

This page intentionally left blank



TRANSCEIVER ARCHITECTURES

With the understanding developed in previous chapters of RF design and communication principles, we are now prepared to move down to the transceiver architecture level. The choice of an architecture is determined by not only the RF performance that it can provide but other parameters such as complexity, cost, power dissipation, and the number of external components. In the past ten years, it has become clear that high levels of integration improve the system performance along all of these axes. It has also become clear that architecture design and circuit design are inextricably linked, requiring iterations between the two. The outline of the chapter is shown below.

Heterodyne Receivers	Direct-Conversion Receivers	Image-Reject and Low-IF Receivers	Transmitter Architectures
■ Problem of Image	■ LO Leakage and Offsets	■ Hartley and Weaver Receivers	■ TX Baseband Processing
■ Mixing Spurs	■ Even-Order Nonlinearity	■ Low-IF Receivers	■ Direct-Conversion TX
■ Sliding-IF RX	■ I/Q Mismatch	■ Polyphase Filters	■ Heterodyne and Sliding-IF TX

4.1 GENERAL CONSIDERATIONS

The wireless communications environment is often called “hostile” to emphasize the severe constraints that it imposes on transceiver design. Perhaps the most important constraint originates from the limited channel bandwidth allocated to each user (e.g., 200 kHz in GSM). From Shannon’s theorem,¹ this translates to a limited rate of information, dictating the use of sophisticated baseband processing techniques such as coding, compression, and bandwidth-efficient modulation.

1. Shannon’s theorem states that the achievable data rate of a communication channel is equal to $B \log_2(1 + SNR)$, where B denotes the bandwidth and SNR the signal-to-noise ratio.

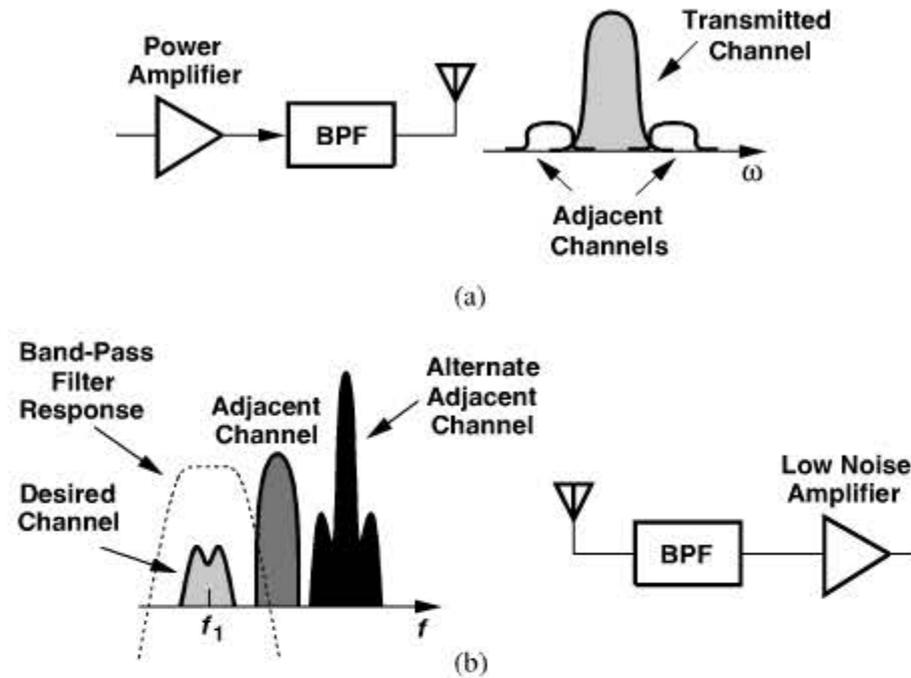


Figure 4.1 (a) Transmitter and (b) receiver front ends of a wireless system.

The narrow channel bandwidth also impacts the RF design of the transceiver. As depicted in Fig. 4.1, the transmitter must employ narrowband modulation and amplification to avoid leakage to adjacent channels, and the receiver must be able to process the desired channel while sufficiently rejecting strong in-band and out-of-band interferers.

