5 Radio Transmitters and Receivers

The advent of reliable amplifiers led to an explosion of new techniques that helped turn radio from a technology for specialised applications into one that has touched the lives of nearly everyone. In particular, it made possible the transmission of high-quality audio that could be received on inexpensive receivers. Radio became a technology for the masses and this changed the nature of the entertainment industry forever. In the current chapter we will look at the technology that made these developments possible. In particular, we will look at the evolution of radio receiver and transmitter designs, culminating in the superheterodyne of Armstrong. We will consider crucial circuit elements such as oscillators, mixers and power amplifiers. Importantly, we will consider various factors that can affect the performance of these elements (factors such as phase noise) and their impact upon the overall performance of receivers and transmitters.

5.1 Feedback and Oscillators

Feedback is the process whereby some of the output of an amplifier is fed back into the input. A general feedback system is illustrated in Figure 5.1, the amplifier having gain A and feedback gain F. Referring to this Figure we have that $v_{out} = A(v_{in} + Fv_{out})$ and so

$$v_{out} = \frac{A}{1 - FA} v_{in}. \tag{5.1}$$

If feedback is positive (i.e. $FA > 0$), and FA is slightly less than 1, we will have a large gain. This fact was recognised by the famous radio engineer Edwin Armstrong around the year 1914. In the early days of radio, valves were very expensive and had low performance. As a consequence, the process of *positive feedback* was able to provide a means of building cheap and sensitive radio receivers with poorly performing valves. The downside, however, was that the feedback had to be adjusted carefully and quite often needed to be readjusted when the radio was tuned to a new station. An amplifier that uses the process of positive feedback has come to be known as a *regenerative amplifier* and a receiver that employs such amplifiers as a *regenerative receiver*. As we shall find out later, there are other advantages to positive feedback besides an increase in gain. Negative feedback (i.e. $FA < 0$) can also be useful. In particular, if the amplifier gain is very large, we will have $v_{out} = -1/F$, i.e. we can control the properties of the amplifier by its feedback.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

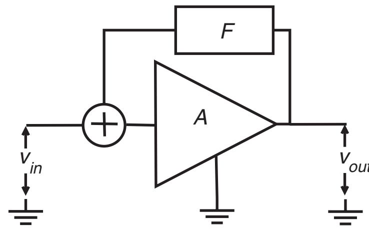


Fig. 5.1 An amplifier with a general feedback network.

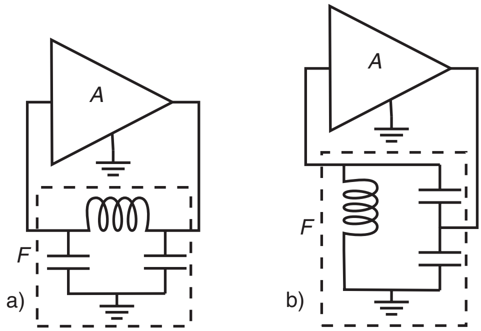


Fig. 5.2 Two forms of Colpitts oscillator.

From Eq. (5.1), we see that, as AF approaches 1, the gain tends towards infinity. Consequently, any small excitation, even circuit noise, will give rise to a substantial output, i.e. the amplifier is unstable. Such instability, however, offers the possibility of signal generation. We consider the situation where the feedback is some kind of frequency-selective circuit. For a harmonic signal ($v_{in} = \Re\{V_{in} \exp(\mathrm{j}\omega t)\}$ and $v_{out} = \Re\{V_{out} \exp(\mathrm{j}\omega t)\}$), Eq. (5.1) can be restated as

$$V_{out} = \frac{A}{1 - FA} V_{in}, \tag{5.2}$$

where F is now a function of frequency and V_{in} and V_{out} are the complex amplitudes of the input and output signals. In order for the amplifier to become unstable, we will need $|FA| \to 1$ and $\arg FA \to 0$ (note that a complex number W can be represented as $|W| \exp(\mathrm{j} \arg W)$ where $\arg W = \arctan(\Im\{W\}/\Re\{W\})$). Let there be a frequency ω_0 at which $\arg(FA) = 0$, then the amplifier will become unstable as $|FA|$ approaches 1, i.e. it will produce oscillations at this frequency. The conditions $|FA| = 1$ and $\arg(FA) = 0$ are known as the *Barkhausen criterion* for oscillation. In a practical amplifier, however, it is difficult to exactly satisfy the condition $|FA| = 1$. Hence, to ensure that the condition can be satisfied, the oscillator designer will choose the gain A such that $|FA|$ is significantly greater than 1. Then, as the amplitude of oscillation rises, gain compression will cause A to fall until the condition is satisfied.

Some examples of feedback are shown in Figure 5.2 and give rise to what is known as a *Colpitts oscillator*. The feedback F in Figure 5.2a is negative and is therefore

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

suitable for use with a common-source (or common-emitter) amplifier (A is negative in such amplifiers). In Figure 5.2b, however, F is positive and so suitable for use with a common-drain (or common-collector) amplifier. We will take a closer look at an oscillator that is based on a common-drain amplifier. A circuit for such an oscillator is shown in Figure 5.3a. It will be noted that we have used a JFET in the design and this allows us to set the quiescent operating point of the transistor by grounding the gate through the inductor L and placing a suitable resistor R_S in the source. In the source circuit there is an RF choke (a large inductance that has effectively infinite impedance at RF frequencies) that isolates the source from the ground at RF frequencies. We will analyse the circuit in terms of the RF model shown in Figure 5.3b (note that the model includes the intrinsic resistance R of the inductor). We assume oscillations at frequency ω and so represent the currents and voltages in terms of their complex amplitudes. From the Kirchhoff current law, we obtain that

$$I_2 = I_1 + V_{GS}g_m. \tag{5.3}$$

Further, from the voltage law we get that

$$RI_1 + j\omega LI_1 + \frac{I_1}{j\omega C_1} + \frac{I_2}{j\omega C_2} = 0 \tag{5.4}$$

and

$$V_{GS} = \frac{I_1}{j\omega C_1}. \tag{5.5}$$

From Eqs. (5.3) and (5.5)

$$I_2 = I_1\left(1 + \frac{g_m}{j\omega C_1}\right) \tag{5.6}$$

and from Eqs. (5.6) and (5.4)

$$I_1\left(R + j\omega L + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} - \frac{g_m}{\omega^2 C_1 C_2}\right) = 0. \tag{5.7}$$

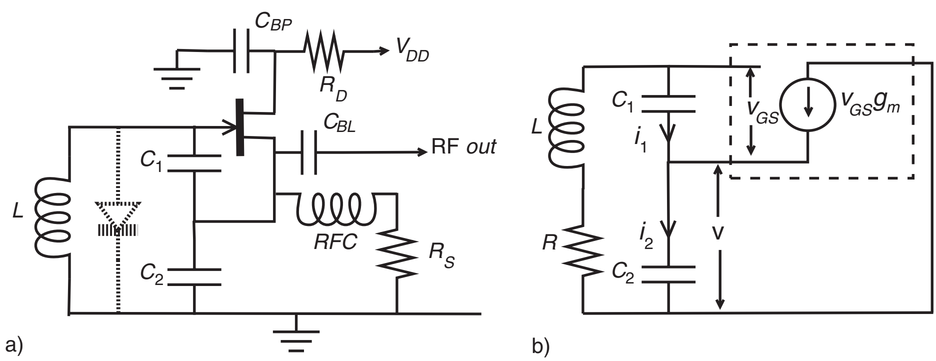


Fig. 5.3 Practical Colpitts oscillator (optional diode shown as dotted lines) and its analysis.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

In order for there to be non-zero current flow, it is necessary that

$$R + \mathrm{j}\omega L + \frac{1}{\mathrm{j}\omega C_1} + \frac{1}{\mathrm{j}\omega C_2} - \frac{g_m}{\omega^2 C_1 C_2} = 0. \tag{5.8}$$

From the imaginary part of Eq. (5.7) we have that

$$\omega^2 = \frac{C_1 + C_2}{LC_1C_2} \tag{5.9}$$

and from the real part that

$$g_m = \omega^2 C_1 C_2 R. \tag{5.10}$$

Equations (5.9) and (5.10) are essentially the Barkhausen criterion. As mentioned earlier, it is difficult to satisfy Eq. (5.10), mainly due to the variation in performance between transistors. If $g_m \neq \omega^2 C_1 C_2 R$, we will need ω to be complex in order for Eq. (5.8) to be satisfied. We can rewrite Eq. (5.8) as

$$1 - \frac{g_m}{R\omega^2 C_1 C_2} + \mathrm{j}\frac{\omega}{\omega_0} Q\left(1 - \frac{\omega_0^2}{\omega^2}\right) = 0, \tag{5.11}$$

where $Q = \omega_0 L/R$ is the unloaded Q of the inductor and $\omega_0 = \sqrt{(C_1 + C_2)/LC_1C_2}$ is the frequency of oscillation when the Barkhausen criterion is satisfied. If we consider small deviations $\delta\omega$ in frequency about ω_0 (i.e. $\omega = \omega_0 + \delta\omega$), then $\omega/\omega_0 \approx 1 + \delta\omega/\omega_0$ and $\omega_0^2/\omega^2 \approx 1 - 2\delta\omega/\omega_0$. Consequently, Eq. (5.11) can be approximately rewritten as

$$1 - \frac{g_m}{R\omega_0^2 C_1 C_2}\left(1 - 2\frac{\delta\omega}{\omega_0}\right) + 2\mathrm{j}Q\frac{\delta\omega}{\omega_0} = 0 \tag{5.12}$$

from which

$$\frac{\delta\omega}{\omega_0} = \frac{\frac{g_m}{R\omega_0^2 C_1 C_2} - 1}{\frac{2g_m}{R\omega_0^2 C_1 C_2} + 2\mathrm{j}Q}. \tag{5.13}$$

It will be noted that $g_m/R\omega_0^2 C_1 C_2$ is positive and, from the Barkhausen criterion, has a value around 1 as the oscillator approaches equilibrium. Further, Q is also positive and usually has a fairly large value (10 or more). We let $\delta\omega = \delta\omega_r + \mathrm{j}\delta\omega_i$ and then the output voltage of the oscillator, the voltage at the source, will be

$$v = \Re\left\{\frac{i_2}{\mathrm{j}\omega C_2}\right\} = \exp(-\delta\omega_i t)\Re\left\{\exp(\mathrm{j}(\omega_0 + \delta\omega_r)t)\frac{I_2}{\mathrm{j}\omega_0 C_2}\right\}. \tag{5.14}$$

From Eq. (5.13), it will be seen that $\delta\omega_i$ (the imaginary part of $\delta\omega$) will be negative if $g_m/R\omega_0^2 C_1 C_2 > 1$ and so the oscillations will grow with time. Consequently, in order to ensure that oscillations start, we need to choose $g_m/R\omega_0^2 C_1 C_2$ to be suitably greater than 1. Then, as the amplitude builds up, the value of g_m will fall due to gain compression and $g_m/R\omega_0^2 C_1 C_2$ will settle down to a value close to 1. From the above considerations, it is obvious that the gain compression in the amplifier is important to the functioning of the oscillator. In order to help this aspect, a reverse-bias diode is often added at the gate of the oscillator of Figure 5.3a in order to enhance the gain compression of the transistor (the effective g_m will now tend to zero as the voltage becomes large). The optional diode

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

is shown as dotted lines in Figure 5.3a and the gate voltage will limit at around 0.8 V for a silicon diode.

