered to amplifier}}$$

In many cases the gain of a power amplifier is so low that the power of the input signal will be a significant part of the total power consumption. In this case it is advisable to refer to the *power-added efficiency*,

$$PAE = \frac{\text{RF power delivered to load - RF power delivered to amplifier}}{\text{DC power delivered to amplifier}}$$

14.1.2 Linearity

Transistors are fundamentally nonlinear, e.g. the bipolar transistor has an exponential transadmittance (and approximately a linear current gain) whereas a FET device has a square-law transadmittance. Furthermore, at high frequencies reactive elements (capacitors) in any transistor will have a significant impact on the behaviour.

In a more general context, systems that have only one of these two properties (nonlinear or reactive) are usually possible to deal with analytically. We can use series expansion of nonlinear systems and from there calculate distortion products with different input signals. Reactive systems can be manipulated in the Laplace-domain to obtain the transfer function and, furthermore, they can be transformed to the time domain to get the impulse response etc. However, for systems having both these properties the analytical methods are scarce and difficult to use. Instead it is common to rely on tools based on numerical methods such as circuit simulation software, e.g. Spice, to investigate the properties of a system. These tools does not give the insight that purely analytical methods might give. On the other hand, complete circuits can be simulated and the results will be as accurate as the models of the components.

While it might be difficult to consider nonlinearities in the design procedure it is much easier to model and measure a given nonlinear system. Even though an amplifier is both nonlinear and reactive it can be modelled as not being reactive (memoryless) under some circumstances. Basically, if the bandwidth of the input signal is small enough compared with the assumed carrier frequency the reactive parts of an amplifier will only contribute to a static phase offset from the input to the output of the amplifier. In other words, the frequency is roughly constant all the time and therefore the impedance of all reactive elements are constant as well. It is therefore justifiable to recapitulate how we can model a nonlinear function with polynomial expansion and what kind of distortion products that come out of it.

14.1.3 Polynomial Expansion

A nonlinear system can be approximated by a polynomial. If x denotes the input signal the output signal is given by

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots \tag{14.1}$$

where more terms increase accuracy. If the system is weakly nonlinear the terms for higher orders decreases rapidly and in many cases (14.1) can be truncated after the third-order term. Subsequent analysis is based on that no higher than third-order products are considered. Three different types of distortion products can be identified, namely harmonic distortion, intermodulation distortion and spurious products. The two former can be derived from (14.1) and this will be discussed briefly below. The latter is the result of more complex processes in a circuit such as parasitic and subharmonic oscillations.

To understand harmonic distortion we apply a single sinusoid $x = A\sin(2\pi ft)$ as input signal to the nonlinear system given by (14.1). The output is given by

$$y = a_0 + a_1 A\sin(2\pi ft) + a_2(A\sin(2\pi ft))^2 + a_3(A\sin(2\pi ft))^3 \tag{14.2}$$

If we expand the trigonometric expressions and approximate by assuming that A is small we get

$$y \approx a_0 + a_1 A\sin(2\pi ft) + \frac{a_2 A^2}{2}\sin(4\pi ft) + \frac{a_3 A^3}{4}\sin(6\pi ft) \tag{14.3}$$

This expression shows that we get what we refer to harmonic distortion, sinusoids at multiplies of the fundamental frequency f. We can disregard the DC term, a_0, in this particular case since we are only interested in RF signals. Equation (14.3) can be illustrated with respect to the input signal amplitude in a dB-dB diagram as shown in figure 14.1. Here the input amplitude is represented by P_{IN} and the output amplitude by P_{OUT}. The fundamental frequency will have a slope equal to one whereas the second order harmonic will have a slope of two, the third-order harmonic a slope of three and so on. Thus, a system that is assumed to be exercised such that this approximation will be

valid is very simple to characterise. It is common practice to refer to the second and third order intercept points, IP_2 and IP_3, that defines the intersections between the harmonics and the fundamental frequency as illustrated in figure 14.1.

Evidently, higher intercept points means better linearity. Note that the intercept points are artifacts as a result of extrapolated characteristics for the fundamental signal and its harmonics and the levels of distortion that they represent cannot be reached. In practice, when the input amplitude is increased these curves will finally saturate as exemplified in figure 14.1. Intercept points are either referred to the input or to the output.

The 1dB-compression point is another important quantity which is defined as the input signal where the actual curve for the fundamental frequency and the extrapolated curve differs 1dB, see figure 14.1.

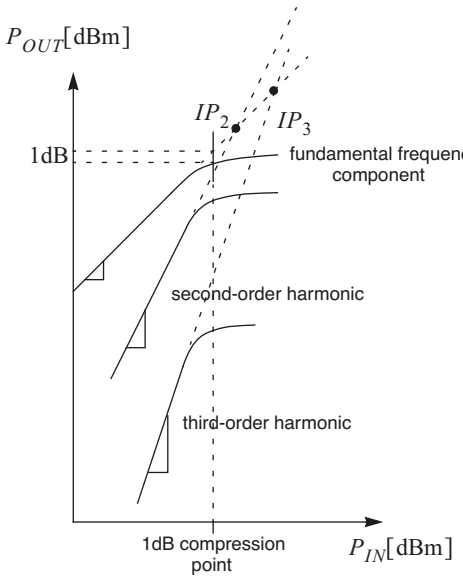


Figure 14.1 Intercept points for harmonic distortion.

Harmonic distortion can be reduced by filtering the amplifier output. A resonance circuit, lumped or distributed, is found at the output in many amplifier designs. This circuit will act as a simple filter that will suppress harmonics, in many cases to acceptable levels.

A larger problem is the intermodulation distortion (IMD) products that appear close to the desired RF signal if the RF signal is amplitude-modulated. This kind of distortion can usually not be filtered because the bandwidth of signal is much smaller than the carrier frequency. The nature of IMD is best understood by considering an input signal consisting of two sinusoids close to each other with equal amplitude (a two-tone test). This corresponds to a DSB-SC (double-sideband-suppressed-carrier) amplitude-modulated signal. If f_c denotes the RF carrier frequency and f_m the modulation frequency there will be one tone at $f_c - f_m$ and one at $f_c + f_m$. If this signal is fed to a nonlinear device the output spectrum will expand close to the input signal as well as in the harmonic zones, see figure 14.2.

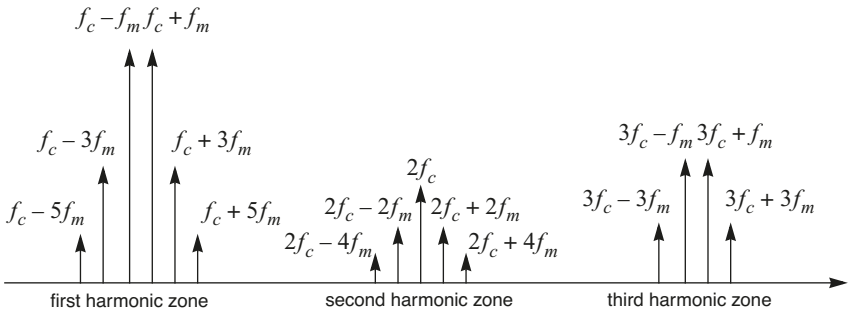


Figure 14.2 Distortion products when the input signal consists of two sinusoids with equal amplitude at $f_c - f_m$ and $f_c + f_m$, respectively.

Expressions for each one of all these frequency components can be developed by applying trigonometric expansion on the polynomial with the input signal, $x = A(\sin(2\pi(f_c - f_m)t) + \sin(2\pi(f_c + f_m)t))$. This is straightforward to do but also tedious. Here we restrict ourselves to assert that we can quantify the IMD products in the same way that we did with harmonic distortion, i.e., we can introduce intercept points for the IMD products, see figure 14.3. However, in this case there will only be odd-order products present, i.e., third-, fifth-order and so on.

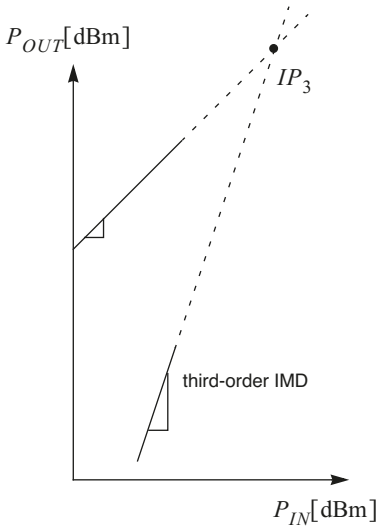


Figure 14.3 Intercept points for IMD.

From the diagrams shown in figures 14.1 and 14.3 it is readily seen that the concept of intercept points provides an easy way to determine distortion products for a given input or output level, or the other way around, to determine an input or output level that corresponds to a certain level of distortion products. Although, one should have in mind that as far as IMD is concerned the intercept points only provides a coarse estimation. The modulating signal is typically not a sinusoid but something very different from a sinusoid that will result in different distortion levels. Still, the two-tone test described above is well-established and in many cases gives a pessimistic estimations of distortion levels.

14.1.4 AM-to-AM and AM-to-PM Conversion

Intercept points are convenient to use for specifying the performance of moderately nonlinear amplifiers. However, many power amplifiers are deliberately nonlinear to such extent that the approximation which the concept of intercept points relies upon is not applicable. Moreover, there is a need for accurate modelling of power amplifiers on system level rather than on circuit level to be able to evaluate effects of nonlinearities in communication systems, e.g. the final power amplifier stage in a transmitter.

One practical way of modelling an amplifier is simply by measuring the output power and the phase shift through the amplifier as a function of the input power with a fixed carrier frequency. This is usually referred to as the AM-to-AM and AM-to-PM conversion for an amplifier. It is also called baseband model because in order to simulate intermodulation products there is no need to calculate the waveform for the modulated carrier signal. The modulating signals (baseband signals) that control the envelope and the phase of the carrier signal are sufficient.

In figure 14.4 below an example of AM-to-AM and AM-to-PM characteristics are shown for a class C amplifier (see section 14.2.2.3) operating at 900MHz. Here the input voltage is the internal peak voltage of the source which has a 50Ω impedance and the output voltage is the peak voltage over a 50Ω load. The phase shift corresponds to the total phase shift in the amplifier including matching networks etc. Three regions can be identified in figure 14.4. For low input amplitudes the gain is very low and the phase shift is approximately constant. The transition to the second region is characterised by a rapid increase in gain as well as phase shift. The second region has an almost linear AM-to-AM conversion and the phase decreases with increasing input amplitude. The last region is characterised by saturation in the AM-to-AM characteristic. Some of the causes to this behaviour is described in section 14.2.2.3 which deals with class C amplifiers.

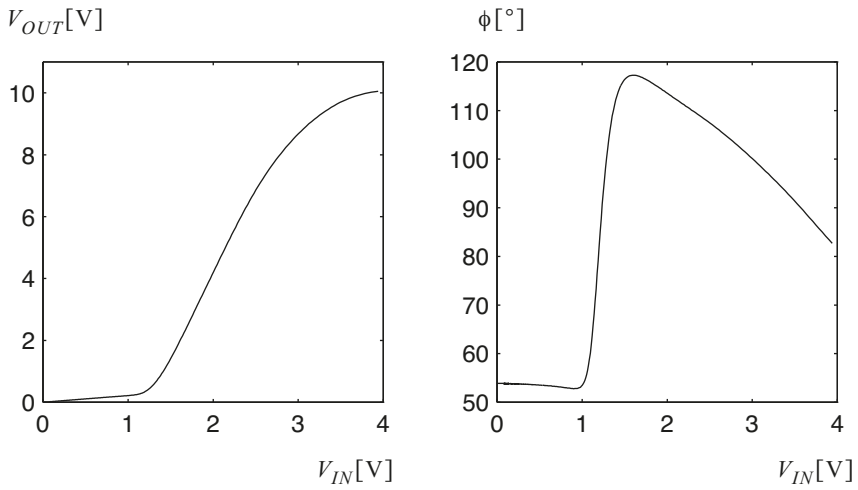


Figure 14.4 Typical AM-to-AM and AM-to-PM characteristics for class C amplifier.

14.2 Amplifier Classes

As mentioned in the beginning of this chapter there are nine different amplifier classes, A, B, C, D, E, F, G, H and S, that are well established. They resemble a basic set of solutions to amplifier operation with various biasing levels and circuit topologies. Other classes is often found in the literature but they are in many cases derivatives of these nine basic classes. Below, the basic properties of the A, B, and C classes are discussed, class A in particular, followed by a short survey of the other classes. For a more detailed treatment please refer to [1].