The reader may recall from Chapter 2 that both nonlinearity and noise increase as we add more stages to a cascade. In particular, we recognized that the linearity of a receiver must be high enough to accommodate interferers without experiencing compression or significant intermodulation. The reader may then wonder if we can simply filter the interferers so as to relax the receiver linearity requirements. Unfortunately, two issues arise here. First, since an interferer may fall only one or two channels away from the desired signal (Fig. 4.2), the filter must provide a very high selectivity (i.e., a high Q). If the interferer level is, say, 50–60 dB above the desired signal level, the required value of Q reaches prohibitively high values, e.g., millions. Second, since a different carrier frequency may be allocated to the

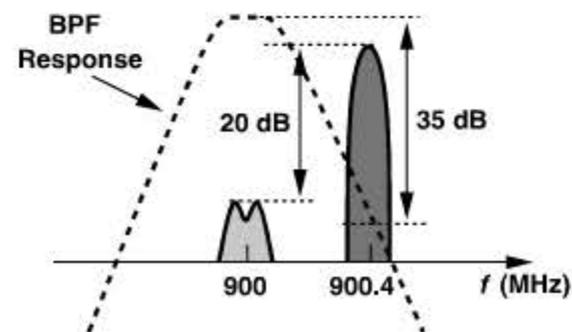


Figure 4.2 Hypothetical filter to suppress an interferer.

user at different times, such a filter would need a *variable*, yet precise, center frequency—a property very difficult to implement.

Example 4.1

A 900-MHz GSM receiver with 200-kHz channel spacing must tolerate an alternate adjacent channel blocker 20 dB higher than the desired signal. Calculate the Q of a second-order LC filter required to suppress this interferer by 35 dB.

Solution:

As shown in Fig. 4.2, the filter frequency response must provide an attenuation of 35 dB at 400 kHz away from center frequency of 900 MHz. For a second-order RLC tank, we write the impedance as

$$Z_T(s) = \frac{RLs}{RLCs^2 + Ls + R} \quad (4.1)$$

and assume a resonance frequency of $\omega_0 = 1/\sqrt{LC} = 2\pi(900 \text{ MHz})$. The magnitude squared of the impedance is thus given by

$$|Z_T(j\omega)|^2 = \frac{L^2\omega^2}{(1 - LC\omega^2)^2 + L^2\omega^2/R^2} \quad (4.2)$$

For an attenuation of 35 dB ($=56.2$) at 900.4 MHz, this quantity must be equal to $R^2/56.2^2$ (why?). Solving for $L^2\omega^2/R^2$, we obtain

$$\frac{L^2\omega^2}{R^2} = 2.504 \times 10^{-10} \quad (4.3)$$

Recall from Chapter 2 that $Q = R/(L\omega) = 63,200$.

Channel Selection and Band Selection The type of filtering speculated above is called “channel-selection filtering” to indicate that it “selects” the desired signal channel and “rejects” the interferers in the other channels. We make two key observations here: (1) all of the stages in the receiver chain that *precede* channel-selection filtering must be sufficiently linear to avoid compression or excessive intermodulation, and (2) since channel-selection filtering is extremely difficult at the input carrier frequency, it must be deferred to some other point along the chain where the center frequency of the desired channel is substantially *lower* and hence the required filter Q 's are more reasonable.²

Nonetheless, most receiver front ends do incorporate a “band-select” filter, which selects the entire *receive band* and rejects “out-of-band” interferers (Fig. 4.3), thereby suppressing components that may be generated by users that do not belong to the standard

2. The Q of a band-pass filter may be roughly defined as the center frequency divided by the -3 -dB bandwidth.

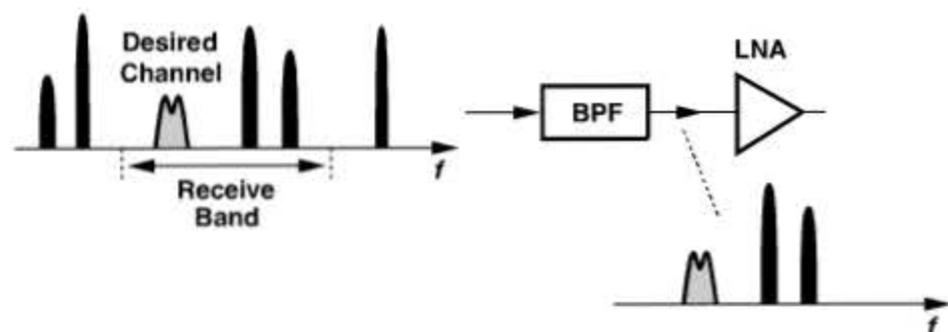


Figure 4.3 Band-selection filtering.

of interest. We thus distinguish between out-of-band interferers and “in-band interferers,” which are typically removed near the end of the receiver chain.