Eq. (5.13) also tells us something about the stability of an oscillator. During startup, the transconductance g_m will steadily fall as the oscillator amplitude settles into its equilibrium state. As can be seen from Eq. (5.13), this will also mean that the oscillation frequency ω will take time to settle into its equilibrium value ω_0, i.e. there will be some initial frequency drift. To this, however, we must add the possibility that there will be additional changes to g_m as a result of the transistor warming up after being switched on. The temperature can take some time to settle down to an equilibrium state and so frequency can drift during this period. What will be noted from Eq. (5.13) is that the larger the value of Q, the smaller the effect of these variations upon frequency. As a consequence it is important to seek the highest possible value of Q in an oscillator circuit. So far we have only considered the inductor resistance to contribute to the Q of the circuit and this is indeed the most problematic resistive component in FET oscillators. However, other resistances can contribute to Q and, in particular, the resistance of any load at the output of the oscillator (the FET source). To reduce the effect of a load, it is prudent to use a high-input impedance buffer amplifier between the oscillator and the load (a common-drain amplifier for example).

Another way of looking at oscillation is through the concept of negative resistance. In Chapter 1, we saw that a tuned circuit will ring at its resonant frequency when excited by an impulse. The oscillations, however, will steadily decay due to the resistance in the circuit. To prevent this decay, we need to introduce a negative resistance that cancels out the inherent resistance in the circuit. How do we create this negative resistance? Consider the circuit shown in Figure 5.4a when analysed through the model in Figure 5.4b. We assume oscillations at frequency ω and so represent the currents and voltages in terms of their complex amplitudes. The voltage at the device input is

$$V = \frac{I}{\mathrm{j}\omega C_1} + \left(I + g_m \frac{I}{\mathrm{j}\omega C_1}\right)\frac{1}{\mathrm{j}\omega C_2} \tag{5.15}$$

so that the impedance looking into the device is given by

$$Z_i = \frac{V}{I} = -\frac{g_m}{C_1 C_2 \omega^2} + \frac{C_1 + C_2}{\mathrm{j}\omega C_1 C_2}. \tag{5.16}$$

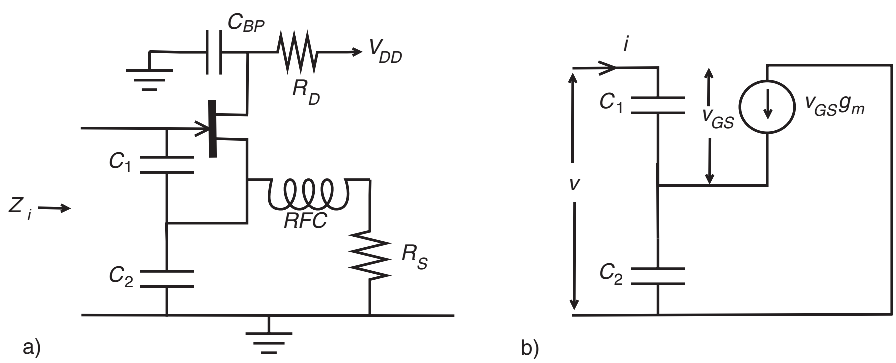


Fig. 5.4 Negative-resistance device and its analysis.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

The impedance is the series combination of resistance

$$R_i = -\frac{g_m}{C_1 C_2 \omega^2} \tag{5.17}$$

and the capacitance $C_i = C_1C_2/(C_1 + C_2)$. If we now place an inductor L across this impedance, the circuit can ring at the resonant frequency of C_i and L. In the real world, the inductor will contain some resistance that can be represented by a resistor R in series with the inductor ($R = Q_U \omega L$ where Q_U is the unloaded Q of the inductor) and this will moderate the growth of oscillations. However, the oscillations will still grow provided that, at the resonant frequency of L and C_i, condition $R + R_i < 0$ is satisfied. The oscillations will continue to grow in amplitude until a point is reached where the nonlinearity in the FET has reduced the transconductance to a point where $R + R_i = 0$. Note that, by adding an inductance L across the input to circuit Figure 5.4a, we have turned it into the Colpitts oscillator of Figure 5.3a, i.e. we have a negative resistance explanation of this oscillator.

The phenomenon of negative resistance can also shed further light on the regenerative amplifier that was discussed at the start of this section. We have already seen that positive feedback in an amplifier can generate a large gain. In addition, we have just seen that feedback can generate negative resistance that can cancel the inherent resistance in a tuned circuit at the input to an amplifier, i.e. it can greatly increase the Q of this tuned circuit and hence its selectivity. This combination of beneficial properties made it possible to build the cheap and effective radios that were the mainstay of the consumer radio industry well into the 1930s. Figure 5.5 shows a modern form of regenerative amplifier that is based on FETs. In this configuration, the lower transistor provides the negative feedback (regeneration) and the upper transistor controls the current through this device (i.e. the level of regeneration). The amplifier of Figure 5.5 can be made to perform the additional function of demodulation by setting the bias at the source so that there is strong asymmetry in response to the upswings and downswings of the RF voltage at the gate. To complete the demodulation, however, there will need to be a low-pass filter at the output in order to remove the RF frequencies. As a consequence, it is possible

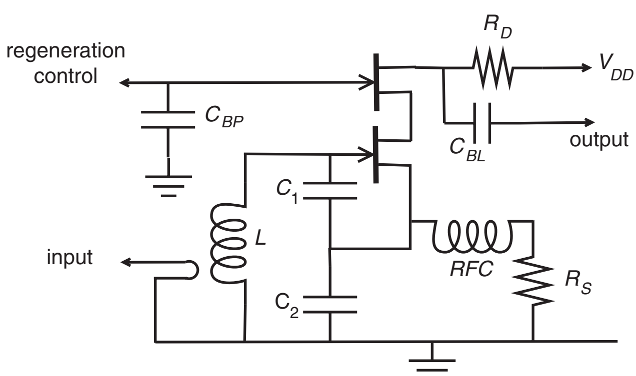


Fig. 5.5 Regenerative amplifier.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

to make the amplifier of Figure 5.5 into a simple receiver that is both highly selective and highly sensitive.

In 1922, Armstrong took his concept of regeneration one step further and invented what is known as the super-regenerative amplifier. In the regenerative amplifier of Figure 5.5, we now replace a constant bias at the gate of the upper transistor by an oscillating voltage (i.e. $V_b + V_q \sin(\omega_q t)$). The constant part of the bias voltage V_b is set such that, on the upswing, the voltage at the gate is high enough to set the lower transistor into oscillation and, on the downswing, is low enough to quench the oscillation. A signal that enters the tuned circuit on the upswing will now set the lower transistor into oscillation, the amplitude of which will grow exponentially until quenched by the downswing. It is clear that the level of the signal at the output will depend on the period of oscillation. This should be as long as possible to get the greatest level of output signal, but not so long that it interferes with any modulation of the RF signal (the *quench frequency* is usually set at several times the highest modulating frequency). Phenomenal gains can be attained with a super-regenerative amplifier, but at the cost of an output that is additionally modulated by the quench frequency. The additional modulation, however, is usually not a problem since it can be filtered out after demodulation.

5.2 Mixers

Thus far we have, almost exclusively, considered rectification as the means of extracting the modulating signal from an amplitude-modulated radio signal. In rectification we extinguish the negative swings of the signal and then smooth out (filter) the remaining signal to leave the modulating signal (see Figure 2.16 of Chapter 2). For amplitude modulation (AM), the signal has the form

$$v_{\mathrm{RF}} = V_{\mathrm{RF}}(1 + a(t))\cos(\omega_c t + \phi), \tag{5.18}$$

where ω_c is the carrier frequency, ϕ is a phase offset and $a(t)$ is the modulating signal (we assume $|a(t)| < 1$). If we multiply v_{RF} by the sinusoidal *local oscillator* (LO) signal $v_{\mathrm{LO}} = \cos(\omega_c t)$, we obtain

$$\begin{aligned} v_{\mathrm{RF}} v_{LO} &= V_{\mathrm{RF}}(1 + a(t))\cos(\omega_c t + \phi)\cos(\omega_c t) \\ &= \frac{V_{\mathrm{RF}}}{2}(1 + a(t))(\cos\phi + \cos(2\omega_c t + \phi)), \end{aligned} \tag{5.19}$$

where we have used the trigonometric identity $\cos A \cos B = \frac{1}{2}(\cos(A-B) + \cos(A+B))$. We could then filter out the high-frequency term and be left with a term whose time varying component is proportional to the modulating signal, i.e.

$$v_{out} = \frac{V_{\mathrm{RF}}}{2}(1 + a(t))\cos\phi. \tag{5.20}$$

An AM demodulator that uses a mixer in this fashion is known as a *synchronous demodulator* and the receiver of Figure 5.6 is based upon this concept. Such designs are known as *direct conversion*, or *homodyne*, receivers.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

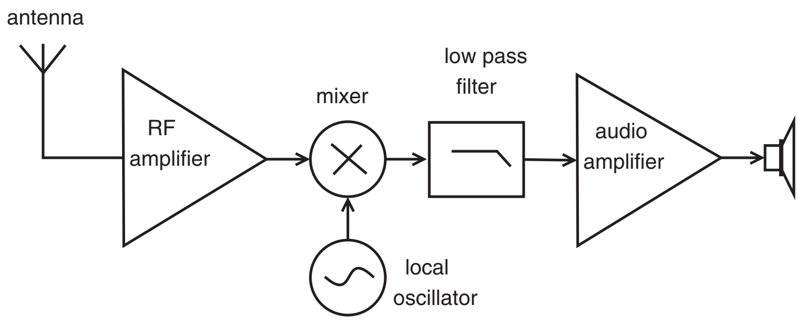


Fig. 5.6 Direct conversion receiver.