14.2.1 Operating Regions

To understand the fundamental operation of the various classes we only need to use a simple and ideal model of an active device. Fortunately, it is not even necessary to consider bipolar and FET devices separately.

Three operating regions can be defined for an active device, namely linear-active, cutoff and saturated regions[1], see figure 14.5. In other words, in the active region we assume that a device is a linearly controlled source and the two other regions together constitute an ideal switch with zero resistance in the saturated region and infinite impedance in the cutoff region. Saturation causes the voltage across a device (collector-emitter or drain-source) to take a limiting value, V_{SAT}. In the ideal model this is assumed to be zero if nothing else is stated. This is of course an extremely simplified picture of a very complex device. The model ignores device capacitors and assumes a perfectly linear active region. Nevertheless, when the model is applied to the various classes of operation it will be possible to find the fundamental limit of performance for a given class.

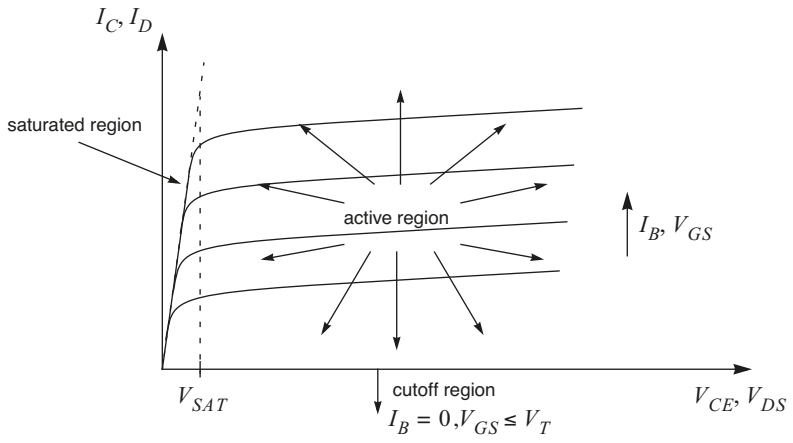


Figure 14.5 Output diagram with the different operating regions. Terms for quantities are given both for bipolar and FET devices.

[1] This is the nomenclature used for bipolar devices. For FET devices the active region is termed saturation region and the region for low drain-source voltages is termed ohmic or triode region.

14.2.2 Classes A, B and C

Common for the A, B and C classes is that they all use active devices that operate in their linear-active region for a significant portion of the RF cycle. High linearity and high efficiency are two conflicting requirements that cannot be obtained simultaneously. The class A, B and C amplifiers are located along a line between these two extremes where class A provides a high degree of linearity, class C high efficiency and class B a little bit of both. Analysis of the performance of these amplifier classes is straightforward, so long as they are operated at power levels below those that cause saturated operation. Saturated operation is more difficult to analyse without resorting to numerical methods and this topic is not discussed here.

14.2.2.1 Class A

A class A amplifier is biased such that operation is confined solely to the active region. Therefore, there is no essential difference between a class A power amplifier and a small-signal amplifier. However, no practical device is perfectly linear and an amplifier that is operated at PEP (peak-envelope-power) will have third-order IMD products no better than 30dB below the desired signal.

A basic common-emitter bipolar class A amplifier for low frequencies is depicted in figure 14.6a.

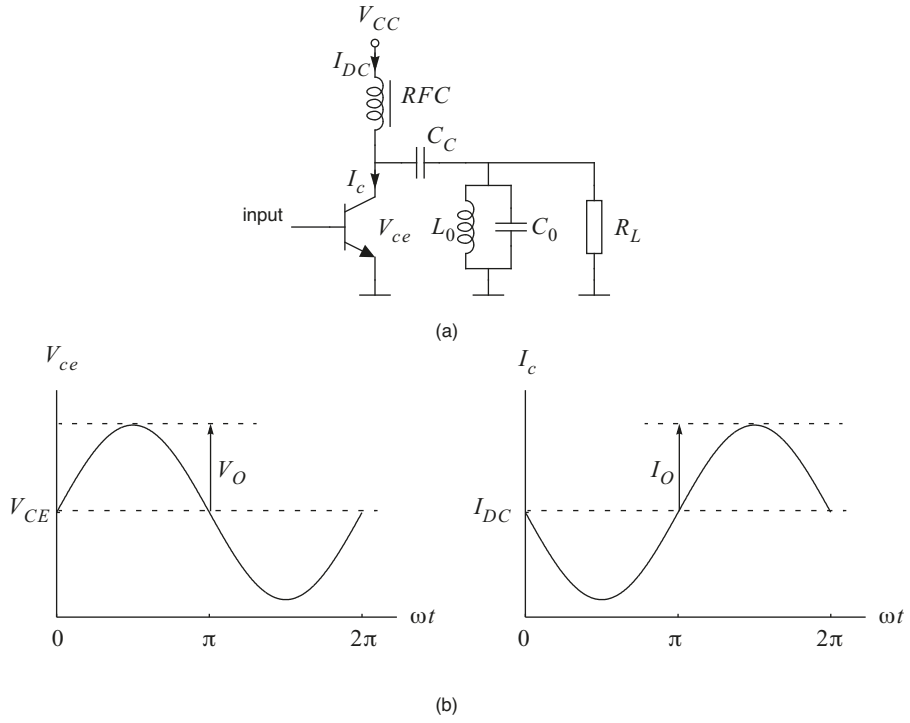


Figure 14.6 Class A amplifier
(a) basic circuit topology (b) class A operation.

The load is ac-coupled to the device and a high impedance RFC connects the device to the supply voltage. In addition to this there is an optional parallel-tuned circuit (L_0, C_0) that will suppress harmonics generated by the moderately nonlinear device. Note that in this cases the inductor in the tuned circuit can replace the RFC to provide the supply voltage to the transistor. In figure 14.6b the collector voltage and current for a sinusoid input is shown where V_O denotes the output voltage amplitude and I_O the output current amplitude.

From figure 14.6b it is evident that output voltage amplitude V_O must be smaller than V_{CC} to avoid transistor cutoff. Furthermore, the output current amplitude I_O must be smaller than I_{DC} to avoid saturated operation. The output current and output voltage are related by $V_O = I_O R_L$ .and it is of course desirable that both the output voltage and output current reach there limits for the same signal amplitude. Thus we can write $V_{CC} = I_{DC} R_L$ that yields the optimal load resistance for a given operating point for the transistor. Once we have an optimal load resistance we will also be able to obtain an optimal power efficiency at peak signal level. The DC input power P_{DC} is given by

$$P_{DC} = V_{CC} I_{DC} = \frac{V_{CC}^2}{R_L} \tag{14.4}$$

The output RF power is

$$P_O = \frac{V_O^2}{2R_L} \leq \frac{V_{CC}^2}{2R_L} \tag{14.5}$$

and hence the instantaneous power efficiency η is given by

$$\eta = \frac{P_O}{P_{DC}} = \frac{V_O^2}{2V_{CC}^2} \leq \frac{1}{2} \tag{14.6}$$

In conclusion, if the input signal is a sinusoid and the load is optimised for maximum output power at PEP the maximum efficiency becomes 50%.

An alternative way of investigating the effect of load resistance is to consider the load line in the output diagram, see figure 14.7, which also shows the effect of the saturation voltage. Thus, from figure 14.7 we see that the optimal load impedance is given by

$$R_{opt} = \frac{V_{CC} - V_{SAT}}{I_{DC}} \tag{14.7}$$

For higher frequencies the behaviour of a transistor is far from what we have assumed above, parasitics dominate the performance of the device. In this case a more pragmatic approach can be used where the optimal source and load impedances are found by an iterative measurement procedure, see section 14.3.

As far as biasing is concerned a class A amplifier is equivalent to a small signal amplifier and the techniques described in an earlier chapter are fully applicable.

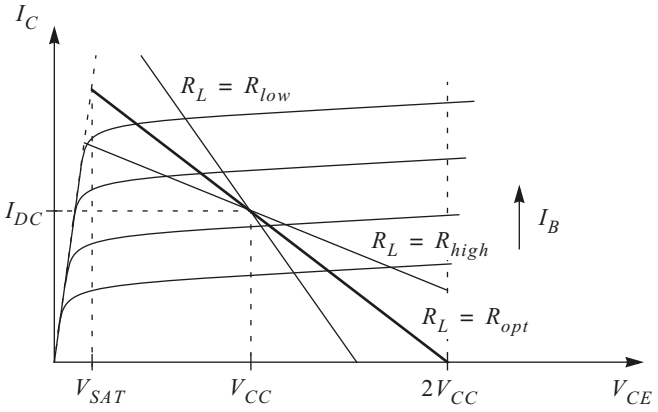


Figure 14.7 Output diagram and load line for three different loads.

14.2.2.2 Class B

In class B operation two devices operate in the active region for exactly half one RF cycle each. When one device operates in the active region the other device is cutoff. At low- and mid-frequencies two alternating devices can be used as show in figure 14.8. Here a transformer-coupled "push-pull" circuit is shown but one can also conceive a version with two complementary devices. One of the most evident benefits with these approaches is that there will be virtually no even-order distortion products if the two devices are well matched. Thus, a tuned circuit might not be necessary and therefore the amplifier can be realised as a broadband amplifier.

A practical limitation for the push-pull class B amplifier is the RF transformer that becomes increasingly difficult to implement due to losses as the frequency is increased. By using complementary devices the transformer is avoided but then again complementary devices are typically poorly matched which can cause crossover distortion as one device starts to conduct at one half-cycle and the other device will be cutoff. For these reasons a tuned single-ended (just one device) version of the class B amplifier is often used at VHF and above at the expense of broadband operation. The circuit topology of a single-ended class B amplifier will be the same as for a class A amplifier with a tuned circuit at the output. The only difference is the biasing level.

In any case, it can be shown that the efficiency of a class B amplifier is given by [1]

$$\eta = \frac{\pi V_C}{4 V_{CC}} \tag{14.8}$$

where V_C is the collector voltage amplitude, see figure 14.8.

As was the case for the class A amplifier the maximum value of V_C possible before saturation occurs is equal to the supply voltage, V_{CC}. Thus, the maximum efficiency becomes $\pi/4$ or 78.5% at PEP. Note that there is no optimal

load impedance as was the case for the class A amplifier which have restricted output voltage as well as restricted output current. Ideally, the output power is given by

$$P_O = \frac{V_C^2}{2R} \leq \frac{V_{CC}^2}{2R} \tag{14.9}$$

where R denotes the resistance as seen by one device with the other device open.

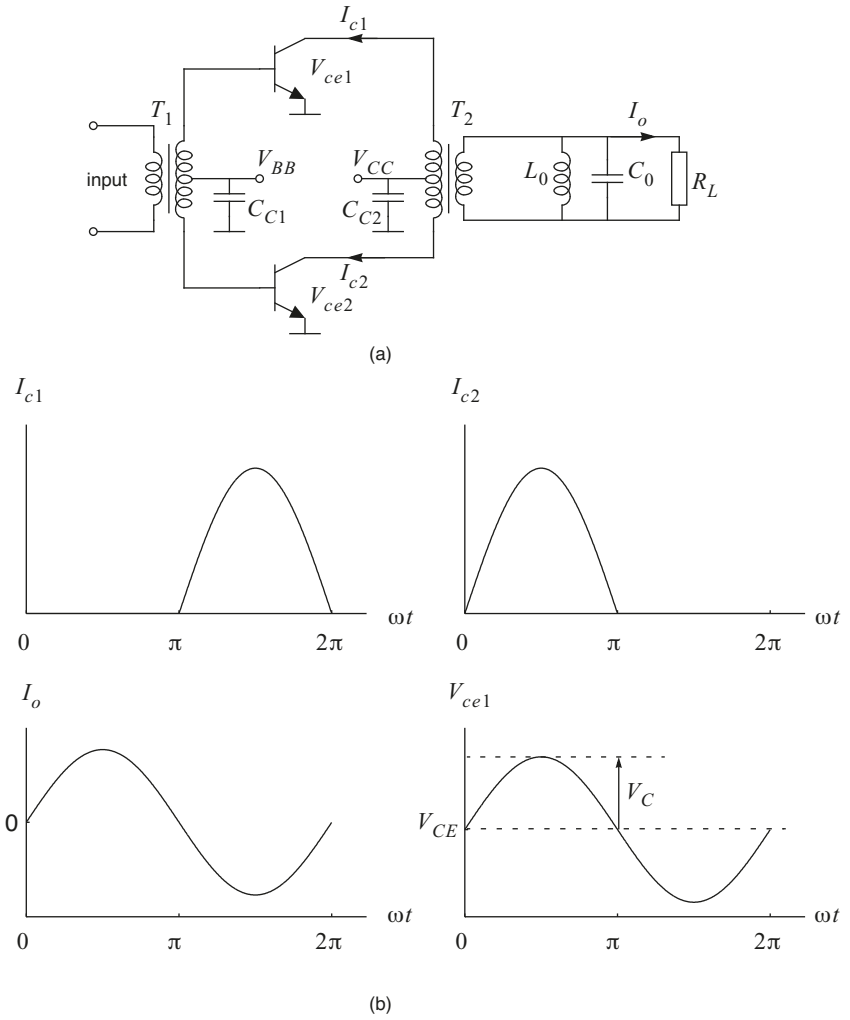


Figure 14.8 Transformer-coupled push-pull class B amplifier (a) circuit topology (b) class B operation.