The front-end band-select (passive) filter suffers from a trade-off between its selectivity and its in-band loss because the edges of the band-pass frequency response can be sharpened only by increasing the order of the filter, i.e., the number of cascaded sections within the filter. Now we note from Chapter 2 that the front-end loss directly raises the NF of the entire receiver, proving very undesirable. The filter is therefore designed to exhibit a small loss (0.5 to 1 dB) and some frequency selectivity.

Figure 4.4 plots the frequency response of a typical duplexer,³ exhibiting an in-band loss of about 2 dB and an out-of-band rejection of 30 dB at 20-MHz “offset” with respect to the receive band. That is, an interferer appearing at f_1 (20 MHz away from the RX band) is attenuated by only 30 dB, a critical issue in the design of both the receive path and the frequency synthesizer (Chapter 10).

The in-band loss of the above duplexer in the *transmit* band also proves problematic as it “wastes” some of the power amplifier output. For example, with 2-dB of loss and a 1-W

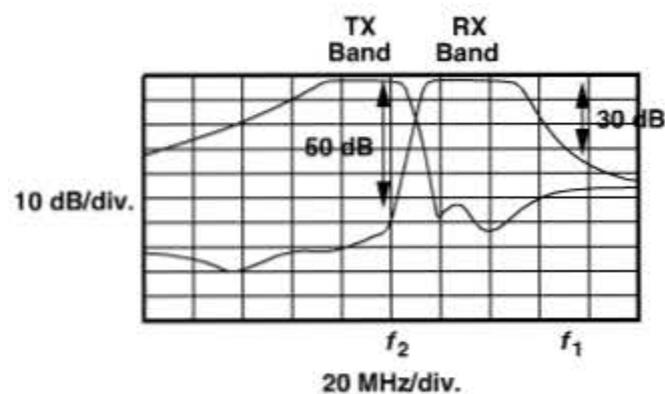


Figure 4.4 Duplexer characteristics.

3. As mentioned in Chapter 3, a duplexer consists of two band-pass filters, one for the TX band and another for the RX band.

PA, as much as 370 mW is dissipated within the duplexer—more than the typical power consumed by the entire receive path!

Our observations also indicate the importance of controlled spectral regrowth through proper choice of the modulation scheme and the power amplifier (Chapter 3). The out-of-channel energy produced by the PA *cannot* be suppressed by the front-end BPF and must be acceptably small by design.

TX-RX Feedthrough As mentioned in Chapter 3, TDD systems activate only the TX or the RX at any point in time to avoid coupling between the two. Also, even though an FDD system, GSM offsets the TX and RX time slots for the same reason. On the other hand, in full-duplex standards, the TX and the RX operate concurrently. (As explained in Chapter 3, CDMA systems require continual power control and hence concurrent TX and RX operation.) We recognize from the typical duplexer characteristics shown in Fig. 4.4 that the transmitter output at frequencies near the upper end of the TX band, e.g., at f_2 , is attenuated by only about 50 dB as it leaks to the receiver. Thus, with a 1-W TX power, the leakage sensed by the LNA can reach -20 dBm (Fig. 4.5), dictating a substantially higher RX compression point. For this reason, CDMA receivers must meet difficult linearity requirements.

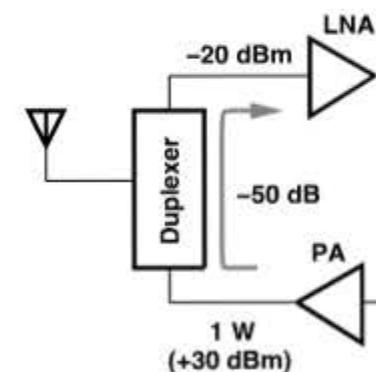


Figure 4.5 TX leakage in a CDMA transceiver.

Example 4.2

Explain how a band-pass filter following the LNA can alleviate the TX-RX leakage in a CDMA system.