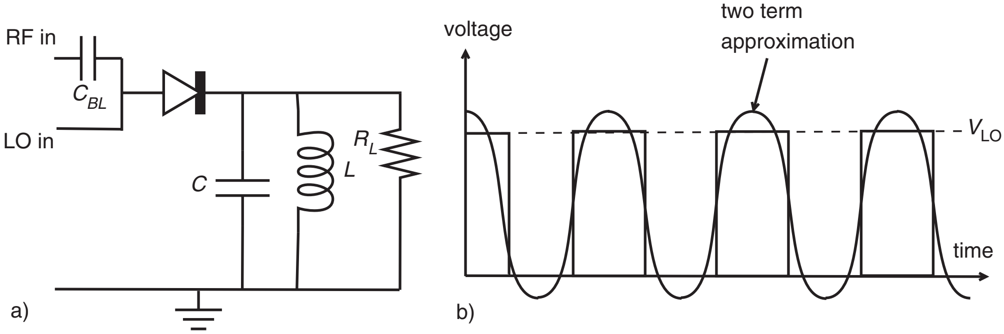


Fig. 5.7 Diode mixer and local oscillator signal.

A device that can achieve the product of two signals is known as a mixer. As we have seen in the previous chapter, a device that contains a square term in its nonlinear behaviour will contain such a product term in its output. Further, all the active devices we have studied thus far have such a term and so could be used as mixers. We will first consider the simplest mixer of all, the single-diode mixer (shown in Figure 5.7a). Diode mixers are usually operated such that the diode is hard on or hard off. This is achieved through a local oscillator voltage that is a sequence of pulses ($v_{LO} = V_{LO}(1 + \mathrm{sgn}(\cos(\omega_{LO}t)))/2$) that has an amplitude V_{LO} that is greater than the voltage that switches the diode on (about 0.7 volts for silicon and 0.3 volts for germanium). This will effectively produce a voltage across the load that is a series of pulses multiplied by the incoming RF voltage ($v_{RF} = V_{RF}\cos(\omega_{RF}t)$) when offset by V_{LO}. The sequence of pulses can be represented by a Fourier series, i.e.

$$v_{LO} = V_{LO}\,(1/2 + 2\cos(\omega_{LO}t)/\pi + 2\cos(3\omega_{LO}t)/3\pi + 2\cos(5\omega_{LO}t)/5\pi + \cdots).$$

As a first approximation we take the first two terms of the Fourier series, i.e. $v_{LO} \approx V_{LO}\,(1/2 + 2\cos(\omega_{LO}t)/\pi)$ (see Figure 5.7b). (In practice, we often make this approximation by using sinusoidal oscillations rather than pulses.) The voltage across the load will now be given by

$$v_{IF} = V_{LO}\left(\frac{1}{2} + \frac{2}{\pi}\cos(\omega_{LO}t)\right)(V_{LO} + V_{RF}\cos(\omega_{RF}t))$$

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

$$
\begin{aligned}
&= V_{\mathrm{LO}}^2\left(\frac{1}{2}+\frac{2}{\pi}\cos(\omega_{\mathrm{LO}}t)\right)\\
&\quad +V_{\mathrm{RF}}V_{\mathrm{LO}}\left(\frac{1}{2}\cos(\omega_{\mathrm{RF}}t)+\frac{1}{\pi}\cos((\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}})t)\right.\\
&\qquad \left.+\frac{1}{\pi}\cos((\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}})t)\right).
\end{aligned}
\tag{5.21}
$$

The major problem with the mixer is that the output contains both the RF signal, LO signal and signals at the frequencies $\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}}$ and $\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}}$. Usually, only one of the product frequencies is required and so all other frequencies are removed by making the parallel tuned circuit resonant at the required frequency. Furthermore, mixer products that were ignored in approximating the LO pulses are also filtered out by this process.

The square-law behaviour of the FET characteristic offers another opportunity for mixing and Figure 5.8 shows an example of a single FET mixer. Let $v_{\mathrm{RF}}=V_{\mathrm{RF}}\cos(\omega_{\mathrm{RF}}t)$ be the incoming RF signal at the FET gate and $v_{\mathrm{LO}}=V_{\mathrm{LO}}\cos(\omega_{\mathrm{LO}}t)$ be the local oscillator signal at the source. If the transistor has a bias current I_D at the drain, then the total current at the drain will be I_D+i_D where i_D is the RF current at the drain. From the characteristic equation of the FET,

$$
\begin{aligned}
I_D+i_D &= K\left(v_{\mathrm{RF}}-v_{\mathrm{LO}}-I_DR_S-V_T\right)^2\\
&= K\left(v_{\mathrm{RF}}^2-2v_{\mathrm{RF}}v_{\mathrm{LO}}+v_{\mathrm{LO}}^2-2(v_{\mathrm{RF}}-v_{\mathrm{LO}})(I_DR_S+V_T)\right.\\
&\qquad \left.+(I_DR_S+V_T)^2\right).
\end{aligned}
\tag{5.22}
$$

The tuned circuit in the drain of the FET is chosen to be resonant at either $\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}}$ or $\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}}$ and so the $v_{\mathrm{RF}}v_{\mathrm{LO}}$ term in the above expansion is the only term of relevance. Since $v_{\mathrm{RF}}v_{\mathrm{LO}} = V_{\mathrm{RF}}V_{\mathrm{LO}}\left(\cos((\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}})t)/2+\cos((\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}})t)/2\right)$, the current through the load R_L will be $-KV_{\mathrm{RF}}V_{\mathrm{LO}}\cos((\omega_{\mathrm{RF}}-\omega_{\mathrm{LO}})t)$ or $-KV_{\mathrm{RF}}V_{\mathrm{LO}}\cos((\omega_{\mathrm{RF}}+\omega_{\mathrm{LO}})t)$, depending on the resonant frequency of the drain circuit.

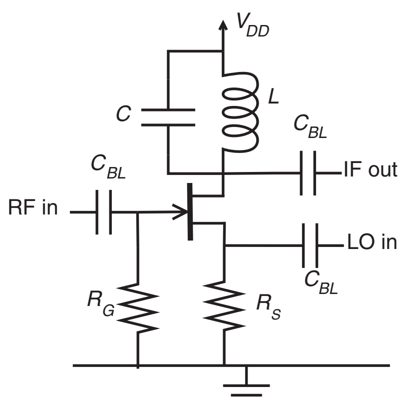


Fig. 5.8 Simple square-law FET mixer.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

The major problem with the mixers that we have discussed thus far is that their output contains all the input frequencies, plus the product frequencies. When only a fixed product frequency is required, the unwanted signals can be removed by suitable filtering, as shown in Figures 5.7 and 5.8. When broadband operation is required, we need to consider other means for eliminating the unwanted signals. A solution is found in what is known as a *balanced mixer*, an example of which is shown in Figure 5.9a. Once again, the local oscillator signal is a sequence of pulses that, in this case, swing between V_{LO} and $-V_{\mathrm{LO}}$ (i.e. $v_{\mathrm{LO}} = V_{\mathrm{LO}}\mathrm{sgn}(\cos(\omega_{\mathrm{LO}}t))$. The voltage amplitude V_{LO} is made large enough so that, at any one time, one side of the circuit will be completely on and the other side completely off. As a consequence, the voltage across the output terminals will be given by

$$v_{\mathrm{IF}} = (I_{bias} + v_{\mathrm{RF}}g_m)R_D\mathrm{sgn}(\cos(\omega_{\mathrm{LO}}t)), \tag{5.23}$$

where I_{bias} is the current that flows as a result of the bias that is applied to gate of the lower FET. As in the case of the diode mixer, we approximate the sequence of pulses by a Fourier series, i.e. $\mathrm{sgn}(\cos(\omega_{\mathrm{LO}}t)) = 4\cos(\omega_{\mathrm{LO}}t)/\pi + 4\cos(3\omega_{\mathrm{LO}}t)/3\pi + 4\cos(5\omega_{\mathrm{LO}}t)/5\pi + \cdots$. Then, truncating this series at the first term, and using the identity $\cos A \cos B = \frac{1}{2}(\cos(A-B) + \cos(A+B))$, we obtain

$$\begin{aligned} v_{IF} &\approx (I_{bias} + V_{\mathrm{RF}}g_m\cos(\omega_{\mathrm{RF}}t))R_D\frac{4}{\pi}\cos(\omega_{\mathrm{LO}}t)) \\ &= I_{bias}R_D\frac{4}{\pi}\cos(\omega_{\mathrm{LO}}t) + \frac{2}{\pi}V_{\mathrm{RF}}g_mR_D\,(\cos((\omega_{\mathrm{RF}} - \omega_{\mathrm{LO}})t) \\ &\quad + \cos((\omega_{\mathrm{RF}} + \omega_{\mathrm{LO}})t))\,. \end{aligned} \tag{5.24}$$

The balance between the two sides of the circuit has eliminated RF components to leave the product terms and the LO signal. To eliminate the LO signal, we can use what is known as a *double balanced mixer*. This consists of two single balanced mixers that are connected in such a way that the unwanted LO signal component cancels itself out.