Linear amplification with class B amplifiers assumes instantaneous transition between active and cutoff regions of operation. Practical devices cannot achieve this and the resulting crossover distortion that is introduced has the effect of reducing the linearity of the amplifier. Crossover distortion is minimised by biasing of the devices. A small quiescent current can be introduced to reduce distortion products at the expense of power efficiency. This mode of operation is sometimes referred to as class AB but class B is also used since a small quiescent current will be needed anyway to have a device conducting in

the neighbourhood of one-half RF cycle. The amount of quiescent current required for minimum distortion is most readily determined experimentally, since theoretical predictions are quite complicated. A quiescent current between 1 to 10 percent of the peak collector or drain current typically results in minimum IMD and the value is usually not critical.

In conclusion an ideal class B amplifier offers improved efficiency over class A operation with preserved linearity. However, in practice the linearity of class B amplifiers is found to be slightly worse than an equivalent class A amplifier.

14.2.2.3 Class C

Class C amplifiers offers even higher efficiency compared to class A and B amplifiers at the expense of linearity. The active device is biased in such a way that is spends significantly less than half an RF cycle in their active region that makes the class C amplifier inherently nonlinear. The circuit topology is the same as for a class A amplifier with a tuned circuit to suppress harmonics, see figure 14.6.

The maximum efficiency at PEP for an ideal class C amplifier is found to be [1]

$$\eta = \frac{2y - \sin 2y}{4(\sin y - y\cos y)} \tag{14.10}$$

where $2y$ denotes the conduction angle, see figure 14.9. As the conduction angle approaches zero the efficiency tends to 100%. At the same time the peak collector (or drain) current tends to infinity to maintain a given output power.

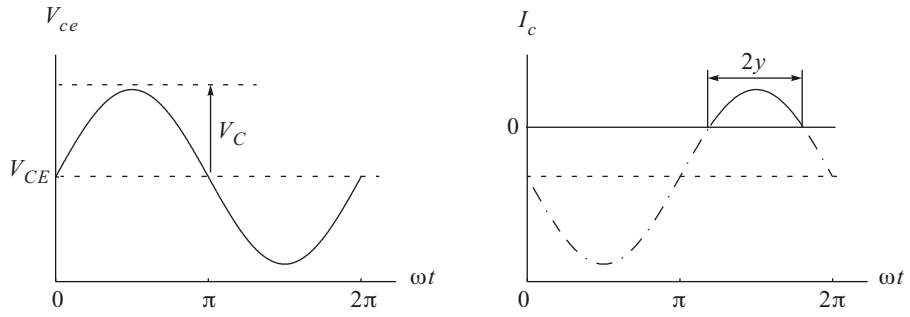


Figure 14.9 Class C operation.

14.2.3 Other Classes of Operation

Class A, B and C amplifiers have one major disadvantage in common. A device that operates in the active region inevitably dissipates power, a current flows through the device at the same time as there is a voltage drop across it. A device that if forced to operate only in cutoff ($I_C = 0$) and saturated

($V_{CE} = V_{SAT} \approx 0$) regions will act as a switch and will therefore ideally not dissipate any power at all. Therefore, an ideal amplifier of this kind has 100% power efficiency but it is also grossly nonlinear. The class D, E and S are based on this technique.

A class D amplifier needs two devices to form a double-pole switch that defines a rectangular voltage or current waveform. The topology for the transformer-coupled class B amplifier in figure 14.8 is one possible solution with the addition of a series-tuned circuit in series with the load that removes harmonics. The switching action in each device makes this class of amplifier operation grossly nonlinear and it is therefore not suitable for amplitude modulated signals. The class E amplifier is a single-ended version of the class D amplifier.

The class F amplifier utilise a single device operating as a current source biased as a class B or C amplifier. The load network have resonances at one or more harmonics as well as at the fundamental frequency. However, only the fundamental frequency is supposed to reach the load. The other resonances are designed to flatten the voltage at the collector to improve power efficiency. The class F amplifier can be found in many forms in the literature and many have their own designated names such as 'biharmonic', 'polyharmonic', 'class CD', 'single-ended class D', 'high efficiency class C' and 'multi-resonator'. The class F amplifier is inherently nonlinear.

Two or more pairs of devices are needed to form a class G amplifier. Each pair operates as a complementary class B amplifier. Furthermore, each pair has its own supply voltage and for low input amplitudes the pair with the lower supply voltage is active. The low voltage pair is cutoff when the input amplitude increases beyond the capability of this circuit and a pair with a higher supply voltage is activated. This technique increases power efficiency. The class G amplifier operates on the instantaneous level of the signal amplitude to decide which pair to activate. This requires the devices to switch rapidly between cutoff and active regions of operation at the signal frequency. For this reason this class of operation is not applicable to high frequency designs.

The class H amplifier consists of a class B amplifier that amplifies the signal. To minimise power dissipation in the device the supply voltage for the class B amplifier is maintained just above the instantaneous level of the signal by a highly efficient, fast and controllable voltage source. This class of operation is also limited to low frequencies due to speed limitations in the voltage source.

Class S amplifiers are highly efficient linear switching amplifiers. Two active devices form a double-pole switch in a similar manner to class D operation but with a switching frequency much higher than the signal frequency. An output filter recovers the slowly varying average value of the switching signal. The average value of the switching waveform is controlled by altering the pulse duty cycle in accordance with the input signal to create a pulse-width modulated waveform. Unfortunately, because the switching frequency needs to be much greater than the signal frequency, practical class S amplifiers are restricted to low frequency applications.

14.3 Measurement Techniques

14.3.1 Measuring Nonlinear Behaviour

Intercept points and AM-to-AM/AM-to-PM conversion represent two ways to quantify a nonlinearity. The former is measured with a spectrum analyser and one or two signal generators as shown in figure 14.10, the latter with a network analyser as illustrated in figure 14.11.

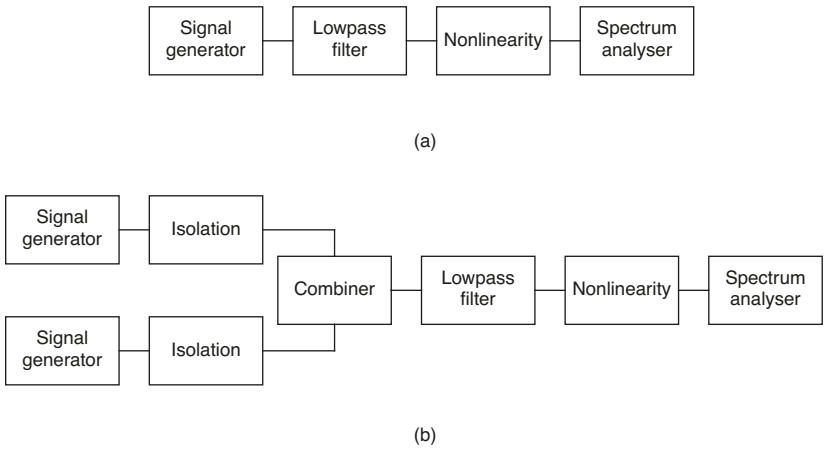


Figure 14.10 Measurement setup to estimate intercept points (a) with one signal generator for characterisation of harmonic distortion (b) with two signal generators for IMD characterisation.

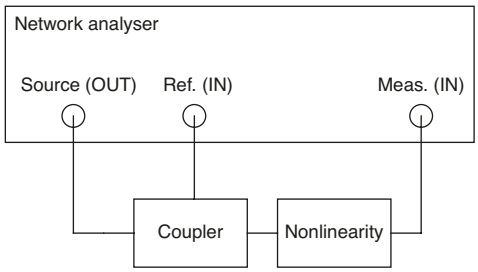


Figure 14.11 Measurement setup for characterisation of AM-to-AM and AM-to-PM conversion.

To estimate intercept points the nonlinear device is exercised with a single tone for harmonics and two tones with a small separation in frequency for IMD characterisation, i.e., equivalent to the analysis in section 14.1.3. Levels of relevant distortion products are measured using a spectrum analyser for different input signal levels. Finally, intercept points are estimated from these measurements by extrapolating the values.

Both measurement setups in figure 14.10 contains a lowpass filter in front of the device. This filter is needed to suppress harmonics from the signal generators which would otherwise interfere with the measurements. Furthermore, if the input impedance of the device is nonlinear there will be an undesired interaction between the device and the signal generator resulting in new distortion products. In this case an isolator is required to separate the generator and the device. An attenuator will be sufficient in many cases. The disadvantage with an attenuator is that the generator power level must be increased accordingly which leads to increased levels of harmonics from the generator. If this is a problem a circulator will do better since it provides a high degree of isolation but low attenuation. For IMD measurements interaction between the two generators should also be avoided. Another source of error is the spectrum analyser. Although the spectrum analyser is quite linear it is still not sufficiently linear to be used without precautions. The instrument specification can be used to chose a suitable input power level that will result in distortion products far below those generated in the measured device. In summary, to be able to succeed with distortion measurements the effects of the measurement system should be verified and compared with the expected behaviour of the device that is to be measured.

A network analyser is preferably used to measure AM-to-AM/AM-to-PM conversion. The network analyser should be set to measure the magnitude and the phase of the complex-gain in power-sweep mode with the desired carrier frequency fixed. Figure 14.12 shows that three instrument ports are required. The source port feeds the device with a test signal and the device output is measured with one of the input ports. The second input port (Ref.) is used for calibration purposes to eliminate the influence of cables etc. in the measurement setup. Typically, network analysers have built-in functions to facilitate the calibration procedure.

14.3.2 Optimal Loading

As mentioned in the introductory section we cannot measure S parameters for a transistor in a nonlinear mode of operation. Such a measurement would only be valid for the specific input power and the specific source and load impedances used in the measurement. Instead we have to emulate the source and load impedances that gives the desired properties of the amplifier in terms of gain, power efficiency, output and maybe also linearity. Such a measurement system is illustrated in figure 14.12.

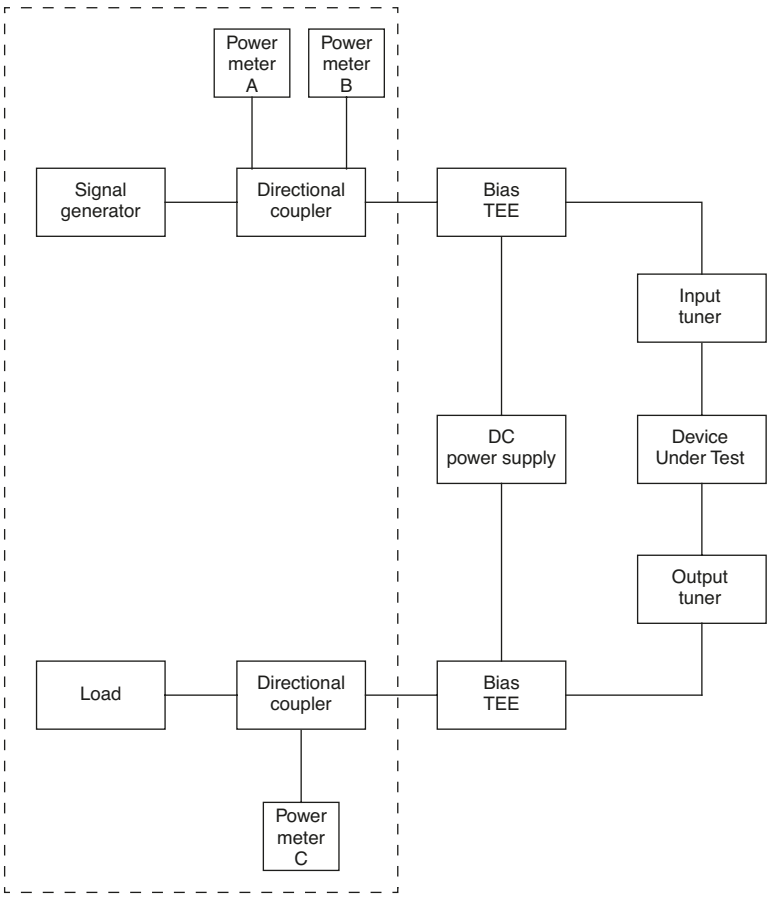


Figure 14.12 Measuring system for large signal parameters. Blocks within dashed frame can be replaced by a network analyser.