Solution:

As depicted in Fig. 4.6, if the BPF provides additional rejection in the TX band, the linearity required of the rest of the RX chain is proportionally relaxed. The LNA compression point, however, must still be sufficiently high.

(Continues)

Example 4.2 (Continued)

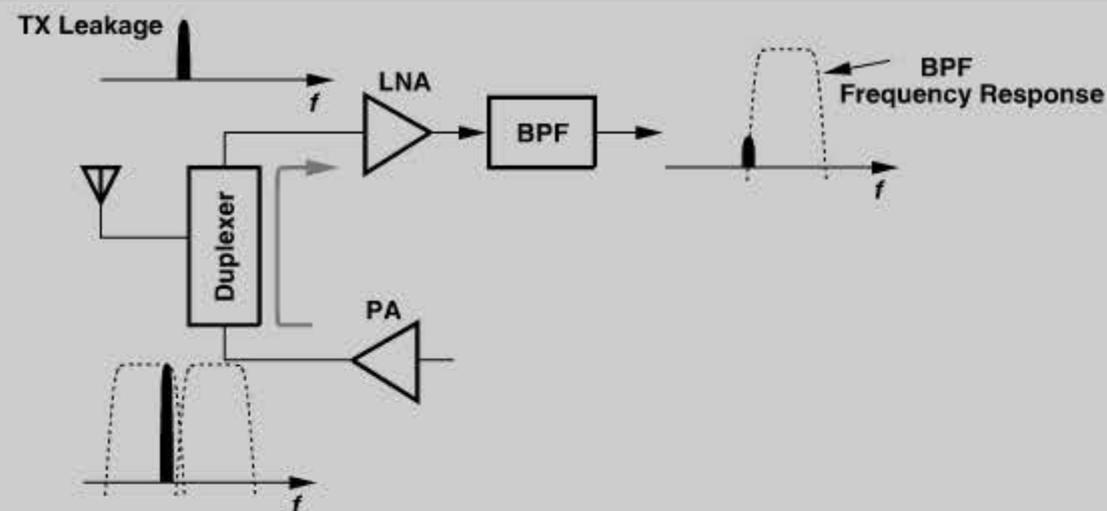


Figure 4.6 Use of BPF after LNA to suppress TX leakage.

4.2 RECEIVER ARCHITECTURES

4.2.1 Basic Heterodyne Receivers

As mentioned above, channel-selection filtering proves very difficult at high carrier frequencies. We must devise a method of “translating” the desired channel to a much lower center frequency so as to permit channel-selection filtering with a reasonable Q . Illustrated in Fig. 4.7(a), the translation is performed by means of a “mixer,” which in this chapter is viewed as a simple analog multiplier. To lower the center frequency, the signal is multiplied by a sinusoid $A_0 \cos \omega_{LO} t$, which is generated by a local oscillator (LO). Since multiplication in the time domain corresponds to convolution in the frequency domain, we observe from Fig. 4.7(b) that the impulses at $\pm \omega_{LO}$ shift the desired channel to $\pm(\omega_{in} \pm \omega_{LO})$. The components at $\pm(\omega_{in} + \omega_{LO})$ are not of interest and are removed by the low-pass filter (LPF) in Fig. 4.7(a), leaving the signal at a center frequency of $\omega_{in} - \omega_{LO}$. This operation is called “downconversion mixing” or simply “downconversion.” Due to its high noise, the downconversion mixer is preceded by a low-noise amplifier [Fig. 4.7(c)].

Called the intermediate frequency (IF), the center of the downconverted channel, $\omega_{in} - \omega_{LO}$, plays a critical role in the performance. “Heterodyne” receivers employ an LO frequency unequal to ω_{in} and hence a nonzero IF.⁴

How does a heterodyne receiver cover a given frequency band? For an N -channel band, we can envision two possibilities. (1) The LO frequency is *constant* and each RF channel