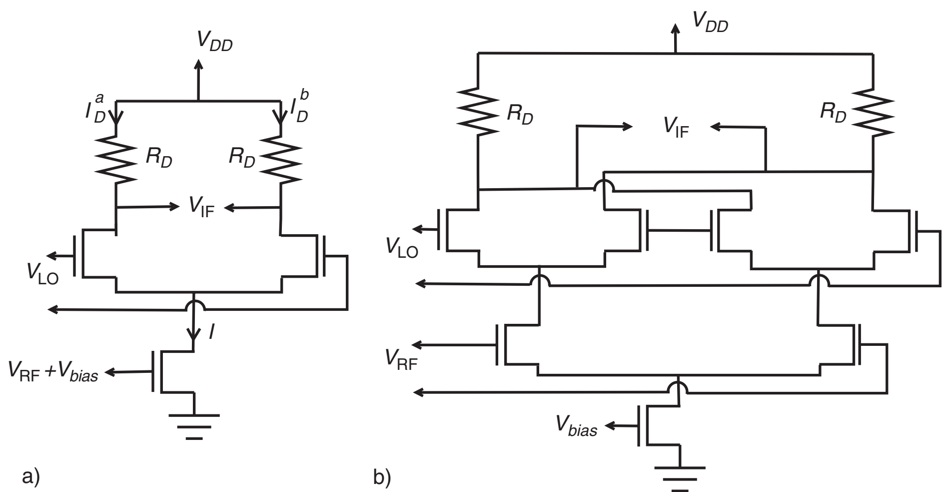


Fig. 5.9 Single and double balanced FET mixers.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

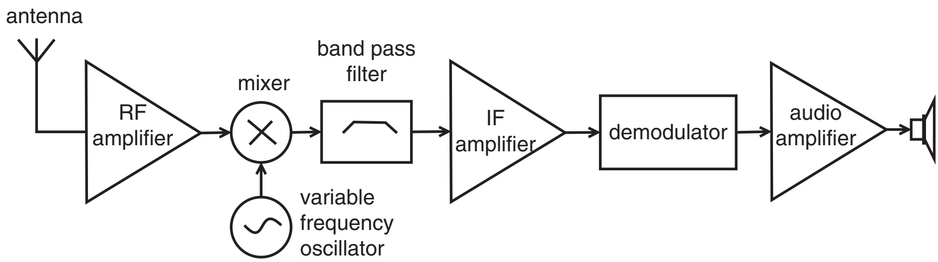


Fig. 5.10 Superheterodyne receiver.

Figure 5.9 shows an example of such a mixer, an architecture that is commonly known as a *Gilbert cell*. The drain resistors can be replaced by transistors, as in the case of the RF amplifier of Figure 4.30b, and so the mixer can be realised completely in terms of transistors. This has important implications. If we take a wafer of silicon, transistors can be constructed on this wafer by infusing impurities into the surface and adding layers of insulation, together with conducting tracks. Complex circuits can then be constructed by connecting these transistors by means of conducting tracks. For example, Gilbert cells and the amplifiers, of Figure 4.30b, can be combined to perform complex radio function on a single silicon wafer. What we have is an integrated circuit (IC) and such circuits have become an essential feature of modern RF electronics.

As we mentioned in Chapter 4, with reference to the TRF receiver, a radio that is designed to receive multiple frequencies needs to be able to simultaneously change several tuned circuits when the frequency is changed. As radio became more popular, the airwaves became more crowded and selectivity became a big issue. Receivers required more and more tuned elements to satisfy the increased selectivity requirements and the cost started to become prohibitive. To solve this problem, Armstrong came up with another of his brilliant ideas, the *superheterodyne receiver*. Consider the receiver architecture shown in Figure 5.10. A signal comes in through the antenna (at frequency ω_{RF}) and is amplified before being fed into a mixer. Here the signal is mixed with an LO signal (at frequency ω_{LO}) from a *variable-frequency oscillator* (VFO). At the output of the mixer, there is a fixed-frequency filter that chooses one of the product signals (either $\omega_{RF} - \omega_{LO}$ or $\omega_{RF} + \omega_{LO}$) and only has a bandwidth that is just wide enough to accommodate the modulation. This *intermediate-frequency* (IF) output is then amplified before being demodulated. Incoming signals are chosen through the correct choice of VFO frequency. The important thing to note is that it is possible to cheaply manufacture highly selective fixed-frequency band-pass filters, unlike the variable variety. The concept of the superheterodyne revolutionised the radio receiver and made it possible to build inexpensive receivers that are highly selective. The superheterodyne receiver, invented by Armstrong in 1918, is still the mainstay of receiver design.

5.3 Modulation and Demodulation

The simplest AM transmitter will have the topology shown in Figure 5.11a. The baseband signal $v_{AF} = 1 + a(t)$, where $a(t)$ is usually an audio signal, is mixed with a carrier

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

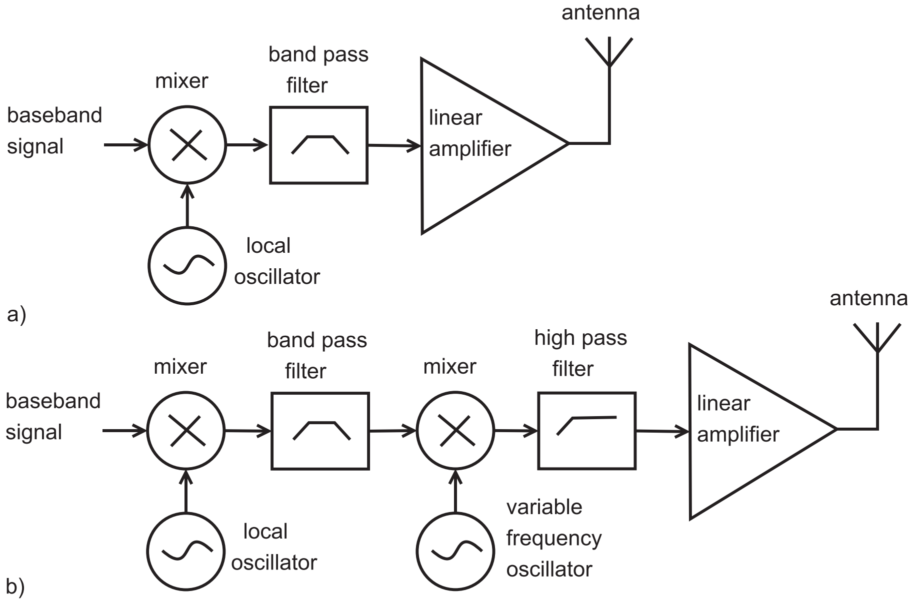


Fig. 5.11 Transmitter topologies.

signal $v_{\mathrm{LO}} = V_{\mathrm{LO}} \cos(\omega_c t)$ at frequency ω_c. The output of the mixer is the signal $v_{\mathrm{RF}} = v_{AF} V_{\mathrm{LO}} \cos(\omega_c t)$ which is then filtered and amplified. A filter is unnecessary for an ideal mixer, but even the best practical mixers will produce undesirable products that will need to be removed. To amplify an AM modulated signal up to the level of power required for transmission, the amplifier will need to be highly linear in order to avoid corrupting the modulation. Consider now the situation where the baseband signal is sinusoidal, i.e. $a(t) = a_0 \cos(\omega_{AF} t)$, then

$$\begin{aligned} v_{\mathrm{RF}} &= V_{\mathrm{LO}}(1 + a_0 \cos(\omega_{AF})) \cos(\omega_c t) \\ &= V_{\mathrm{LO}} \Big(\cos(\omega_c t) \\ &\qquad + \frac{a_0}{2} \cos((\omega_c + \omega_{AF})t) + \frac{a_0}{2} \cos((\omega_c - \omega_{AF})t) \Big). \end{aligned} \tag{5.25}$$

It will be noted that there is a signal at frequency ω_c that contains no information about the modulation (the *carrier signal*), a signal at frequency $\omega_c + \omega_{AF}$ and a signal at the mirror frequency $\omega_c - \omega_{AF}$. In general, the modulating signal will be far more complex (i.e. it will be the sum of sinusoidal signals at a variety of frequencies), but the structure of the modulated RF will be much the same. There will be a carrier at frequency ω_c and two sidebands. The first sideband will consist of a signal with spectrum that is the spectrum of $a(t)$ when translated upwards in frequency by ω_c. The second sideband will consist of a signal with spectrum that is the mirror spectrum of $a(t)$ when translated upwards in frequency by ω_c (see Figure 5.12). The signal below the carrier frequency is known as the *lower sideband* and that above as the *upper sideband*. Importantly, it will be seen that the carrier carries no information about the modulation and the sidebands carry exactly the same information. From an information viewpoint, it is senseless to

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

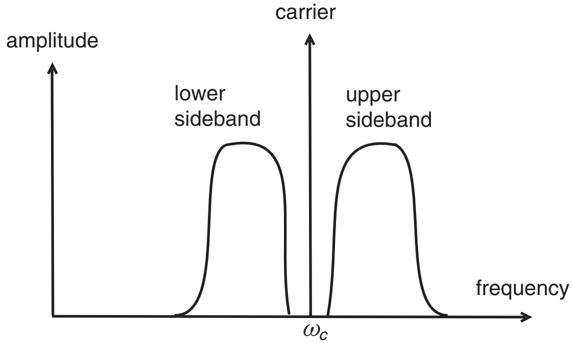


Fig. 5.12 Sideband structure in an AM signal.

transmit anything more than one of the sidebands. This brings us to the concept of single-sideband (*SSB*) modulation in which we merely transmit one of the sidebands (either upper or lower). This mode of operation saves both bandwidth (it is half that of AM) and power. The simplest way to generate SSB is with the transmitter of Figure 5.11a. The band-pass filter is now chosen such that it removes the carrier and the unwanted sideband. Obviously, such a transmitter will require a band-pass filter with very high selectivity and so the transmitter of Figure 5.11a will only be practical for a single frequency. Figure 5.11b, however, shows a transmitter that is suitable for operation over a range of frequencies. This is the transmitter equivalent of the superheterodyne receiver. It consists of a low power SSB transmitter that produces an SSB modulated output at a frequency ω_{IF}. This is then mixed with the output of a VFO at frequency ω_{VFO} to form two signals, one at frequency $\omega_{\mathrm{IF}} + \omega_{\mathrm{VFO}}$ and one at frequency $\omega_{\mathrm{IF}} - \omega_{\mathrm{VFO}}$. The mixer is then followed by a filter that chooses the correct product and then a linear power amplifier brings the signal up to a suitable power level.

Single sideband can be demodulated using synchronous demodulation but, unfortunately, it will respond to both sidebands. Consequently, if the receiver is tuned into a desired sideband, there could be an unwanted signal within the other sideband and this will also be demodulated, hence interfering with the desired signal. A solution is to use the superheterodyne receiver of Figure 5.10 with a demodulator of the synchronous variety. The band-pass filter, however, is now chosen narrow enough to only allow through a single sideband and so there will be no competing signal at the synchronous detector.