With this measuring system it is possible to alter both DC biasing, through bias-TEEs, and impedance levels with tuners, both on the source and the load side. In addition to this we must be able to alter the input power level. To obtain a certain output power with a given available source power the procedure is as follows; The output tuning circuit is adjusted so that a certain power level at the output is obtained (power meter C) and at the same time the input tuner is adjusted to obtain zero reflected power (conjugate match) which can be read from power meter B. The incident input power can be read from power meter A. If other specifications must be met such as power efficiency and linearity the procedure becomes tedious and even more iterative.

Once the desired specification has been met the device is disconnected and a network analyser is used to measure the output impedance of the source and the load tuners. Note that a network analyser can replace all blocks within the dashed frame in figure 14.12 in the tuning procedure described above.

The impedances that were measured should be realised in the design using matching networks to obtain the desired properties again. Note that this measurement technique applies to any class of amplifier operation. Actually, the device in figure 14.12 could be single transistor or a more or less complete design with biasing circuitry etc. whose optimal source and load impedance should be found.

14.4 Amplifier Topologies

In high power applications a single transistor device might not be enough to obtain a certain level of output power. It can also be that a single transistor amplifier is not the most cost-effective solution due to high cost of high power devices. Instead we can use several transistors in parallel and we then have two options. Either we can connect the transistors directly in parallel as shown in figure 14.13 or use a structure based on hybrid combiners and dividers as illustrated in figure 14.14. The former is theoretically possible to use but in many cases impractical. As the number of parallel-coupled transistors are increased the input and output impedances will decrease. This effectively increases the losses in the matching networks because the equivalent series resistance of the components in the matching networks will be in the same order of magnitude as the input resistance of the parallel-coupled transistors. Another disadvantage is that if one transistor fails the complete amplifier fails as well. The hybrid approach overcomes these problems.

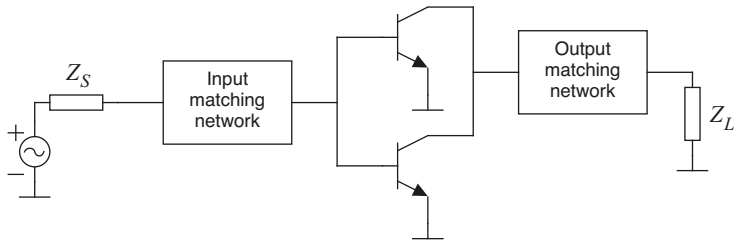


Figure 14.13 Simple but not always a feasible structure for high-power amplifiers.

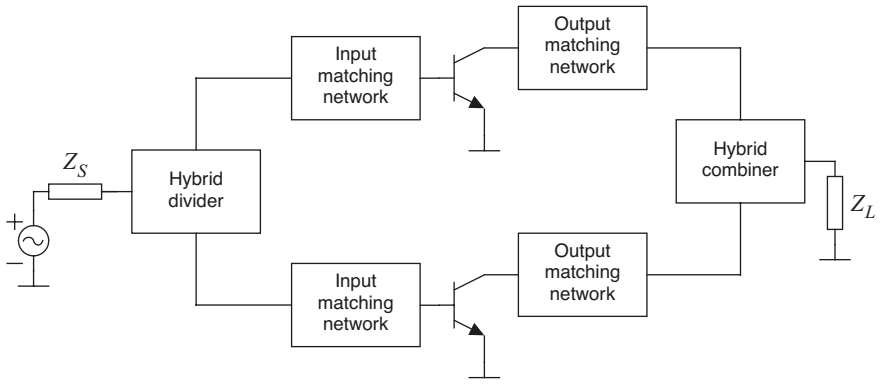


Figure 14.14 High power amplifier structure based on hybrid combiner/ divider.

The most common arrangement for a power amplifier based hybrid combiners and dividers is the balanced amplifier configuration in figure 14.15. Combiner and divider circuits are configured such that the signals are 90° out of phase between ports 2 and 3 (i.e. the signals are in quadrature). Furthermore, if the divider at the input side provides 90° phase shift for the upper branch then the combiner at the output will contribute 90° to the lower branch, i.e.,

both branches will have equal phase shifts. As we will see later there are good reasons to have this arrangement. The Lange coupler and the quadrature coupler are two combiner/divider solutions that provide the desired phase relationships directly but it is also possible to use a Wilkinson coupler where the signals are in phase at ports 2 and 3. One of the ports of the Wilkinson coupler is then extended with a quarter-wavelength transmission line. The net effect will be the same as with the previously mentioned couplers.

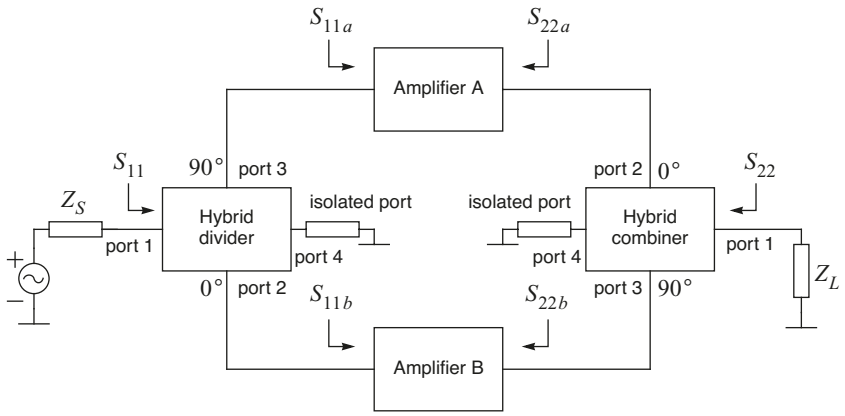


Figure 14.15 Balanced amplifier configuration

If the amplifiers found in each branch is in figure 14.15 are linear enough to be characterised by S parameters, we can write the S parameters for the complete amplifier where we assume that there are no additional phase shifts between the blocks in figure 14.15. The S parameters for the complete amplifier are given by

$$S_{11} = \frac{1}{2}(S_{11b} - S_{11a}) \tag{14.11}$$

$$S_{22} = \frac{1}{2}(S_{22a} - S_{22b}) \tag{14.12}$$

$$S_{12} = j\frac{1}{2}(S_{12a} + S_{12b}) \tag{14.13}$$

$$S_{21} = j\frac{1}{2}(S_{21a} + S_{21b}) \tag{14.14}$$

These expressions are quite intuitive. If the amplifies are identical both branches will provide the same gain and no reflection at the input and the output. The latter property is an effect of the quarter-wavelength difference between the branches at the input and output sides, respectively. For example, consider what happens with a signal that is fed to the input of the complete amplifier. If each branch amplifier reflects the signal equally the signal in the upper branch will be half a wavelength delayed compared with the signal in

the lower branch when they return to the divider again. Due to the phase properties of the divider these signal will be in anti-phase at the divider input but in-phase at the isolation port where the signal power will be dissipated.

Due to the isolating properties of the hybrid combiners and dividers this arrangement will not fail completely if one device breaks. In a balanced amplifier structure as the one shown in figure 14.14 there will only be a reduction in gain.

This amplifier structure can be extended to have more than two branches. For example, the two outputs of one divider can feed the inputs of two other dividers to split the signals to four ports. The combining circuitry can be designed in the same way. In this case it is important to note that the innermost dividers/combiners should provide the $90°$ phase shift whereas the others should be symmetrical in phase. One problem with this scheme is that as the number of branches increases the power efficiency of the divider and the combiner circuits will go down.

14.5 References

[1] H. L. Krauss, C. W. Bostian and F. H. Raab, *Solid state radio engineering*, John Wiley, 1980.

Appendix A

Applying the Indefinite Admittance Matrix to Calculate the Boundaries for Oscillation

The indefinite admittance matrix equation is a matrix form for calculating the transfer function and impedances in an electronic circuit. The circuit is represented by the admittance between different nodes and the corresponding node voltages

$$I = Y \cdot V \qquad \begin{bmatrix} i_1 \\ \dots \\ i_i \end{bmatrix} = \begin{bmatrix} y_{11} & \cdots & y_{1j} \\ \dots & \dots & \dots \\ y_{i1} & \cdots & y_{ij} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \dots \\ v_i \end{bmatrix}$$

The indefinite matrix, Y, has some characteristic properties:

1. $det(Y) = 0$
2. The sum of the elements in a row equals zero $\sum_i y_{ij} = 0$
3. The sum of the elements in a column equals zero $\sum_j y_{ij} = 0$
4. All subdeterminants Y_i^j are equal, deleting an arbitrary row i or column j of the matrix does not make any difference to the resulting determinant.

From linear algebra we also have the following important result:
A determinant of a matrix A that differs from zero, $det(A) \neq 0$, implies that the equation $A \cdot X = 0$ has only the trivial solution $X = 0$. If on the other hand $det(A) = 0$ there exist non trivial solutions $X \neq 0$ to $A \cdot X = 0$

How does this relate to the indefinite admittance matrix?

If one node, noted m, is assumed to be the reference (ground) node, the corresponding row and column is deleted from the indefinite admittance matrix and we get a definite admittance matrix, Y_m^m. If the admittances are chosen so the determinant of this sub-matrix equals zero

$$det(Y_m^m) = 0 \tag{A.1}$$

We will have a non trivial solution to the equation $Y_m^m V = I = 0$. The non trivial solution $V \neq 0$ implies that there can exist voltages in the system without having an external current source driving the system i.e. an self generating circuit.

If we set up the indefinite admittance matrix for a transistor and three surrounding admittances, see figure A.1, like in the generalized oscillator model in chapter 10, it will have the corresponding indefinite admittance matrix as follows:

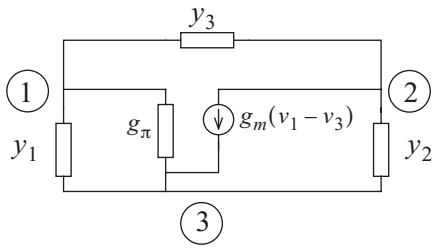


Figure A.1 Simplified transistor model in oscillator circuit with admittances

Attach a fictitious external current source to the node you want to set up the Kirchoff current equations for. This to help get all current directions and voltage references correct. The current source is not necessary since you can apply the Kirchoff current law to the node without it but the authors opinion is that it simplifies the procedure a lot.

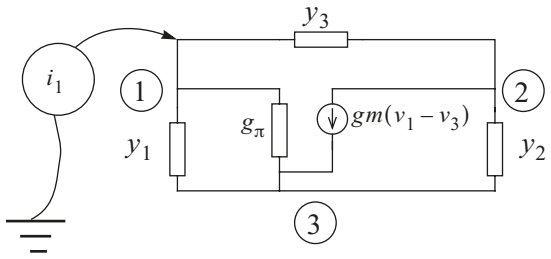


Figure A.2 Fictitious external current source i_1 connected to node 1

The current can then be calculated as

$$i_1 = y_1(v_1 - v_3) + y_3(v_1 - v_2) + g_\pi(v_1 - v_3)$$

Extracting the voltages gives us the first row in the matrix Y

$$\begin{bmatrix} i_1 \end{bmatrix} = \begin{bmatrix} y_1 + y_3 + g_\pi & -y_3 & -g_\pi - y_1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

A quick check shows that the rowsum equals zero and the diagonal element (y_{11}) contains all the admittances connected to the node with positive sign. This is a very convenient way of checking that the matrix has been set up right.