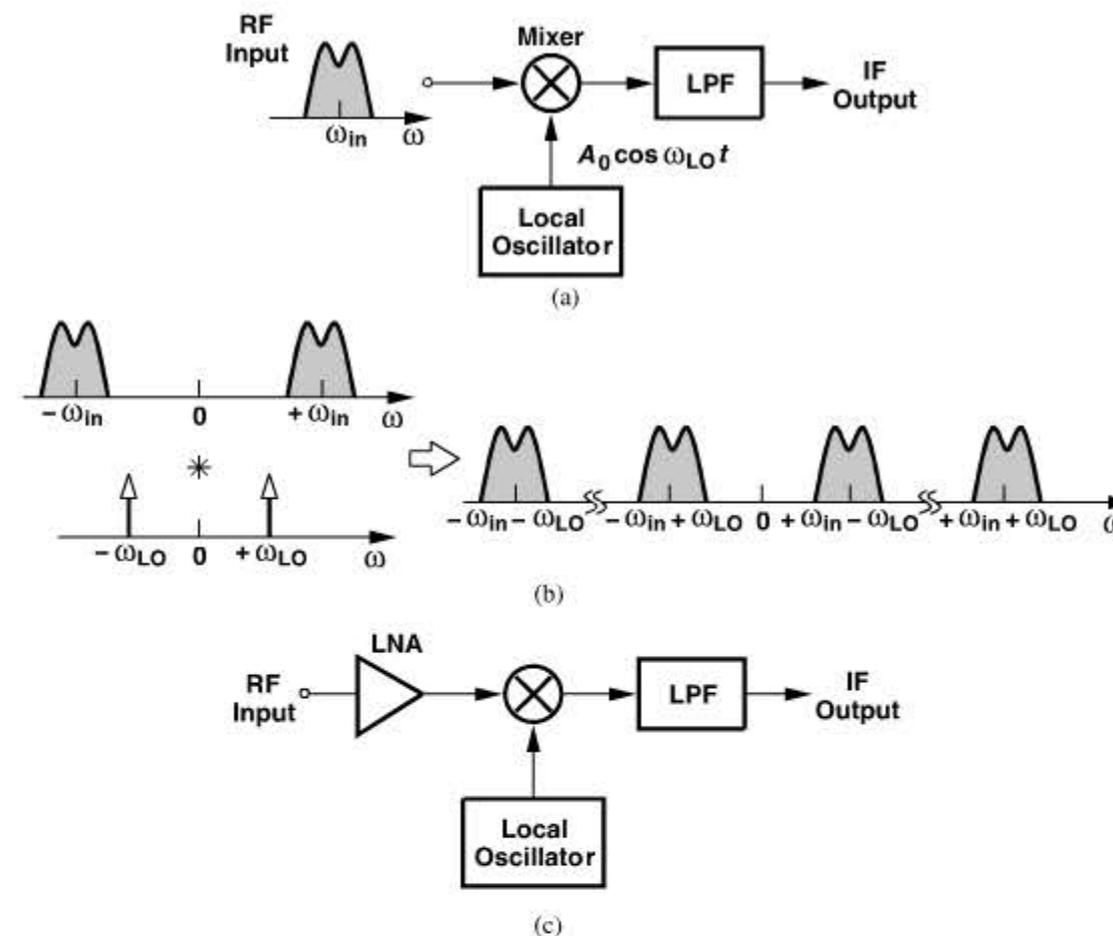


Figure 4.7 (a) Downconversion by mixing, (b) resulting spectra, (c) use of LNA to reduce noise.

is downconverted to a different IF channel [Fig. 4.8(a)], i.e., $f_{IFj} = f_{RFj} - f_{LO}$. (2) The LO frequency is *variable* so that all RF channels within the band of interest are translated to a single value of IF [Fig. 4.8(b)], i.e., $f_{LOj} = f_{RFj} - f_{IF}$. The latter case is more common as it simplifies the design of the IF path; e.g., it does not require a filter with a variable center frequency to select the IF channel of interest and reject the others. However, this approach demands a feedback loop that precisely defines the LO frequency steps, i.e., a “frequency synthesizer” (Chapters 9–11).

Problem of Image Heterodyne receivers suffer from an effect called the “image.” To understand this phenomenon, let us assume a sinusoidal input and express the IF component as

$$A \cos \omega_{IF} t = A \cos(\omega_{in} - \omega_{LO}) t \quad (4.4)$$

$$= A \cos(\omega_{LO} - \omega_{in}) t. \quad (4.5)$$

That is, whether $\omega_{in} - \omega_{LO}$ is positive or negative, it yields the same intermediate frequency. Thus, whether ω_{in} lies *above* ω_{LO} or *below* ω_{LO} , it is translated to the same IF. Figure 4.9

4. In this book, we do not distinguish between heterodyne and “super heterodyne” architectures. The term *heterodyne* derives from *hetero* (different) and *dyne* (to mix).