The above approach to the generation and demodulation of SSB relies upon the use of highly selective filters, but there is another approach. Figure 5.13a shows what is known as a *phasing SSB generator*. In essence, this circuit consists of two mixer-based AM generators, but with one having both its inputs advanced by 90°. If we consider the modulating signal to be sinusoidal, i.e. $v_B(t) = V_B \cos(\omega_B t)$, the sum of the generator outputs will be

$$\begin{aligned} v_{SSB} &= V_{\mathrm{LO}} V_B \cos(\omega_c t)\cos(\omega_B t) \\ &\quad + V_{\mathrm{LO}} V_B \cos\left(\omega_c t + \frac{\pi}{2}\right)\cos\left(\omega_B t + \frac{\pi}{2}\right) \end{aligned}$$

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

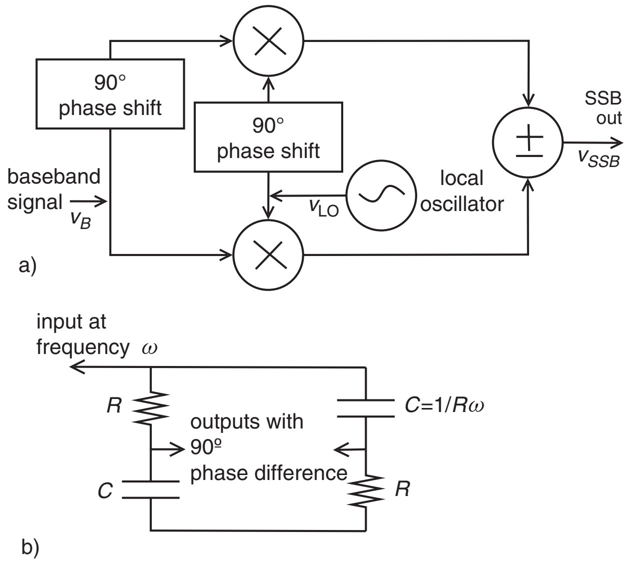


Fig. 5.13 The phasing method for generating SSB.

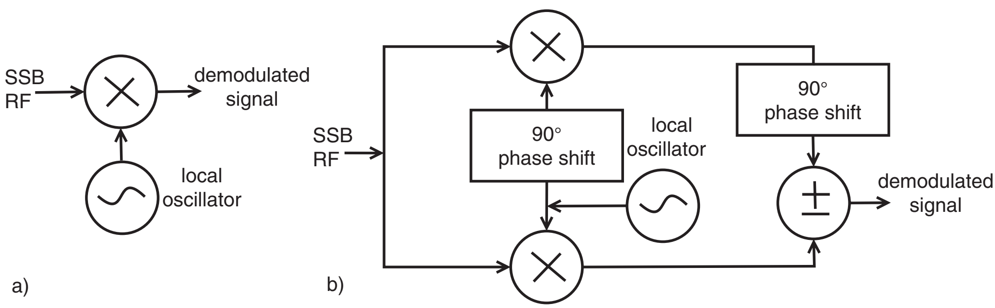


Fig. 5.14 Methods for demodulating SSB.

$$
\begin{aligned}
&= \frac{V_{\mathrm{LO}} V_B}{2}\left(\cos((\omega_c - \omega_B)t) + \cos((\omega_c + \omega_B)t)\right) \\
&\quad + \frac{V_{\mathrm{LO}} V_B}{2}\left(\cos((\omega_c - \omega_B)t) - \cos((\omega_c + \omega_B)t)\right) \\
&= V_{\mathrm{LO}} V_B \cos((\omega_c - \omega_B)t). \qquad (5.26)
\end{aligned}
$$

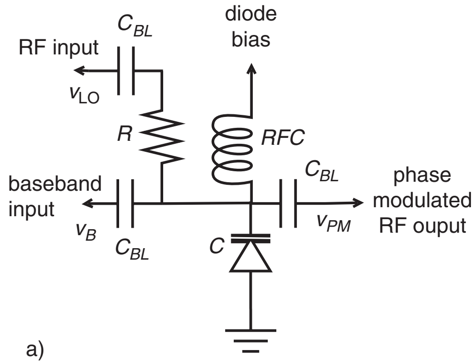
Consequently, we have produced the lower sideband alone without the need for filtering (we can produce the upper sideband by merely taking a difference of modulator outputs).

The phasing approach can also be used for the demodulation of SSB by means of the process shown in Figure 5.14b. The demodulator consists of two synchronous detectors, one having its oscillator and output advanced in phase by 90°. If we sum the outputs, the upper sideband will be demodulated and the lower sideband eliminated (if take the difference between the outputs the lower sideband will be demodulated). The major

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

practical problem with the phasing approach comes from the production of the 90° phase change. For the local oscillator, this can be achieved through the circuit of Figure 5.13b. In the case of the baseband signal, however, its bandwidth is often too wide for such a circuit and a polyphase network must be used. In modern communications, the signal is often modulated, and/or demodulated, within a computer. Within the digital domain, it is easy to produce the 90° phase shift over a broad range of frequencies and so, increasingly, the *in-phase* and *quadrature* (90° out-of-phase) baseband signals are produced and/or processed within the digital domain.

Thus far, we have concentrated on AM and its derivative modulation SSB, but we now turn our attention to *angle modulation* (FM and PM). Phase modulation can be generated through the circuit shown in Figure 5.15a. This circuit includes a reverse-biased diode which acts as a voltage-controlled reactance. The reverse-biased PN junction of a diode acts as a capacitor across its depletion region and its capacitance varies with bias due to the fact the width of the depletion region varies with this bias voltage. As a consequence, the capacitance in the circuit will vary with the incoming baseband voltage ($C \approx C_0 + v_B C_1$). Let the RF input to the modulator have the form $v_{\mathrm{LO}} = \Re\{V_{\mathrm{LO}} \exp(\mathrm{j}\omega_{\mathrm{LO}} t)\}$ and the output have the form $v_{PM} = \Re\{V_{PM} \exp(\mathrm{j}\omega_{\mathrm{LO}} t)\}$. The circuit is essentially a baseband-controlled



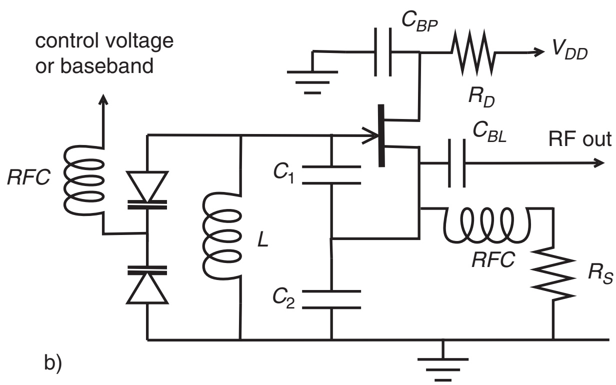


Fig. 5.15 A phase modulator and a voltage controlled oscillator for generating FM.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

voltage divider for which the complex amplitude of the output signal is given by

$$V_{PM} = V_{\mathrm{LO}} \frac{\frac{1}{\mathrm{j}\omega_{\mathrm{LO}}C}}{R + \frac{1}{\mathrm{j}\omega_{\mathrm{LO}}C}} = \frac{V_{\mathrm{LO}}}{\mathrm{j}\omega_{\mathrm{LO}}RC + 1}. \tag{5.27}$$

If $\omega_{\mathrm{LO}}RC \ll 1$, then $V_{PM} \approx V_{\mathrm{LO}}(1 - \mathrm{j}v_B\omega_{\mathrm{LO}}RC)$ and, on noting that $\exp x \approx 1 + x$ for small x, we find that

$$\begin{aligned} v_{PM} &= \Re\{V_{PM}\exp(\mathrm{j}\omega_{\mathrm{LO}}t)\} \\ &= \Re\{V_{\mathrm{LO}}\exp(\mathrm{j}(\omega_{\mathrm{LO}}t - v_B\omega_{\mathrm{LO}}RC_1 - \omega_{\mathrm{LO}}RC))\}, \end{aligned} \tag{5.28}$$

i.e. the baseband signal modulates the phase. A synchronous demodulator can be used to demodulate a PM signal since its output is proportional to the cosine of the phase difference between the input and local oscillator signals (see (5.20)). It will be noted, however, that this phase difference needs to be small in order for the demodulation to be linear.

FM can be generated by using what is known as a *voltage-controlled oscillator* (VCO). Figure 5.15b shows the circuit of the Colpitts VCO in which diodes have been added to allow the frequency to be varied through the variation of their bias, and hence capacitance. To generate FM, we simply use the baseband signal as the biasing voltage and then the frequency of the oscillator will vary in sympathy with the baseband signal. Demodulating FM can be done by means of what is known as slope detection. Consider the circuit in Figure 5.16a and for which the tuned circuit has a very high Q. The frequency response of the tuned circuit will look like that shown in Figure 5.16b. If the peak response of the tuned circuit is now offset from the centre of the FM frequencies, as the frequency of the FM signal varies the amplitude of its voltage will vary. Essentially, we have converted an FM signal into an AM signal. The AM signal will now be demodulated by the diode detector shown in Figure 5.16a.

Since we are dealing with angle modulation, the amplitude of the signal is irrelevant. Consequently, it is normal in an FM receiver to remove any possibility of AM interference by placing an amplitude-limiting amplifier prior to the demodulator. A simple example of such an amplifier is shown in Figure 5.17. In this amplifier, negative feedback is used to control the gain with the diode feedback circuit limiting the gain when a threshold output voltage is reached.

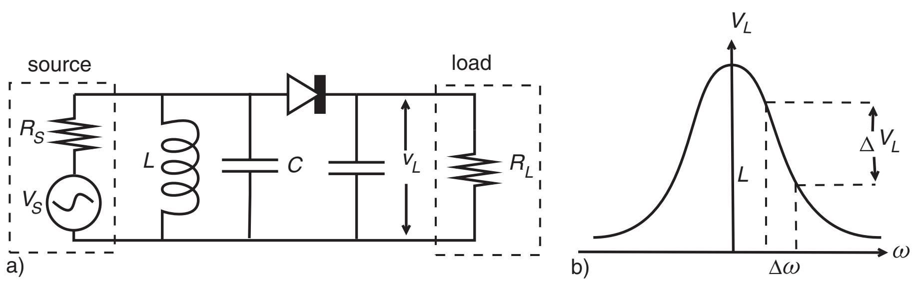


Fig. 5.16 Slope detector for demodulating FM.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

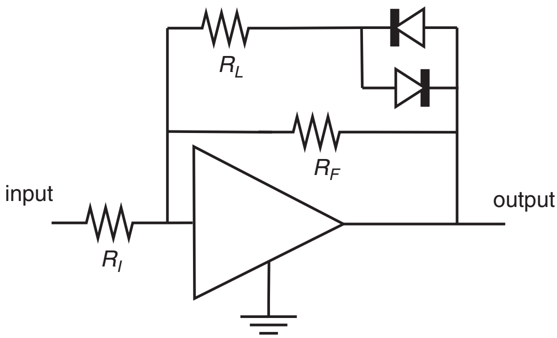


Fig. 5.17 A limiting amplifier.