The second node contains a current generator that draws current from the node. This should then be subtracted from the external test generator i_2 since its direction is opposite regarding the node.

$$i_2 - g_m(v_1 - v_3) = y_2(v_2 - v_3) + y_3(v_2 - v_1)$$

$$\begin{bmatrix} i_2 \end{bmatrix} = \begin{bmatrix} g_m - y_3 & y_2 + y_3 & -g_m - y_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

The third and final node current is calculated in the same way. Now the current generators have the same direction so $g_m(v_1 - v_3)$ and i_3 are added.

$$i_3 + g_m(v_1 - v_3) = (y_1 + g_\pi)(v_3 - v_1) + y_2(v_3 - v_2)$$

$$\begin{bmatrix} i_3 \end{bmatrix} = \begin{bmatrix} -y_1 - g_m - g_\pi & -y_2 & g_m + g_\pi + y_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

This finally yields a full Y matrix

$$Y = \begin{bmatrix} y_1 + y_3 + g_\pi & -y_3 & -g_\pi - y_1 \\ g_m - y_3 & y_2 + y_3 & -g_m - y_2 \\ -y_1 - g_m - g_\pi & -y_2 & g_m + g_\pi + y_2 \end{bmatrix}$$

A final check that all diagonal passive elements are positive and that each row and column sum are zero shows that the matrix (most likely) has been set up correct. This circuit is free-floating and to calculate gain and impedances one needs a reference node. In this case it is convenient to use node 3, the ground node, as reference. If node 3 is grounded that means that $v_3 = 0$ and column 3 can be deleted from Y.

If we connect all the external current generators to the circuit, (A.3, it is necessary that the sum of the incoming currents must equal zero. One current can be therefore be written as a linear combination of the others. If it is so, the corresponding row can be deleted from Y since it does not contribute any new information.

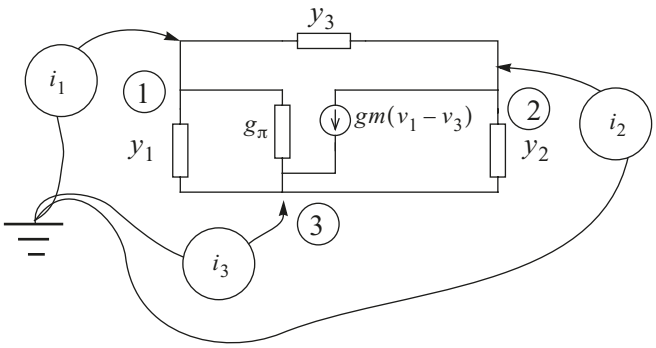


Figure A.3 Current sources connected to circuit.

From the statements in the beginning it follows that any row or column can be deleted if we are only interested in the determinant. Here we eliminate the reference row (3) and column (3). We have then left the submatrix, denoted Y_3^3

$$Y_3^3 = \begin{bmatrix} y_1 + y_3 + g_\pi & -y_3 \\ g_m - y_3 & y_2 + y_3 \end{bmatrix}$$

The determinant of Y_3^3 now becomes

$$y_1y_2 + y_2y_3 + g_\pi y_2 + y_1y_3 + y_3^2 + g_\pi y_3 + g_m y_3 - y_3^2$$

If all reactances are considered lossless, $y_{1\ldots3}$ are purely imaginary in an oscillator. The determinant can now be written as one real and one imaginary part which both must be zero to ensure oscillation (equation (A.1)).

$$Re(det(Y_3^3)) = y_1y_2 + y_2y_3 + y_1y_3 = 0 \tag{A.2}$$

$$\begin{aligned} Im(det(Y_3^3)) &= g_\pi\left(y_2 + y_3\left(1 + \frac{g_m}{g_\pi}\right)\right) \\ &= g_\pi(y_2 + y_3(1 + \beta)) = 0 \end{aligned} \tag{A.3}$$

With $x = 1/y$ and by dividing the terms in the real part with $y_1 y_2 y_3$ this is now be rewritten as

$$x_3 + x_1 + x_2 = 0 \tag{A.4}$$

which is recognised as the criterion for oscillations from chapter 10 Now you can not only use the formula, you can prove it as well and that may give some relief and satisfaction.

For the imaginary part, $g_\pi \neq 0$ and therefore

$$y_2 + y_3(1+\beta) = 0 \Rightarrow 1 + \beta = -\frac{y_2}{y_3} = -\frac{-(x_1 + x_2)}{x_2} \Rightarrow \beta = \frac{x_1}{x_2} \tag{A.5}$$

With a positive β this demands that x_1 and x_2 are of the same type, i.e. capacitors or inductors and their mutual magnitude is determined by β.

As can be seen from these calculations it is easy to use a more complex transistor model with various capacitors and resistors as well as applying a load to the oscillator. The problem is that the expressions tend to grow tremendously if there are more than three nodes in the net and already at five nodes some kind of computer assistance is absolutely needed.

Another way of calculating oscillation would be to find the impedance of a node and see if it is negative or calculate the transfer function and see if the poles are in the right half of the complex plane. More on indefinite admittance matrix and its applications and equations can be found in [1].

A.1 References

[1] H. Floberg, *Symbolic Analysis in Analog Integrated Circuit Design*, Kluwer Academic Press, 1997.

Appendix B

Filter Tables

B.1 Pole Locations for Normalised Filters

Below poles are given for normalised Butterworth, Bessel and Chebyshev filters. The definition of *normalised* is unique for each filter approximation. For the Butterworth filter it means 3dB attenuation at $\omega/\omega_0 = 1$ whereas for the Chebyshev filters it means that the attenuation is equal to the ripple at $\omega/\omega_0 = 1$. For the Bessel filter it is completely different, here the poles are given for unity delay at $\omega/\omega_0 = 0$.

B.1.1 Butterworth Poles

1	2	3	4	5	6	7	8	9	10
-1.0000	-0.7071 ±j0.7071	-0.5000 ±j0.8660	-0.9239 ±j0.3827	-0.8090 ±j0.5878	-0.9659 ±j0.2588	-0.9010 ±j0.4339	-0.9808 ±j0.1951	-0.9397 ±j0.3420	-0.9877 ±j0.1564
		-1.0000	-0.3827 ±j0.9239	-0.3090 ±j0.9511	-0.7071 ±j0.7071	-0.6235 ±j0.7818	-0.8315 ±j0.5556	-0.7660 ±j0.6428	-0.8910 ±j0.4540
				-1.0000	-0.2588 ±j0.9659	-0.2225 ±j0.9749	-0.5556 ±j0.8315	-0.5000 ±j0.8660	-0.7071 ±j0.7071
						-1.0000	-0.1951 ±j0.9808	-0.1736 ±j0.9848	-0.4540 ±j0.8910
								-1.0000	-0.1564 ±j0.9877

B.1.2 Bessel Poles

1	2	3	4	5	6	7	8	9	10
-1.0000	-1.5000 ±j0.8660	-1.8389 ±j1.7544	-2.8962 ±j0.8672	-3.3520 ±j1.7427	-4.2484 ±j0.8675	-4.7583 ±j1.7393	-5.5879 ±j0.8676	-6.1294 ±j1.7378	-6.9220 ±j0.8677
		-2.3222	-2.1038 ±j2.6574	-2.3247 ±j3.5710	-3.7357 ±j2.6263	-4.0701 ±j3.5172	-5.2048 ±j2.6162	-5.6044 ±j3.4982	-6.6153 ±j2.6116
				-3.6467	-2.5159 ±j4.4927	-2.6857 ±j5.4207	-4.3683 ±j4.4144	-4.6384 ±j5.3173	-5.9675 ±j4.3849
						-4.9718	-2.8390 ±j6.3539	-2.9793 ±j7.2915	-4.8862 ±j6.2250
								-6.2970	-3.1089 ±j8.2327

B.1.3 Chebyshev Poles, 0.1dB Ripple

1	2	3	4	5	6	7	8	9	10
-6.5522	-1.1862 ±j1.3809	-0.4847 ±j1.2062	-0.6377 ±j0.4650	-0.4360 ±j0.6677	-0.4280 ±j0.2831	-0.3395 ±j0.4637	-0.3216 ±j0.2053	-0.2729 ±j0.3562	-0.2575 ±j0.1617
		-0.9694	-0.2642 ±j1.1226	-0.1665 ±j1.0804	-0.3133 ±j0.7734	-0.2349 ±j0.8355	-0.2727 ±j0.5847	-0.2225 ±j0.6694	-0.2323 ±j0.4692
				-0.5389	-0.1147 ±j1.0565	-0.0838 ±j1.0418	-0.1822 ±j0.8750	-0.1452 ±j0.9018	-0.1844 ±j0.7307
						-0.3768	-0.0640 ±j1.0322	-0.0504 ±j1.0255	-0.1184 ±j0.9208
								-0.2905	-0.0408 ±j1.0207

B.1.4 Chebyshev Poles, 0.25dB Ripple

1	2	3	4	5	6	7	8	9	10
-4.1081	-0.8983 ±j1.1432	-0.3836 ±j1.0915	-0.5131 ±j0.4377	-0.3535 ±j0.6414	-0.3485 ±j0.2752	-0.2771 ±j0.4539	-0.2630 ±j0.2020	-0.2235 ±j0.3516	-0.2110 ±j0.1600
		-0.7672	-0.2125 ±j1.0568	-0.1350 ±j1.0379	-0.2552 ±j0.7517	-0.1918 ±j0.8180	-0.2230 ±j0.5752	-0.1822 ±j0.6607	-0.1904 ±j0.4642
				-0.4370	-0.0934 ±j1.0269	-0.0684 ±j1.0200	-0.1490 ±j0.8609	-0.1189 ±j0.8902	-0.1511 ±j0.7231
						-0.3076	-0.0523 ±j1.0154	-0.0413 ±j1.0123	-0.0970 ±j0.9111
								-0.2378	-0.0334 ±j1.0100

B.1.5 Chebyshev Poles, 0.5dB Ripple

1	2	3	4	5	6	7	8	9	10
-2.8628	-0.7128 ±j1.0040	-0.3132 ±j1.0219	-0.4233 ±j0.4209	-0.2931 ±j0.6252	-0.2898 ±j0.2702	-0.2308 ±j0.4479	-0.2193 ±j0.1999	-0.1864 ±j0.3487	-0.1761 ±j0.1589
		-0.6265	-0.1754 ±j1.0163	-0.1120 ±j1.0116	-0.2121 ±j0.7382	-0.1597 ±j0.8071	-0.1859 ±j0.5693	-0.1520 ±j0.6553	-0.1589 ±j0.4612
				-0.3623	-0.0777 ±j1.0085	-0.0570 ±j1.0064	-0.1242 ±j0.8520	-0.0992 ±j0.8829	-0.1261 ±j0.7183
						-0.2562	-0.0436 ±j1.0050	-0.0345 ±j1.0040	-0.0810 ±j0.9051
								-0.1984	-0.0279 ±j1.0033

B.1.6 Chebyshev Poles, 1dB Ripple

1	2	3	4	5	6	7	8	9	10
-1.9652	-0.5489 ±j0.8951	-0.2471 ±j0.9660	-0.3369 ±j0.4073	-0.2342 ±j0.6119	-0.2321 ±j0.2662	-0.1851 ±j0.4429	-0.1760 ±j0.1982	-0.1497 ±j0.3463	-0.1415 ±j0.1580
		-0.4942	-0.1395 ±j0.9834	-0.0895 ±j0.9901	-0.1699 ±j0.7272	-0.1281 ±j0.7982	-0.1492 ±j0.5644	-0.1221 ±j0.6509	-0.1277 ±j0.4586
				-0.2895	-0.0622 ±j0.9934	-0.0457 ±j0.9953	-0.0997 ±j0.8448	-0.0797 ±j0.8769	-0.1013 ±j0.7143
						-0.2054	-0.0350 ±j0.9965	-0.0277 ±j0.9972	-0.0650 ±j0.9001
								-0.1593	-0.0224 ±j0.9978