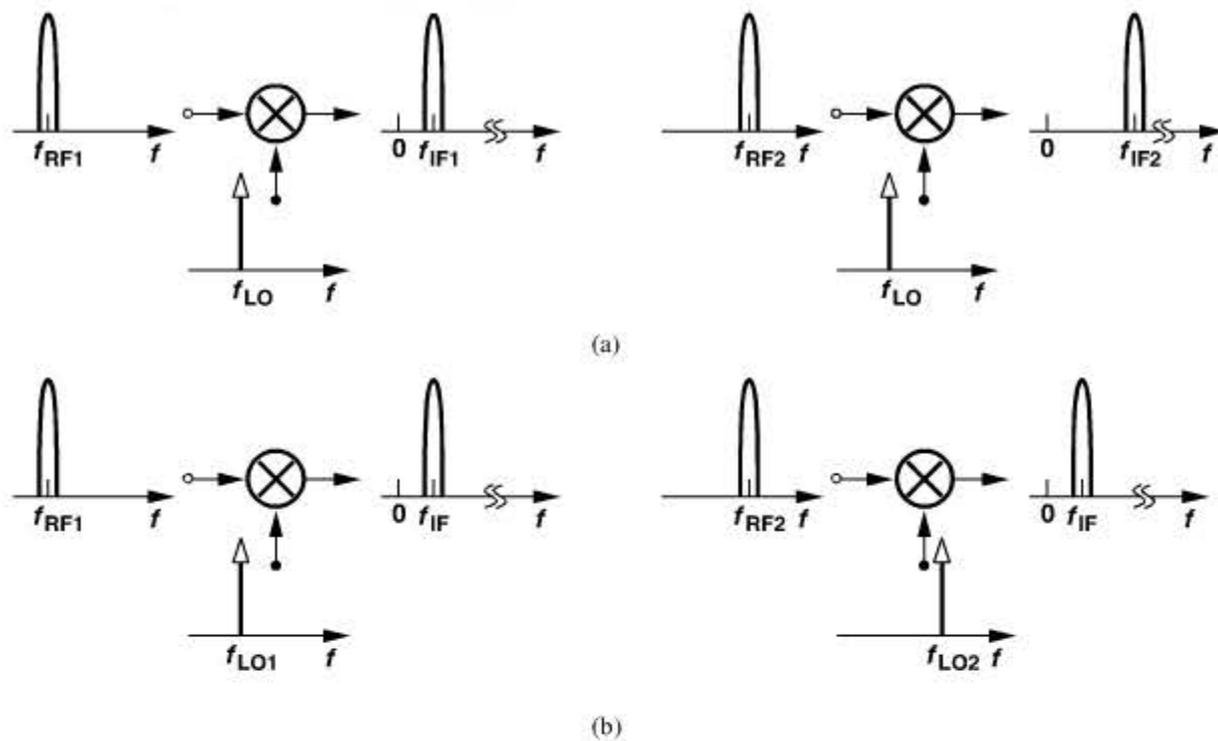


Figure 4.8 (a) Constant-LO and (b) constant-IF downconversion mixing.

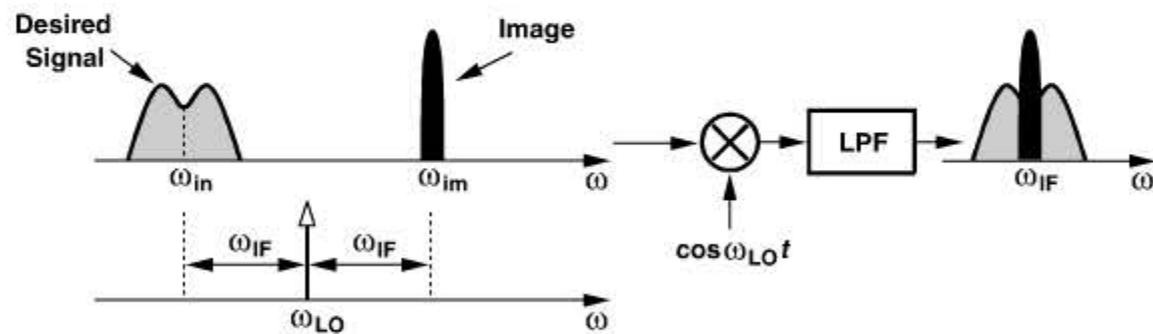


Figure 4.9 Problem of image in heterodyne downconversion.