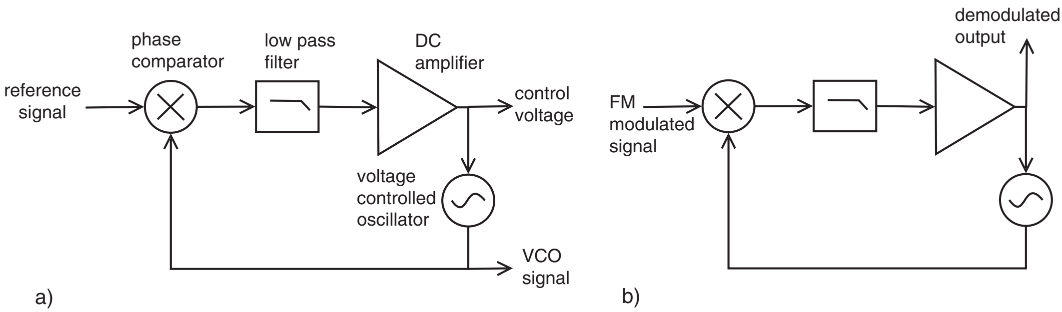


Fig. 5.18 Demodulating FM by means of a phase-locked loop.

FM signals can also be demodulated by means of what is known as a phase-locked loop, an example of which is shown in Figure 5.18a. In such a loop, the phase of the output of a VCO is compared with the phase of a reference signal (this operation is shown as a mixer since this is the simplest example of a phase-comparison device). After filtering and amplification, the phase difference is then passed to the VCO as a control voltage. Any small change in VCO phase is thus fed back as a control voltage and, by this means, the VCO is locked onto the reference signal. If the incoming reference signal is FM-modulated (see Figure 5.18b), the VCO will be locked onto this signal and the control voltage will be the demodulated FM.

FM radio was invented by Edwin Armstrong in 1933 in an effort to overcome some of the drawbacks of AM, and *static noise* in particular. This noise (originating from lightning and man-made sources such as electric switching) detracted from the quality of broadcasts and, in particular, that of music. Armstrong was able to demonstrate that, by increasing the bandwidth of the FM modulation, it was possible to obtain interference-free reception for a relatively low carrier-to-noise ratio. The signal-to-noise ratio after demodulation is given by

$$\mathrm{SNR} \approx 3 \left(\frac{B_{FM}}{2B} \right)^3 \mathrm{CNR}, \tag{5.29}$$

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

where B is the bandwidth of the baseband, B_{FM} is the bandwidth of the FM signal and CNR is the carrier-to-noise ratio in the bandwidth B_{FM}. Unfortunately, the above relation breaks down when CNR is less than 13 dB and the demodulated signal is rapidly overtaken by the noise at lower CNRs.

5.4 Oscillator Noise and Reciprocal Mixing

A particularly useful way of viewing an oscillator is as a filter of the random noise that excites the oscillations. We will consider the Colpitts oscillator of Section 5.1 from this viewpoint. The noise will be represented by a voltage source v_n at the gate of the amplifier and we will study the behaviour of the oscillator in terms of the model of Figure 5.19 (note that the mean square noise voltage per unit bandwidth is given by $\overline{v_n^2} = (F-1)kTR$, where F is the noise factor of the amplifier). We consider noise input at frequency ω and so represent the currents and voltages in terms of their complex amplitudes. The analysis proceeds as for the Colpitts oscillator, except that Eq. (5.5) is replaced by

$$V_{GS} = \frac{I_1}{\mathrm{j}\omega C_1} + V_n, \tag{5.30}$$

where V_n is the noise amplitude. From (5.3) and (5.30)

$$I_2 = I_1\left(1 + \frac{g_m}{\mathrm{j}\omega C_1}\right) + V_n g_m \tag{5.31}$$

and hence Eq. (5.7) will be replaced by

$$I_1\left(R + \mathrm{j}\omega L + \frac{1}{\mathrm{j}\omega C_1} + \frac{1}{\mathrm{j}\omega C_2} - \frac{g_m}{\omega^2 C_1 C_2}\right) + \frac{V_n g_m}{\mathrm{j}\omega C_2} = 0. \tag{5.32}$$

From (5.31), the complex amplitude V of the RF voltage at the output of the amplifier (i.e. the transistor source) is given by

$$V = \frac{I_2}{\mathrm{j}\omega C_2} = \frac{I_1}{\mathrm{j}\omega C_2}\left(1 + \frac{g_m}{\mathrm{j}\omega C_1}\right) + \frac{V_n g_m}{\mathrm{j}\omega C_2} \tag{5.33}$$

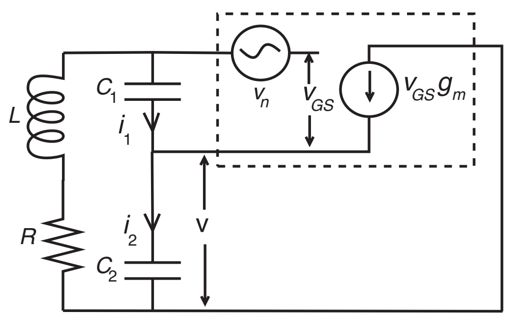


Fig. 5.19 Colpitts oscillator with noise source.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

and so, from (5.32) and (5.33), we obtain

$$V = \frac{V_n g_m}{j\omega C_2} \frac{R + j\omega L + \frac{1}{j\omega C_1}}{R + j\omega L + \frac{1}{j\omega C_1} + \frac{1}{j\omega C_2} - \frac{g_m}{\omega^2 C_1 C_2}}. \tag{5.34}$$

This can be rewritten as

$$V = \frac{V_n g_m}{j\omega C_2} \frac{1 + jQ\frac{\omega}{\omega_0}\left(1 - \frac{\omega_0^2}{\omega^2}\right) - \frac{1}{jR\omega C_2}}{1 - \frac{g_m}{R\omega^2 C_1 C_2} + jQ\frac{\omega}{\omega_0}\left(1 - \frac{\omega_0^2}{\omega^2}\right)}, \tag{5.35}$$

where $Q = \omega_0 L/R$ and $\omega_0 = \sqrt{(C_1 + C_2)/LC_1C_2}$. After the oscillator has settled into equilibrium, the parameter $\Delta = 1 - g_m/R\omega_0^2 C_1 C_2$ will be extremely small. However, Δ cannot be exactly zero since this would make V infinite. If we now consider frequencies close to ω_0, the denominator in (5.35) will dominate and so we can set the numerator to its value at $\omega = \omega_0$. Noting that $1 - \omega_0^2/\omega^2 \approx 2\delta\omega/\omega_0$, where $\delta\omega = \omega - \omega_0$, we then find that

$$V \approx \frac{V_n g_m}{j\omega_0 C_2 \Delta} \frac{1 - \frac{1}{jR\omega_0 C_2}}{1 + 2jQ\frac{\delta\omega}{\Delta\omega_0}} \tag{5.36}$$

It will be seen that the amplitude will fall off as $1/\delta\omega$ as we move away from resonance. Since the noise source will produce output on all frequencies, it is clear from (5.36) that the oscillator will produce a spread of frequencies and not a single frequency. Further, the spread in frequency will be of the order of $\Delta\omega_0/Q$.

The spread in frequency at the oscillator output can be regarded as a modulation of the desired signal at frequency ω_0 and, due to the random nature of the noise source, the modulation itself will be random. However, due to gain compression, the amplitude will remain fairly constant and the modulation will essentially consist of phase fluctuations (note that frequency is the rate of change of phase). For this reason, oscillator noise is often termed *phase noise*. From (5.36) it will be noted that the higher the value of Q, the narrower the spread in frequency. As a consequence, it can be seen that a high Q is desirable for low phase noise.

The spread of frequencies in a practical oscillator is obviously a major problem as it prevents us from transmitting a totally clean signal by adding some random modulation onto the signal. Further, at the receive end, the local oscillator can also add additional random modulation to the incoming signal. There is, however, an additional problem at the receiver. Consider a superheterodyne receiver with IF frequency f_{IF} that is tuned to a desired signal S1 on frequency f_1. If there is also a strong undesired signal S2 on another frequency f_2, then this signal will also be present at the IF if frequency $f_{LO} + f_2 - f_1$ is present within the spectrum of the LO. This phenomenon, illustrated in Figure 5.20, is known as *reciprocal mixing*.

From the above considerations, and those of Section 5.1, it is clear that a high-Q resonant circuit is essential for an oscillator to produce a clean and stable output. In a resonant circuit that is formed with capacitors and inductors, a major problem arises due to the fact that it is very difficult to produce inductors with self Q greater than

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

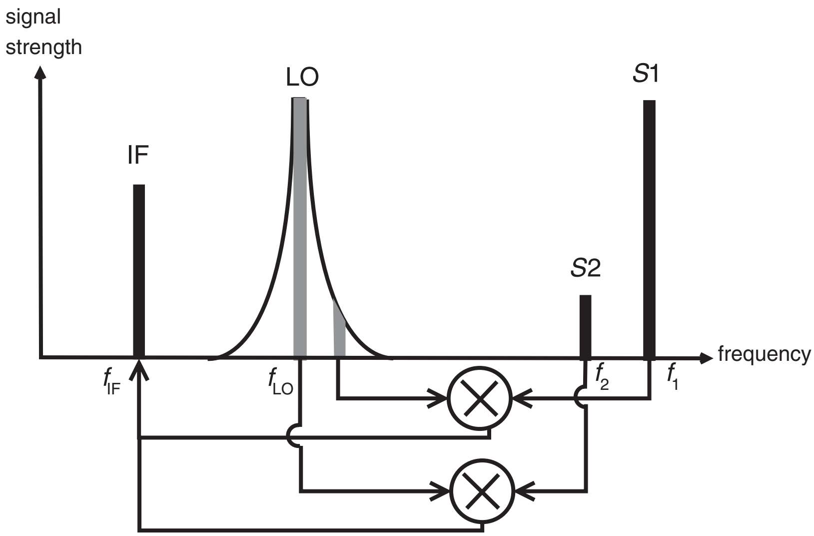


Fig. 5.20 Reciprocal mixing.