B.1.7 Chebyshev Poles, 2dB Ripple

1	2	3	4	5	6	7	8	9	10
-1.3076	-0.4019 ±j0.8133	-0.1845 ±j0.9231	-0.2532 ±j0.3968	-0.1766 ±j0.6016	-0.1753 ±j0.2630	-0.1400 ±j0.4391	-0.1332 ±j0.1969	-0.1134 ±j0.3445	-0.1072 ±j0.1574
		-0.3689	-0.1049 ±j0.9580	-0.0675 ±j0.9735	-0.1283 ±j0.7187	-0.0969 ±j0.7912	-0.1129 ±j0.5607	-0.0924 ±j0.6474	-0.0967 ±j0.4567
				-0.2183	-0.0470 ±j0.9817	-0.0346 ±j0.9866	-0.0754 ±j0.8391	-0.0603 ±j0.8723	-0.0767 ±j0.7113
						-0.1553	-0.0265 ±j0.9898	-0.0209 ±j0.9919	-0.0493 ±j0.8962
								-0.1206	-0.0170 ±j0.9935

B.1.8 Chebyshev Poles, 3dB Ripple

1	2	3	4	5	6	7	8	9	10
-1.0024	-0.3224 ±j0.7772	-0.1493 ±j0.9038	-0.2056 ±j0.3920	-0.1436 ±j0.5970	-0.1427 ±j0.2616	-0.1140 ±j0.4373	-0.1085 ±j0.1963	-0.0923 ±j0.3437	-0.0873 ±j0.1570
		-0.2986	-0.0852 ±j0.9465	-0.0549 ±j0.9659	-0.1044 ±j0.7148	-0.0789 ±j0.7881	-0.0920 ±j0.5590	-0.0753 ±j0.6459	-0.0788 ±j0.4558
				-0.1775	-0.0382 ±j0.9764	-0.0281 ±j0.9827	-0.0614 ±j0.8365	-0.0491 ±j0.8702	-0.0625 ±j0.7099
						-0.1265	-0.0216 ±j0.9868	-0.0171 ±j0.9896	-0.0401 ±j0.8945
								-0.0983	-0.0138 ±j0.9915

B.2 Ladder Filters

In subsequent sections tabulated element values are given for normalised Butterworth, Chebyshev and Bessel filters. Values are given both for even-order and odd-order filters in which odd-order filter element symbols are denoted by L_i' and C_i'. Elements are numbered starting at the load as shown figure B.1. Note that element values are given for normalised frequency and the load resistance normalised to unity.

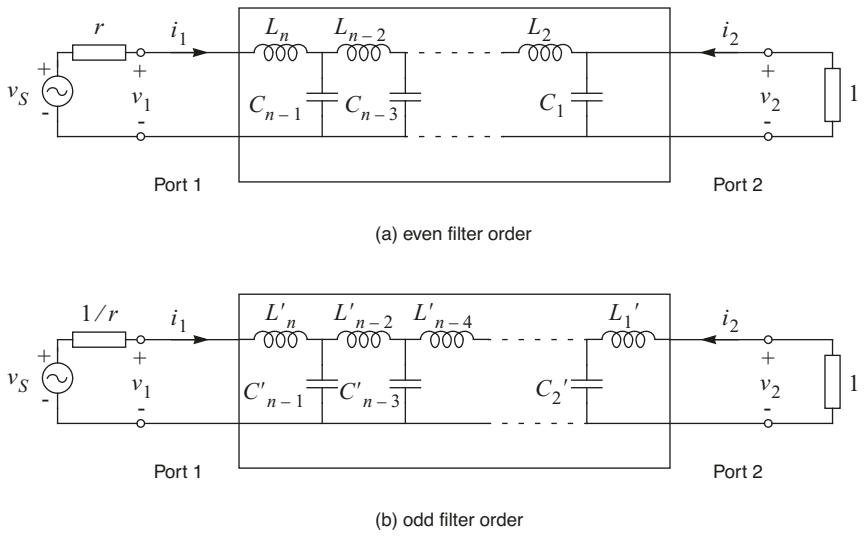


Figure B.1 (a) Even-order filter structure (b) odd-order filter structure.

Different tables exist for a few different source-load resistance ratios (r); 0.1, 0.25, 0.5 and 1. For even-order filters the ratio is defined as $r = R_S/R_L$ and for odd-order filters $r = R_L/R_S$. If the ratio is larger than unity the tables can still be used because of the reciprocity of the filter, i.e., the filter topology can be flipped between the source and the load. One special case is not defined by figure B.1. If the filter-order is odd and the source resistance is zero figure B.1b applies with the source resistance omitted and $r = 0$. The component values should be denormalised as follows:

- Butterworth and Chebyshev filters
  All inductor values should be multiplied by R_S/ω_0 and all capacitor values should be multiplied by $1/\omega_0 R_S$ where ω_0 is the angular normalisation frequency.
- Bessel filters
  All inductor values should be multiplied by $R_S\tau_0$ and all capacitor values should be multiplied by τ_0/R_S where τ_0 is the delay of the filter and is equal to $1/\omega_0$. Thus a normalised Bessel filter always has a group delay equal to unity.

The gain of the filter at $f = 0$ is simply determined by the voltage division between R_S and R_L. From this we conclude that power matching is only obtained for $R_L/R_S = 1$.

B.2.1 Butterworth Ladder Filters

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	1.0000									
2	0.7071	1.4142								
3	0.5000	1.3333	1.5000							
4	0.3827	1.0824	1.5772	1.5307						
5	0.3090	0.8944	1.3820	1.6944	1.5451					
6	0.2588	0.7579	1.2016	1.5529	1.7593	1.5529				
7	0.2225	0.6560	1.0550	1.3972	1.6588	1.7988	1.5576			
8	0.1951	0.5776	0.9370	1.2588	1.5283	1.7287	1.8246	1.5607		
9	0.1736	0.5155	0.8414	1.1408	1.4037	1.6202	1.7772	1.8424	1.5628	
10	0.1564	0.4654	0.7626	1.0406	1.2921	1.5100	1.6869	1.8121	1.8552	1.5643
r=1/8										
1	9.0000									
2	11.9764	0.0939								
3	12.4442	0.1735	4.1674							
4	12.5685	0.2032	8.9296	0.0493						
5	12.6076	0.2169	11.3305	0.1146	2.5343					
6	12.6190	0.2243	12.6794	0.1533	6.1898	0.0330				
7	12.6199	0.2287	13.5040	0.1778	8.5907	0.0835	1.8121			
8	12.6166	0.2314	14.0417	0.1940	10.2279	0.1190	4.6929	0.0248		
9	12.6117	0.2333	14.4102	0.2053	11.3856	0.1446	6.8248	0.0653	1.4086	
10	12.6064	0.2346	14.6730	0.2135	12.2305	0.1635	8.4293	0.0965	3.7698	0.0198
r=1/4										
1	5.0000									
2	6.2741	0.1992								
3	6.3870	0.3608	2.1699							
4	6.3840	0.4180	4.6024	0.1018						
5	6.3636	0.4435	5.8036	0.2350	1.2992					
6	6.3425	0.4567	6.4673	0.3130	3.1601	0.0675				
7	6.3238	0.4641	6.8671	0.3618	4.3727	0.1700	0.9225			
8	6.3078	0.4687	7.1244	0.3940	5.1943	0.2417	2.3838	0.0503		
9	6.2941	0.4716	7.2984	0.4162	5.7720	0.2932	3.4607	0.1325	0.7143	
10	6.2825	0.4735	7.4209	0.4321	6.1916	0.3312	4.2683	0.1955	1.9090	0.0401
r=1/3										
1	4.0000									
2	4.8284	0.2761								
3	4.8473	0.4934	1.6725							
4	4.8105	0.5676	3.5233	0.1386						
5	4.7743	0.5997	4.4239	0.3186	0.9912					
6	4.7446	0.6156	4.9155	0.4233	2.4042	0.0913				
7	4.7206	0.6244	5.2085	0.4882	3.3200	0.2294	0.7006			
8	4.7012	0.6295	5.3950	0.5308	3.9376	0.3258	1.8075	0.0678		
9	4.6853	0.6326	5.5200	0.5601	4.3702	0.3948	2.6209	0.1785	0.5410	
10	4.6720	0.6346	5.6071	0.5809	4.6833	0.4454	3.2293	0.2630	1.4445	0.0540
r=1/2										
1	3.0000									
2	3.3461	0.4483								
3	3.2612	0.7789	1.1811							
4	3.1868	0.8826	2.4524	0.2175						
5	3.1331	0.9237	3.0510	0.4955	0.6857					
6	3.0938	0.9423	3.3687	0.6542	1.6531	0.1412				
7	3.0640	0.9513	3.5532	0.7512	2.2726	0.3536	0.4799			
8	3.0408	0.9558	3.6678	0.8139	2.6863	0.5003	1.2341	0.1042		
9	3.0223	0.9579	3.7426	0.8565	2.9734	0.6046	1.7846	0.2735	0.3685	
10	3.0072	0.9588	3.7934	0.8864	3.1795	0.6808	2.1943	0.4021	0.9818	0.0825
r=1										
1	2.0000									
2	1.4142	1.4142								
3	1.0000	2.0000	1.0000							
4	0.7654	1.8478	1.8478	0.7654						
5	0.6180	1.6180	2.0000	1.6180	0.6180					
6	0.5176	1.4142	1.9319	1.9319	1.4142	0.5176				
7	0.4450	1.2470	1.8019	2.0000	1.8019	1.2470	0.4450			
8	0.3902	1.1111	1.6629	1.9616	1.9616	1.6629	1.1111	0.3902		
9	0.3473	1.0000	1.5321	1.8794	2.0000	1.8794	1.5321	1.0000	0.3473	
10	0.3129	0.9080	1.4142	1.7820	1.9754	1.9754	1.7820	1.4142	0.9080	0.3129

B.2.2 Chebyshev Ladder Filters, 0.1dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.1526									
2	0.4215	0.7159								
3	0.5158	1.0864	1.0895							
4	0.5544	1.1994	1.4576	1.2453						
5	0.5734	1.2490	1.5562	1.5924	1.3759					
6	0.5841	1.2752	1.5999	1.6749	1.7236	1.4035				
7	0.5906	1.2908	1.6236	1.7107	1.7987	1.7395	1.4745			
8	0.5949	1.3008	1.6380	1.7302	1.8302	1.8070	1.8163	1.4660		
9	0.5978	1.3076	1.6476	1.7423	1.8473	1.8343	1.8814	1.7991	1.5182	
10	0.6000	1.3124	1.6542	1.7503	1.8579	1.8489	1.9068	1.8600	1.8585	1.4964
r=1/8										
1	1.3736									
2	5.9892	0.0567								
3	8.9466	0.1403	4.3787							
4	10.0512	0.1866	9.8722	0.0733						
5	11.1128	0.2008	12.7123	0.1602	4.8368					
6	11.2235	0.2179	13.5071	0.2041	10.4600	0.0770				
7	11.8455	0.2179	14.5445	0.2153	13.2359	0.1652	4.9726			
8	11.6822	0.2287	14.4866	0.2312	13.9349	0.2087	10.6567	0.0783		
9	12.1681	0.2248	15.1596	0.2297	14.9251	0.2191	13.4207	0.1672	5.0298	
10	11.9040	0.2336	14.8765	0.2401	14.8149	0.2346	14.0887	0.2106	10.7462	0.0790
r=1/4										
1	0.7631									
2	3.0912	0.1220								
3	4.5446	0.2886	2.3272							
4	5.0046	0.3815	5.0696	0.1559						
5	5.5547	0.4037	6.4880	0.3281	2.5577					
6	5.5377	0.4403	6.7916	0.4167	5.3580	0.1632				
7	5.8904	0.4353	7.3489	0.4330	6.7472	0.3379	2.6256			
8	5.7441	0.4601	7.2417	0.4675	7.0032	0.4258	5.4537	0.1659		
9	6.0374	0.4478	7.6338	0.4594	7.5404	0.4404	6.8374	0.3419	2.6542	
10	5.8435	0.4690	7.4193	0.4839	7.4072	0.4742	7.0778	0.4295	5.4970	0.1672
r=1/3										
1	0.6105									
2	2.3497	0.1712								
3	3.4253	0.3914	1.8216							
4	3.7120	0.5161	3.8671	0.2172						
5	4.1422	0.5389	4.9341	0.4442	1.9966					
6	4.0823	0.5906	5.1046	0.5639	4.0826	0.2272				
7	4.3777	0.5785	5.5512	0.5786	5.1294	0.4572	2.0481			
8	4.2247	0.6155	5.4212	0.6284	5.2644	0.5760	4.1536	0.2308		
9	4.4804	0.5942	5.7535	0.6119	5.6978	0.5885	5.1967	0.4623	2.0697	
10	4.2930	0.6266	5.5455	0.6490	5.5485	0.6374	5.3199	0.5810	4.1857	0.2325
r=1/2										
1	0.4579									
2	1.5715	0.2880								
3	2.2746	0.6035	1.3341							
4	2.3545	0.7973	2.6600	0.3626						
5	2.6921	0.8042	3.3882	0.6853	1.4572					
6	2.5561	0.8962	3.3962	0.8761	2.8071	0.3785				
7	2.8260	0.8560	3.7594	0.8685	3.5246	0.7050	1.4932			
8	2.6324	0.9285	3.5762	0.9619	3.5095	0.8950	2.8547	0.3843		
9	2.8839	0.8762	3.8788	0.9121	3.8660	0.8836	3.5703	0.7127	1.5084	
10	2.6688	0.9429	3.6461	0.9887	3.6707	0.9765	3.5472	0.9027	2.8761	0.3870
r=1										
1	0.3052									
2										
3	1.0316	1.1474	1.0316							
4										
5	1.1468	1.3712	1.9750	1.3712	1.1468					
6										
7	1.1812	1.4228	2.0967	1.5734	2.0967	1.4228	1.1812			
8										
9	1.1957	1.4426	2.1346	1.6167	2.2054	1.6167	2.1346	1.4426	1.1957	
10										