depicts a more general case, revealing that two spectra located *symmetrically* around ω_{LO} are downconverted to the IF. Due to this symmetry, the component at ω_{im} is called the image of the desired signal. Note that $\omega_{im} = \omega_{in} + 2\omega_{IF} = 2\omega_{LO} - \omega_{in}$.

What creates the image? The numerous users in all standards (from police to WLAN bands) that transmit signals produce many interferers. If one interferer happens to fall at $\omega_{im} = 2\omega_{LO} - \omega_{in}$, then it corrupts the desired signal after downconversion.

While each wireless standard imposes constraints upon the emissions by its own users, it may have no control over the signals in other bands. The image power can therefore be much higher than that of the desired signal, requiring proper “image rejection.”

Example 4.3

Suppose two channels at ω_1 and ω_2 have been received and $\omega_1 < \omega_2$. Study the downconverted spectrum as the LO frequency varies from below ω_1 to above ω_2 .

Solution:

Shown in Fig. 4.10(a) is the case of $\omega_{LO} < \omega_1$. We note that the impulse at $-\omega_{LO}$ shifts the components at $+\omega_1$ and $+\omega_2$ to the left. Similarly, the impulse at $+\omega_{LO}$ shifts

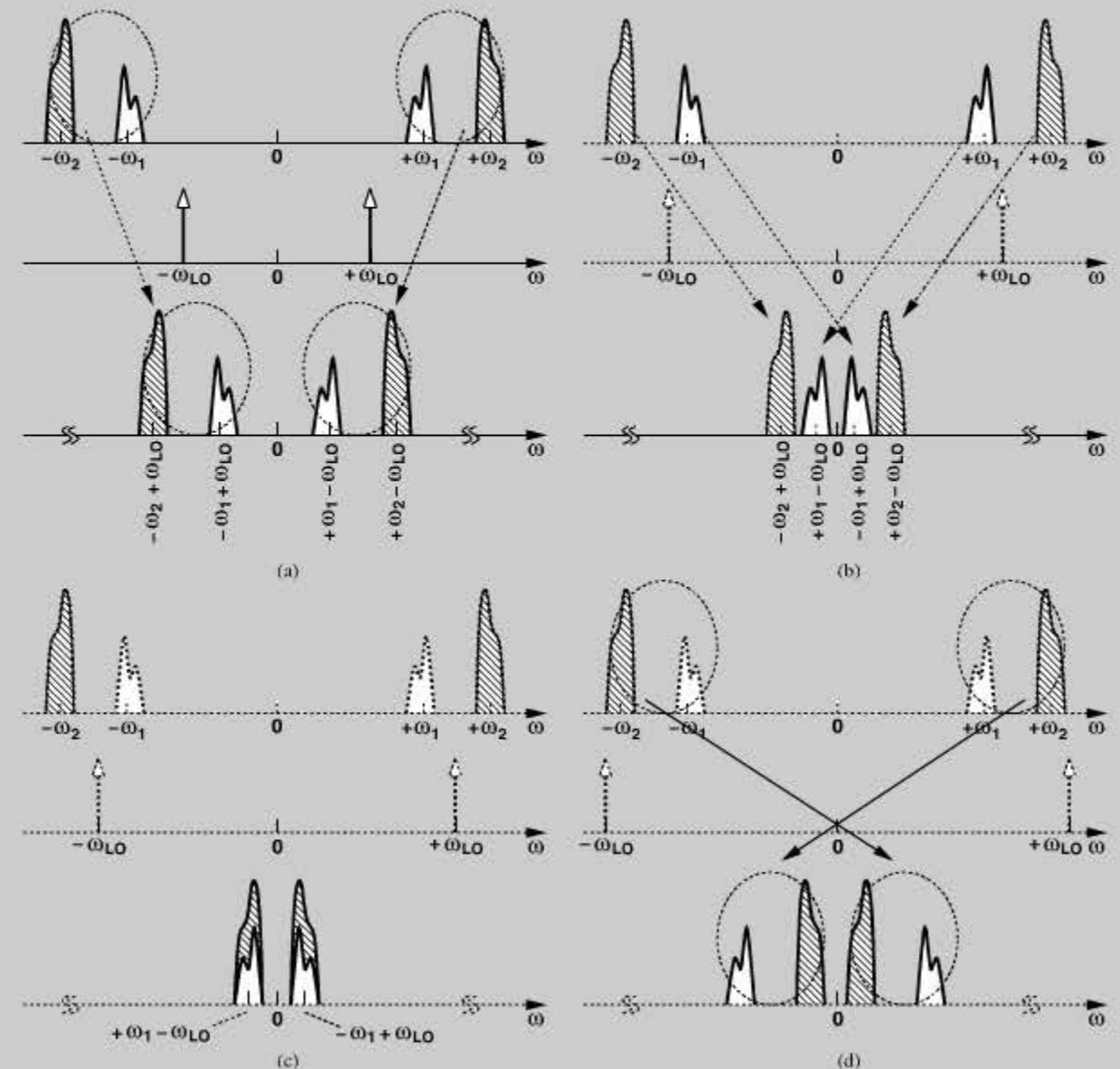


Figure 4.10 Downconversion of two channels for (a) $\omega_{LO} < \omega_1$, (b) ω_{LO} slightly above ω_1 , (c) ω_{LO} midway between ω_1 and ω_2 , and (d) $\omega_{LO} > \omega_2$.

(Continues)

Example 4.3 (Continued)

the components at $-\omega_1$ and $-\omega_2$ to the right. Since $\omega_{LO} < \omega_1$, the positive input frequencies remain positive after downconversion, and the negative input frequencies remain negative.