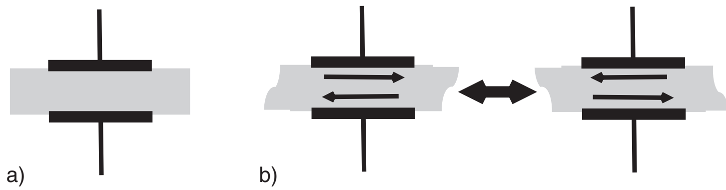


Fig. 5.21 Quartz crystal resonator and shear vibration mode.

about 200. As a consequence, such resonant circuits are often inadequate if we require a pure and stable sinusoid from the oscillator. A solution to this problem can be found in what is known as a *crystal resonator*. The mechanical vibrations of a crystalline solid can often exhibit a very high Q and this can be exploited to produce a very high-Q electrical resonator. This is possible through a phenomenon (discovered by Pierre Curie in 1880) known as the piezoelectric effect. Under this effect, deformations of some crystalline solids will induce an electric field (the deformations cause a realignment of the electric dipoles within the solid and this causes a field). The reverse is also true and an electric field across the same solid will induce a deformation (the fields cause a realignment of the electric dipoles within the solid). Consider a crystalline material that exhibits the piezoelectric effect (quartz is an example of such a material) when placed between two electrodes (see Figure 5.21). If a voltage is placed across the electrodes, this will cause a transverse deformation of the crystal. When the voltage is switched off, the crystal will vibrate at its mechanical resonant frequency ω_0 (this frequency depends on the dimensions and the elastic properties of the crystal). Shear waves will travel backwards and forwards between the upper and lower electrodes, the deformations being transverse to the propagation direction. (The distance between the electrodes will be an odd number of wavelengths, i.e. the crystal is capable of many vibration modes.) Through the piezoelectric effect, this will now mean that there is now an oscillating voltage across the electrodes, i.e. the quartz crystal will behave like a tuned circuit with resonant frequency ω_0. Importantly, this tuned circuit will have an extremely high Q (of the order of 10^7 for a quartz crystal). Figure 5.22a shows a model of the equivalent electrical resonator and Figure 5.22b the behaviour of the reactance of the circuit. It will be noted that the resonator has both series and parallel resonances. Quartz is the material often used for crystal resonators since it exhibits very little change in resonance properties with temperature.

A crystal oscillator circuit can be derived from one which uses inductors and capacitors by replacing the inductor with a crystal. Figure 5.23a shows a crystal oscillator that was derived from the Colpitts oscillator of Figure 5.3a. The fundamental mode of crystal vibration is usually only used for frequencies below 32 MHz, and above this the higher modes (overtones) tend to be used due to difficulty of making fundamental mode crystals for these frequencies. To make sure the circuit oscillates at the desired overtone, we modify the oscillator of Figure 5.23a to that of Figure 5.23b. By choosing capacitor C and inductor L to have a series resonance at the desired overtone, we then encourage oscillation at the desired overtone and block off oscillations at the fundamental and undesired overtones.

The problem with crystal oscillators is that it is difficult to vary the frequency. Some variation can be achieved in the above Colpitts oscillator by adding a variable capacitance in parallel with the crystal, but the achievable variation is small. Phase-lock loops, however, offer some other possibilities for producing stable oscillators. We first note that it is much easier to produce stable variable frequency oscillators (VFOs) at low frequencies. This can then be used to produce a stable oscillation at higher frequencies through the circuit shown in Figure 5.24a. The final signal (frequency f) is produced by a voltage-controlled oscillator (VCO), but a sample of this signal is converted down to a much lower frequency ($f - f_{XTAL}$) using a mixer and a stable crystal oscillator (frequency f_{XTAL}). The phase of this signal will then be compared with that from a stable VFO (frequency f_{VFO}) using a mixer followed by a low-pass filter. This will produce an

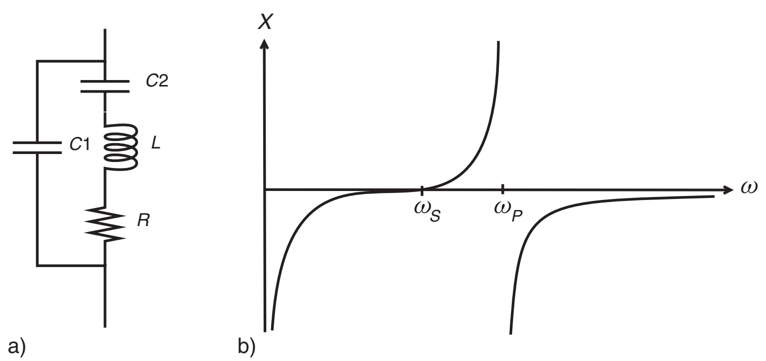


Fig. 5.22 Model of Quartz crystal resonator.

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

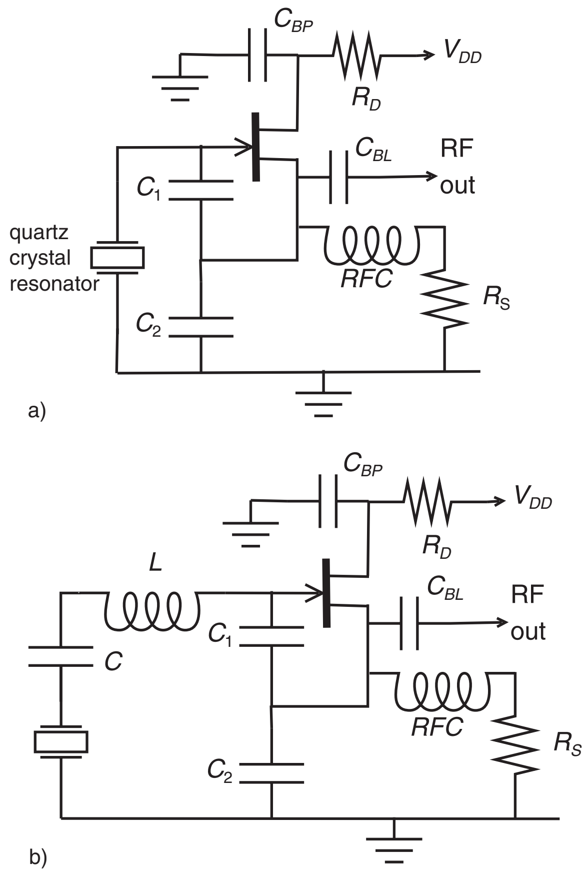


Fig. 5.23 Crystal controlled Colpitts oscillator and an overtone oscillator.

output that is a function of difference in phase, which, after amplification, produces a control voltage for the VCO. At the desired frequency, set by the stable low-frequency VFO, there will be a given phase difference that ensures the VCO runs at the correct frequency of $f_{\mathrm{VFO}} + f_{\mathrm{XTAL}}$. If the VCO drifts, there will be a change in phase difference that causes the VCO frequency to move back towards the desired frequency. In this way the VCO is locked onto the VFO.

Another possibility has come about through developments in digital electronics and, in particular, digital devices that act as programmable frequency dividers. By means of such a device, a phase-locked loop can be turned into a programmable frequency source. Figure 5.24b shows a suitable topology for such a *frequency synthesiser*. As before, the desired signal is generated by a voltage-controlled oscillator (VCO) and this is then divided down to the frequency of a reference oscillator by the programmable divider (the desired frequency needs to be a multiple of the reference frequency). If the VCO starts to drift from the desired frequency, there will be a phase difference between the frequency-divided VCO signal and the reference signal. This will result in a voltage at the

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

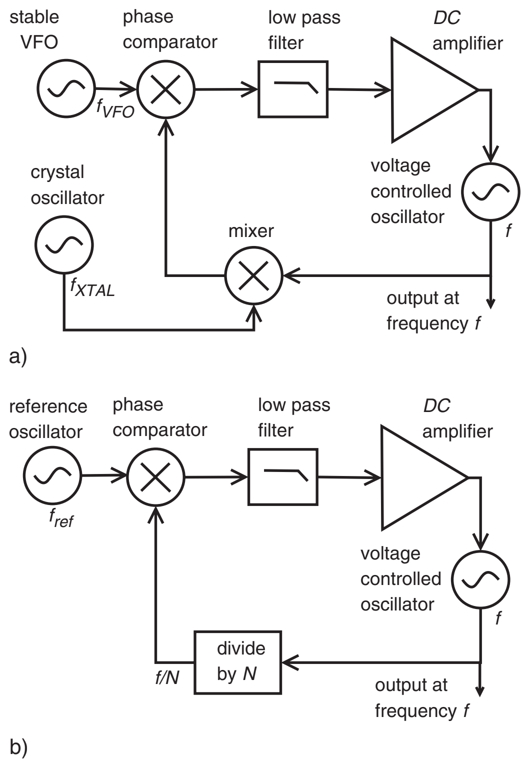


Fig. 5.24 Frequency stabilisation through phase-locked loops.

output of the phase detector and, after amplification, this will provide a control voltage for the VCO. This voltage will then provide the feedback that moves the frequency-divided signal back towards that of the reference frequency. In this way the VCO is locked onto the reference. Consequently, arbitrary frequencies that are a multiple of the reference frequency can be generated by choosing a suitable frequency division. In both circuits of Figure 5.24, a problem arises when the circuit is first switched and the VCO is not locked onto the reference oscillator. A simple solution to this problem is to add some positive feedback to the DC amplifier so that it oscillates at a low frequency until the correct control voltage for the VCO is found. Once lock is achieved, these oscillations will then cease.