B.2.3 Chebyshev Ladder Filters, 0.25dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.2434									
2	0.5566	0.8499								
3	0.6517	1.2198	1.2248							
4	0.6891	1.3215	1.5979	1.3003						
5	0.6912	1.3538	1.6741	1.6371	1.4480					
6	0.7173	1.3868	1.7271	1.7144	1.8105	1.4193				
7	0.7234	1.3999	1.7475	1.7450	1.8816	1.7497	1.5323			
8	0.7274	1.4083	1.7598	1.7612	1.9099	1.8124	1.8806	1.4647		
9	0.7302	1.4140	1.7678	1.7711	1.9248	1.8365	1.9439	1.7927	1.5648	
10	0.7322	1.4180	1.7733	1.7776	1.9338	1.8490	1.9676	1.8505	1.9119	1.4864
r=1/8										
1	2.1908									
2	7.0446	0.0755								
3	10.0648	0.1562	5.5746							
4	10.4126	0.2051	10.8046	0.0922						
5	11.8024	0.2054	13.8236	0.1738	6.0183					
6	11.2734	0.2298	13.7315	0.2211	11.3063	0.0958				
7	12.3557	0.2179	15.2643	0.2183	14.2825	0.1780	6.1472			
8	11.5987	0.2378	14.4405	0.2423	14.0925	0.2250	11.4700	0.0970		
9	12.5942	0.2228	15.7237	0.2287	15.6054	0.2214	14.4391	0.1797	6.2011	
10	11.7539	0.2414	14.7148	0.2489	14.7268	0.2452	14.2166	0.2267	11.5438	0.0976
r=1/4										
1	1.2171									
2	3.5907	0.1647								
3	5.1234	0.3182	2.9867							
4	5.1282	0.4214	5.4989	0.1990						
5	5.9266	0.4097	7.0768	0.3529	3.2121					
6	5.5080	0.4669	6.8423	0.4540	5.7437	0.2062				
7	6.1786	0.4322	7.7470	0.4354	7.3060	0.3611	3.2774			
8	5.6503	0.4815	7.1583	0.4930	7.0209	0.4618	5.8228	0.2087		
9	6.2867	0.4410	7.9582	0.4543	7.9206	0.4415	7.3830	0.3643	3.3046	
10	5.7179	0.4880	7.2798	0.5050	7.3032	0.4989	7.0810	0.4650	5.8583	0.2099
r=1/3										
1	0.9737									
2	2.6983	0.2337								
3	3.8716	0.4279	2.3508							
4	3.7640	0.5729	4.1610	0.2812						
5	4.4396	0.5433	5.3965	0.4742	2.5236					
6	4.0210	0.6299	5.1006	0.6179	4.3438	0.2910				
7	4.6162	0.5713	5.8745	0.5783	5.5708	0.4851	2.5735			
8	4.1167	0.6480	5.3173	0.6668	5.2361	0.6285	4.4023	0.2945		
9	4.6917	0.5822	6.0240	0.6017	6.0086	0.5864	5.6288	0.4893	2.5944	
10	4.1620	0.6561	5.4002	0.6817	5.4294	0.6749	5.2808	0.6328	4.4285	0.2962
r=1/2										
1	0.7303									
2	1.7288	0.4104								
3	2.5965	0.6465	1.7402							
4	2.2884	0.9039	2.7832	0.4930						
5	2.9282	0.7984	3.7341	0.7177	1.8648					
6	2.4162	0.9771	3.2941	0.9837	2.9094	0.5100				
7	3.0294	0.8341	4.0204	0.8546	3.8585	0.7340	1.9007			
8	2.4631	1.0000	3.4072	1.0463	3.3925	1.0015	2.9490	0.5161		
9	3.0724	0.8478	4.1088	0.8843	4.1199	0.8670	3.8989	0.7404	1.9157	
10	2.4852	1.0100	3.4501	1.0651	3.4927	1.0606	3.4235	1.0085	2.9666	0.5190
r=1										
1	0.4868									
2										
3	1.3034	1.1463	1.3034							
4										
5	1.4144	1.3180	2.2414	1.3180	1.4144					
6										
7	1.4468	1.3560	2.3476	1.4689	2.3476	1.3560	1.4468			
8										
9	1.4604	1.3704	2.3800	1.5000	2.4414	1.5000	2.3800	1.3704	1.4604	
10										

B.2.4 Chebyshev Ladder Filters, 0.5dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.3493									
2	0.7014	0.9403								
3	0.7981	1.3001	1.3465							
4	0.8352	1.3916	1.7279	1.3138						
5	0.8529	1.4291	1.8142	1.6426	1.5388					
6	0.8627	1.4483	1.8494	1.7101	1.9018	1.4042				
7	0.8686	1.4596	1.8675	1.7371	1.9712	1.7254	1.5982			
8	0.8725	1.4666	1.8750	1.7508	1.9980	1.7838	1.9571	1.4379		
9	0.8752	1.4714	1.8856	1.7591	2.0116	1.8055	2.0203	1.7571	1.6238	
10	0.8771	1.4748	1.8905	1.7645	2.0197	1.8165	2.0432	1.8119	1.9816	1.4539
r=1/8										
1	3.1438									
2	7.6905	0.0965								
3	11.1053	0.1646	6.8796							
4	10.3991	0.2234	11.2532	0.1135						
5	12.5367	0.2039	14.9223	0.1801	7.3211					
6	11.0346	0.2434	13.5532	0.2386	11.6839	0.1170				
7	12.9745	0.2131	16.0900	0.2154	15.3437	0.1837	7.4478			
8	11.2694	0.2497	14.0704	0.2556	13.8641	0.2422	11.8216	0.1182		
9	13.1608	0.2167	16.4502	0.2230	16.4112	0.2180	15.4833	0.1852	7.5006	
10	11.3804	0.2524	14.2670	0.2607	14.3247	0.2583	13.9670	0.2436	11.8832	0.1188
r=1/4										
1	1.7466									
2	3.8432	0.2145								
3	5.6859	0.3308	3.7139							
4	5.0293	0.4646	5.6377	0.2504						
5	6.3476	0.4023	7.6862	0.3613	3.9411					
6	5.2985	0.5009	6.6492	0.4963	5.8461	0.2576				
7	6.5476	0.4187	8.2283	0.4252	7.8993	0.3683	4.0061			
8	5.3972	0.5122	6.8706	0.5272	6.8027	0.5035	5.9120	0.2602		
9	6.6323	0.4250	8.3941	0.4388	8.3940	0.4303	7.9689	0.3710	4.0331	
10	5.4437	0.5172	6.9544	0.5364	6.9993	0.5328	6.8522	0.5064	5.9413	0.2614
r=1/3										
1	1.3972									
2	2.8282	0.3109								
3	4.3200	0.4405	2.9371							
4	3.6172	0.6399	4.1985	0.3620						
5	4.7896	0.5293	5.8898	0.4809	3.1130					
6	3.7922	0.6851	4.8770	0.6852	4.3536	0.3722				
7	4.9305	0.5495	6.2770	0.5603	6.0535	0.4901	3.1632			
8	3.8560	0.6990	5.0230	0.7235	4.9937	0.6953	4.4022	0.3759		
9	4.9901	0.5572	6.3947	0.5770	6.4061	0.5671	6.1064	0.4936	3.1841	
10	3.8860	0.7051	5.0780	0.7348	5.1229	0.7314	5.0307	0.6993	4.4237	0.3776
r=1/2										
1	1.0479									
2	1.5132	0.6538								
3	2.9431	0.6503	2.1903							
4	1.8158	1.1328	2.4881	0.7732						
5	3.2228	0.7645	4.1228	0.7116	2.3197					
6	1.8786	1.1884	2.7589	1.2403	2.5976	0.7976				
7	3.3055	0.7899	4.3575	0.8132	4.2419	0.7252	2.3566			
8	1.9012	1.2053	2.8152	1.2864	2.8479	1.2628	2.6310	0.8063		
9	3.3403	0.7995	4.4283	0.8341	4.4546	0.8235	4.2795	0.7304	2.3719	
10	1.9117	1.2127	2.8366	1.2999	2.8964	1.3054	2.8744	1.2714	2.6456	0.8104
r=1										
1	0.6986									
2										
3	1.5963	1.0967	1.5963							
4										
5	1.7058	1.2296	2.5408	1.2296	1.7058					
6										
7	1.7373	1.2582	2.6383	1.3443	2.6383	1.2582	1.7373			
8										
9	1.7504	1.2690	2.6678	1.3673	2.7239	1.3673	2.6678	1.2690	1.7504	
10										

B.2.5 Chebyshev Ladder Filters, 1dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.5088									
2	0.9110	0.9957								
3	1.0118	1.3332	1.5088							
4	1.0495	1.4126	1.9093	1.2817						
5	1.0674	1.4441	1.9938	1.5908	1.6652					
6	1.0773	1.4601	2.0270	1.6507	2.0491	1.3457				
7	1.0832	1.4694	2.0437	1.6736	2.1192	1.6489	1.7118			
8	1.0872	1.4751	2.0537	1.6850	2.1453	1.7021	2.0922	1.3691		
9	1.0899	1.4790	2.0601	1.6918	2.1583	1.7213	2.1574	1.6707	1.7317	
10	1.0918	1.4817	2.0645	1.6961	2.1658	1.7306	2.1803	1.7215	2.1111	1.3801
r=1/8										
1	4.5796									
2	7.9318	0.1286								
3	12.5563	0.1657	8.8038							
4	9.9024	0.2517	11.1584	0.1467						
5	13.7259	0.1945	16.5650	0.1789	9.2596					
6	10.3304	0.2677	12.7878	0.2668	11.5115	0.1503				
7	14.0719	0.2009	17.5013	0.2045	16.9660	0.1819	9.3890			
8	10.4856	0.2725	13.1313	0.2802	13.0465	0.2701	11.6220	0.1516		
9	14.2174	0.2033	17.7827	0.2097	17.8168	0.2066	17.0949	0.1830	9.4427	
10	10.5585	0.2746	13.2602	0.2842	13.3503	0.2828	13.1287	0.2715	11.6709	0.1522
r=1/4										
1	2.5442									
2	3.7779	0.3001								
3	6.5048	0.3264	4.7927							
4	4.5699	0.5428	5.3680	0.3406						
5	7.0522	0.3776	8.6301	0.3520	5.0313					
6	4.7366	0.5716	6.0240	0.5764	5.5353	0.3486				
7	7.2126	0.3888	9.0689	0.3973	8.8368	0.3577	5.0989			
8	4.7966	0.5803	6.1592	0.6005	6.1501	0.5836	5.5869	0.3515		
9	7.2800	0.3930	9.2001	0.4064	9.2344	0.4015	8.9024	0.3598	5.1270	
10	4.8247	0.5841	6.2098	0.6076	6.2689	0.6063	6.1890	0.5864	5.6096	0.3528
r=1/3										
1	2.0354									
2	2.5721	0.4702								
3	4.9893	0.4286	3.8075							
4	3.0355	0.7929	3.7589	0.5347						
5	5.3830	0.4915	6.6673	0.4622	3.9944					
6	3.1307	0.8287	4.1451	0.8467	3.8812	0.5475				
7	5.4978	0.5050	6.9839	0.5177	6.8280	0.4696	4.0473			
8	3.1647	0.8395	4.2237	0.8764	4.2404	0.8580	3.9186	0.5520		
9	5.5459	0.5101	7.0783	0.5288	7.1141	0.5232	6.8785	0.4724	4.0693	
10	3.1806	0.8442	4.2532	0.8851	4.3088	0.8857	4.2691	0.8623	3.9349	0.5541
r=1/2										
1	1.5265									
2										
3	3.4774	0.6153	2.8540							
4										
5	3.7211	0.6949	4.7448	0.6650	2.9936					
6										
7	3.7916	0.7118	4.9425	0.7348	4.8636	0.6757	3.0331			
8										
9	3.8210	0.7182	5.0013	0.7485	5.0412	0.7429	4.9004	0.6797	3.0495	
10										
r=1										
1	1.0177									
2										
3	2.0236	0.9941	2.0236							
4										
5	2.1349	1.0911	3.0009	1.0911	2.1349					
6										
7	2.1666	1.1115	3.0936	1.1735	3.0936	1.1115	2.1666			
8										
9	2.1797	1.1192	3.1214	1.1897	3.1746	1.1897	3.1214	1.1192	2.1797	
10										