5.5 Power Amplifiers

For effective communications, we need to produce RF signals at considerable power. However, the small-signal amplifiers of the previous chapter are no longer suitable. Figure 5.25 shows a typical topology of a single-transistor power-amplifier circuit (an NPN BJT device is shown, but the topology can apply equally well to a FET). To simplify our discussion, we will assume that the transistor amplifies in a linear manner. In general,

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

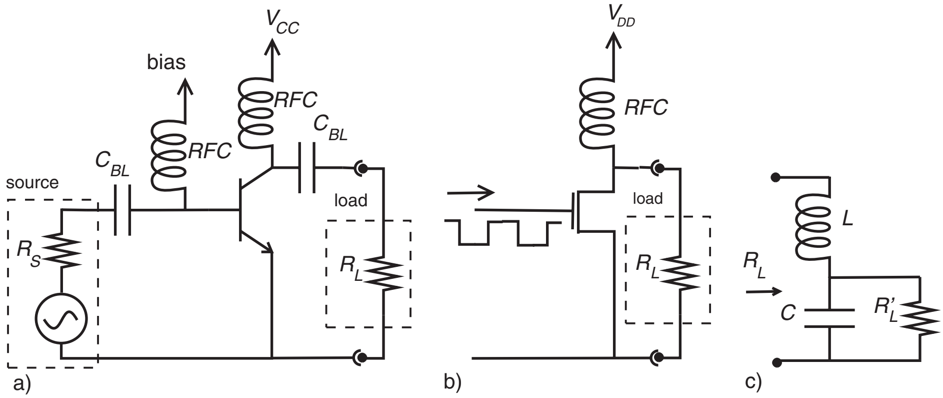


Fig. 5.25 Single-transistor amplifier topology and matching network.

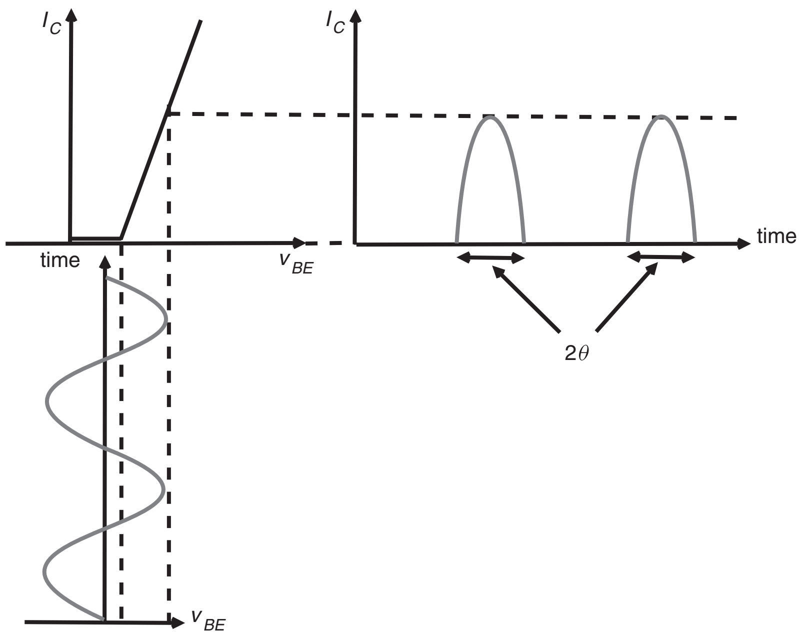


Fig. 5.26 Relation between input and output for a single transistor power amplifier.

the voltage at the base of the transistor is given by

$$\mathcal{V}_{\mathrm{BE}} = V_{\mathrm{bias}} + V_{\mathrm{in}} \sin(\omega t) \tag{5.37}$$

and the current at the collector can be represented by

$$\mathcal{I}_{\mathrm{CE}} = \max(I_{\mathrm{P}}\left(\sin(\omega t) - cos(\theta)\right), 0) \tag{5.38}$$

if we assume the amplifier behaves in a linear manner when it conducts. This is illustrated Figure 5.26 where the transistor is shown to conduct over an angle 2θ of each period of

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

the output. The collector emitter current is not necessarily sinusoidal but, since it is still periodic in nature, it can be represented as a Fourier series, i.e.

$$\mathcal{I}_{\mathrm{CE}} = I_0 + I_1 \sin(\omega t) + I_2 \sin(2\omega t) + I_3 \sin(3\omega t) + \cdots \tag{5.39}$$

The harmonic content (terms with coefficients I_2, I_3, etc.) are usually of no interest and are filtered out by the means of a low-pass filter. Consequently, after filtering, the output current will have the form

$$\mathcal{I}_{\mathrm{CE}} = I_{\mathrm{DC}} + I_{\mathrm{RF}} \sin(\omega t), \tag{5.40}$$

where we have, from Fourier theory, that $I_{DC} = I_P(\sin\theta - \theta\cos\theta)/\pi$ and $I_{\mathrm{RF}} = I_P(2\theta - \sin 2\theta)/2\pi$. We will assume that the voltage swing is V_{CC} (the maximum possible) and so $I_{\mathrm{RF}} = V_{CC}/R_L$. One of the major considerations with a power amplifier is its efficiency; this is the ratio of the average output power $P_{\mathrm{RF}} = I_{\mathrm{RF}}^2 R_L/2 = V_{CC} I_{\mathrm{RF}}/2$ to the DC input power $P_{DC} = I_{DC} V_{CC}$, i.e.

$$\text{efficiency} = \frac{P_{\mathrm{RF}}}{P_{DC}} = \frac{2\theta - \sin 2\theta}{4(\sin\theta - \theta\cos\theta)}. \tag{5.41}$$

If we set the transistor bias so that $\theta = \pi$, the transistor will conduct over the whole cycle and the output waveform will be sinusoidal. The amplifier is known as a *class A amplifier* and, from (5.41), we see that it has an efficiency of 50%. If we now set the bias so that $\theta = \pi/2$, the transistor will now only conduct for half a cycle. In this case, the amplifier is known as a *class B amplifier* and, from (5.41), we see that it has an efficiency of 78.5%. The downside is that the output is now highly distorted and needs filtering to remove unwanted harmonics (signals at multiples of the input frequency). We can get even higher efficiencies with amplifier that conducts for less than half a cycle, a *class C amplifier*. In theory, by reducing the conduction period to almost nothing, we can get to almost 100% efficiency. There is a catch, however, in that there will be a very large current during the transistor's on period. The maximum current during the on period will be

$$I_{\max} = \frac{2\pi V_{CC}}{R_L} \frac{1 - \cos\theta}{2\theta - \sin 2\theta} \tag{5.42}$$

and this will become infinite as θ tends to zero. All transistors will have a finite limit of current that they can sustain and so the achievable efficiency of a class C amplifier is limited by the transistor itself. Another problem with the class C amplifier is the high degree of distortion in the output and this precludes it being used in any application that requires linear amplification (SSB for example). For angle modulation (i.e. FM and PM), however, class C can provide extremely efficient power amplification.

The advent of high-speed power-switching transistors has made possible another avenue for highly efficient power amplification: the *class D amplifier* (see Figure 5.25b). The transistor is now driven by a series of pulses (these can be frequency-, phase- or width-modulated). In fact, this is not an amplifier but a high-speed switch. When the transistor is off there is obviously no power dissipated. Further, when on, there is no voltage drop across the transistor and hence no power is dissipated. Obviously, this is an ideal and a real device will dissipate some power. At switch-on the current will take time to rise and at switch-off it will take time to fall. Nevertheless, these amplifiers can

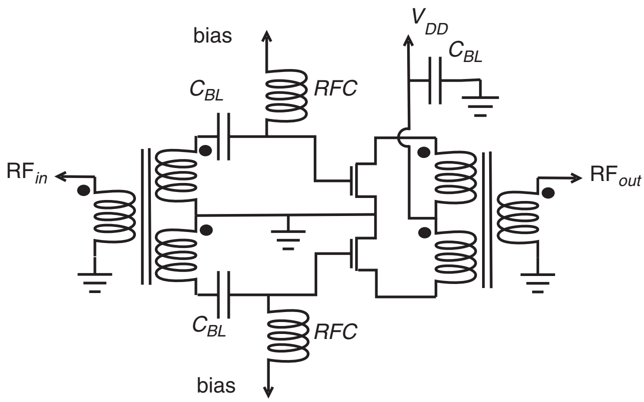


Fig. 5.27 Class B push–pull amplifier.

be extremely efficient and the necessary pulse drive is easily generated with modern digital techniques. As with the class C amplifier, the output will be rich in harmonics and therefore require filtering to remove unwanted harmonics.

The question now arises as to the nature of the load R_L. If the average power output of P_o is specified, we will need to choose R_L to be $V_{CC}^2/2P_o$. This, however, might not be convenient (the impedance of the antenna is often 50 Ω) and so it will be necessary to include a matching network in the amplifier. Figure 5.25c shows a simple L network that might be used. It will be noted that we have chosen the reactances to form a low-pass filter so that the network will also help to remove unwanted harmonics, besides matching the power amplifier to the load R_L'.

Both class A and B amplifiers retain the amplitude information of the original signal, but a class B amplifier will require filtering to remove its unwanted harmonics. This filtering, however, can be a problem if operation is required over a wide range of frequencies. Fortunately, a push–pull amplifier can help solve this problem (see Figure 5.27). Such an amplifier is essentially two class B amplifiers that are 90° out of phase. These generate two half sine waves that combine to form a full sine wave and so, in principle, avoid the problem of harmonics. However, as we have discussed in the previous chapter, a transistor will have a nonlinear characteristic and so the full sinusoid will itself be distorted, especially if we are dealing with large signals. Since a FET has a quadratic characteristic, the distortion will mainly consist of the harmonic at frequency 2ω when the input signal is at frequency ω. Fortunately, a push–pull amplifier has the property that it will cancel out this harmonic between the two sides on the amplifier. Consequently, the push–pull amplifier of Figure 5.27 will produce an output that is relatively free from harmonics.

5.6 Conclusion

In the current chapter we have discussed some of the developments that turned radio into a key technology of the twentieth century. Circuit components such as amplifiers, mixers

https://doi.org/10.1017/9781108684514.006 Published online by Cambridge University Press

and oscillators have made possible high-performance radios through radio architectures such as the superheterodyne. However, from the late twentieth century onwards, there have been major advances in computer technology that have now made it possible to build radios in computer software. By digitising a radio signal, it is now possible to carry out within a computer the signal processing that was hitherto carried out in the analogue domain. In the next chapter we will consider the use of digital techniques in radio and the practical realisation of software radio.