B.2.6 Chebyshev Ladder Filters, 2dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.7648									
2	1.2441	0.9766								
3	1.3553	1.2740	1.7717							
4	1.3962	1.3389	2.2169	1.1727						
5	1.4155	1.3640	2.3049	1.4468	1.9004					
6	1.4261	1.3765	2.3383	1.4974	2.3304	1.2137				
7	1.4328	1.3836	2.3551	1.5159	2.4063	1.4836	1.9379			
8	1.4366	1.3881	2.3645	1.5251	2.4332	1.5298	2.3646	1.2284		
9	1.4395	1.3911	2.3707	1.5304	2.4463	1.5495	2.4386	1.4959	1.9553	
10	1.4416	1.3932	2.3748	1.5337	2.4538	1.5536	2.4607	1.5419	2.3794	1.2353
r=1/8										
1	6.8830									
2	7.2895	0.1875								
3	14.9900	0.1541	11.9205							
4	8.5051	0.3093	9.9546	0.2088						
5	15.9745	0.1729	19.4874	0.1646	12.4250					
6	8.7527	0.3224	10.9256	0.3260	10.2196	0.2130				
7	16.2581	0.1769	20.2574	0.1811	19.8989	0.1669	12.5671			
8	8.8412	0.3263	11.1205	0.3368	11.1251	0.3295	10.3007	0.2144		
9	16.3765	0.1784	20.4848	0.1844	20.5920	0.1828	20.0271	0.1678	12.6259	
10	8.8824	0.3280	11.1932	0.3399	11.2954	0.3396	11.1858	0.3308	10.3363	0.2151
r=1/4										
1	3.8239									
2										
3	7.9106	0.2955	6.5423							
4										
5	8.3859	0.3285	10.3300	0.3156	6.8118					
6										
7	8.5220	0.3354	10.7009	0.3443	10.5472	0.3199	6.8877			
8										
9	8.5787	0.3380	10.8103	0.3499	10.8798	0.3475	10.6142	0.3215	6.9191	
10										
r=1/3										
1	3.0591									
2										
3	6.1471	0.3816	5.2161							
4										
5	6.4974	0.4219	8.0681	0.4076	5.4294					
6										
7	6.5974	0.4302	8.3415	0.4425	8.2389	0.4131	5.4893			
8										
9	6.6391	0.4334	8.4220	0.4493	8.4834	0.4467	8.2913	0.4152	5.5141	
10										
r=1/2										
1	2.2943									
2										
3	4.3975	0.5326	3.9184							
4										
5	4.6265	0.5835	5.8503	0.5698	4.0790					
6										
7	4.6917	0.5941	6.0293	0.6136	5.9780	0.5776	4.1242			
8										
9	4.7187	0.5980	6.0821	0.6220	6.1370	0.6195	6.0168	0.5805	4.1429	
10										
r=1										
1	1.5296									
2										
3	2.7107	0.8327	2.7107							
4										
5	2.8310	0.8985	3.7827	0.8985	2.8310					
6										
7	2.8650	0.9120	3.8774	0.9537	3.8774	0.9120	2.8650			
8										
9	2.8790	0.9171	3.9056	0.9643	3.9597	0.9643	3.9056	0.9171	2.8790	
10										

B.2.7 Chebyshev Ladder Filters, 3dB Ripple

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	0.9976									
2	1.5506	0.9109								
3	1.6744	1.1739	2.0302							
4	1.7195	1.2292	2.5272	1.0578						
5	1.7409	1.2501	2.6227	1.3015	2.1491					
6	1.7522	1.2606	2.6578	1.3455	2.6309	1.0876				
7	1.7591	1.2666	2.6750	1.3614	2.7141	1.3282	2.1827			
8	1.7638	1.2701	2.6852	1.3690	2.7436	1.3687	2.6618	1.0982		
9	1.7670	1.2726	2.6916	1.3733	2.7577	1.3827	2.7414	1.3380	2.1970	
10	1.7692	1.2744	2.6958	1.3761	2.7655	1.3893	2.7683	1.3774	2.6753	1.1032
r=1/8										
1	8.9787									
2	6.1219	0.2596								
3	17.4070	0.1392	14.8205							
4	6.9104	0.3884	8.2760	0.2861						
5	18.3377	0.1530	22.4760	0.1481	15.3856					
6	7.0661	0.4011	8.8887	0.4087	8.4796	0.2913				
7	18.6027	0.1559	23.1920	0.1600	22.9195	0.1499	15.5441			
8	7.1213	0.4048	9.0096	0.4190	9.0452	0.4128	8.5411	0.2931		
9	18.7129	0.1569	23.4023	0.1623	23.5592	0.1614	23.0554	0.1506	15.6097	
10	7.1470	0.4064	9.0546	0.4219	9.1496	0.4223	9.0917	0.4144	8.5679	0.2939
r=1/4										
1	4.9881									
2										
3	9.3059	0.2625	8.1669							
4										
5	9.7676	0.2866	12.0571	0.2791	8.4724					
6										
7	9.8986	0.2915	12.4111	0.2998	12.2946	0.2826	8.5581			
8										
9	9.9530	0.2934	12.5151	0.3037	12.6097	0.3024	12.3669	0.2839	8.5935	
10										
r=1/3										
1	3.9905									
2										
3	7.2903	0.3358	6.5207							
4										
5	7.6371	0.3652	9.4808	0.3571	6.7635					
6										
7	7.7352	0.3712	9.7463	0.3822	9.6687	0.3615	6.8315			
8										
9	7.7760	0.3734	9.8243	0.3870	9.9043	0.3856	9.7256	0.3632	6.8597	
10										
r=1/2										
1	2.9929									
2										
3	5.2910	0.4618	4.8991							
4										
5	5.5259	0.4993	6.9460	0.4917	5.0821					
6										
7	5.5922	0.5069	7.1256	0.5235	7.0869	0.4979	5.1335			
8										
9	5.6197	0.5098	7.1785	0.5296	7.2454	0.5282	7.1292	0.5002	5.1547	
10										
r=1										
1	1.9953									
2										
3	3.3487	0.7117	3.3487							
4										
5	3.4813	0.7619	4.5375	0.7619	3.4813					
6										
7	3.5185	0.7722	4.6390	0.8038	4.6390	0.7722	3.5185			
8										
9	3.5339	0.7760	4.6691	0.8118	4.7270	0.8118	4.6691	0.7760	3.5339	
10										

B.2.8 Bessel Ladder Filters

n	C_1 or L'_1	L_2 or C'_2	C_3 or L'_3	L_4 or C'_4	C_5 or L'_5	L_6 or C'_6	C_7 or L'_7	L_8 or C'_8	C_9 or L'_9	L_{10} or C'_{10}
r=0										
1	1.0000									
2	0.3333	1.0000								
3	0.1667	0.4800	0.8333							
4	0.1000	0.2899	0.4627	0.7101						
5	0.0667	0.1948	0.3103	0.4215	0.6231					
6	0.0476	0.1400	0.2246	0.3005	0.3821	0.5595				
7	0.0357	0.1055	0.1704	0.2288	0.2877	0.3487	0.5111			
8	0.0278	0.0823	0.1338	0.1806	0.2227	0.2639	0.3212	0.4732		
9	0.0222	0.0660	0.1077	0.1463	0.1811	0.2129	0.2465	0.2986	0.4424	
10	0.0182	0.0541	0.0886	0.1209	0.1549	0.1880	0.2057	0.2209	0.2712	0.4161
r=1/8										
1	9.0000									
2	8.6533	0.0433								
3	7.1426	0.0615	1.3652							
4	6.0700	0.0589	2.3569	0.0127						
5	5.3229	0.0535	2.5118	0.0246	0.5401					
6	4.7803	0.0484	2.4267	0.0283	1.1309	0.00601				
7	4.3691	0.0442	2.2790	0.0288	1.3738	0.0133	0.2881			
8	4.0462	0.0407	2.1256	0.0280	1.4536	0.0168	0.6627	0.00350		
9	3.7848	0.0378	1.9841	0.0267	1.4558	0.0184	0.8666	0.00830	0.1788	
10	3.5682	0.0354	1.8591	0.0254	1.4215	0.0189	0.9718	0.0111	0.4348	0.00228
r=1/4										
1	5.0000									
2	4.6409	0.0898								
3	3.7994	0.1258	0.6973							
4	3.2221	0.1198	1.1956	0.0258						
5	2.8247	0.1084	1.2690	0.0498	0.2731					
6	2.5375	0.0980	1.2231	0.0571	0.5703	0.0121				
7	2.3202	0.0893	1.1470	0.0580	0.6915	0.0268	0.1451			
8	2.1496	0.0823	1.9689	0.0563	0.7306	0.0338	0.3333	0.00704		
9	2.0114	0.0764	0.9973	0.0537	0.7310	0.0369	0.4354	0.0167	0.0899	
10	1.8967	0.0716	0.0342	0.0509	0.7132	0.0379	0.4878	0.0224	0.2184	0.00459
r=1/3										
1	4.0000									
2	3.6330	0.1223								
3	2.9601	0.1700	0.5298							
4	2.5075	0.1613	0.9046	0.0347						
5	2.1981	0.1457	0.9577	0.0669	0.2063					
6	1.9750	0.1316	0.9217	0.0765	0.4300	0.0163				
7	1.8064	0.1199	0.8636	0.0776	0.5207	0.0358	0.1093			
8	1.6740	0.1104	0.8044	0.0753	0.5497	0.0453	0.2509	0.00942		
9	1.5667	0.1026	0.7503	0.0718	0.5496	0.0494	0.3275	0.0223	0.0676	
10	1.4777	0.0962	0.7027	0.0680	0.5360	0.0506	0.3668	0.0299	0.1642	0.00614
r=1/2										
1	3.0000									
2	2.6180	0.1910								
3	2.1156	0.2613	0.3618							
4	1.7893	0.2461	0.6127	0.0530						
5	1.5686	0.2217	0.6456	0.1015	0.1393					
6	1.4102	0.1999	0.6196	0.1158	0.2894	0.0246				
7	1.2904	0.1821	0.5797	0.1171	0.3497	0.0542	0.0735			
8	1.1964	0.1676	0.5395	0.1135	0.3685	0.0683	0.1684	0.0142		
9	1.1202	0.1558	0.5030	0.1081	0.3680	0.0744	0.2195	0.0336	0.0453	
10	1.0569	0.1460	0.4710	0.1024	0.3586	0.0763	0.2456	0.0450	0.1100	0.00925
r=1										
1	2.0000									
2	1.5774	0.4226								
3	1.2550	0.5528	0.1922							
4	1.0598	0.5116	0.3181	0.1104						
5	0.9303	0.4577	0.3312	0.2090	0.0718					
6	0.8377	0.4116	0.3158	0.2364	0.1480	0.0505				
7	0.7677	0.3744	0.2944	0.2378	0.1778	0.1104	0.0375			
8	0.7125	0.3446	0.2735	0.2297	0.1867	0.1387	0.0855	0.0289		
9	0.6678	0.3203	0.2547	0.2184	0.1859	0.1506	0.1111	0.0682	0.0230	
10	0.6305	0.3002	0.2384	0.2066	0.1808	0.1539	0.1240	0.0911	0.0557	0.0187

Index

E

F

G

H

I

J

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z