Now consider the case depicted in Fig. 4.10(b), where ω_{LO} is slightly greater than ω_1 . Here, after downconversion the channel at $+\omega_1$ slides to *negative* frequencies while that at $+\omega_2$ remains positive. If ω_{LO} reaches $(\omega_1 + \omega_2)/2$, then the received channels are translated such that they completely overlap each other at the IF output [Fig. 4.10(c)]. That is, ω_1 and ω_2 are images of each other. Finally, if ω_{LO} is greater than ω_2 , both positive input frequencies are shifted to negative values, and both negative input frequencies are shifted to positive values [Fig. 4.10(d)].

Example 4.4

Formulate the downconversion shown in Fig. 4.9 using expressions for the desired signal and the image.

Solution:

The two components contain modulation, assuming the forms $A_{in}(t) \cos[\omega_{in}t + \phi_{in}(t)]$ and $A_{im}(t) \cos[\omega_{im}t + \phi_{im}(t)]$, where $\omega_{im} = 2\omega_{LO} - \omega_{in}$. Upon multiplication by $A_{LO} \cos \omega_{LO}t$, they yield

$$\begin{aligned} x_{IF}(t) = & \frac{1}{2}A_{in}(t)A_{LO} \cos[(\omega_{in} + \omega_{LO})t + \phi_{in}(t)] - \frac{1}{2}A_{in}(t)A_{LO}[\cos(\omega_{in} - \omega_{LO})t + \phi_{in}t] \\ & + \frac{1}{2}A_{im}(t)A_{LO} \cos[(\omega_{im} + \omega_{LO})t + \phi_{im}(t)] \\ & - \frac{1}{2}A_{im}(t)A_{LO}[\cos(\omega_{im} - \omega_{LO})t + \phi_{im}t]. \end{aligned} \quad (4.6)$$

We observe that the components at $\omega_{in} + \omega_{LO}$ and $\omega_{im} + \omega_{LO}$ are removed by low-pass filtering, and those at $\omega_{in} - \omega_{LO} = -\omega_{IF}$ and $\omega_{im} - \omega_{LO} = +\omega_{IF}$ coincide. The corruption is given by the ratio of the rms values of $A_{im}(t)$ and $A_{in}(t)$.

High-Side and Low-Side Injection In the case illustrated in Fig. 4.9, the LO frequency is *above* the desired channel. Alternatively, ω_{LO} can be chosen below the desired channel frequency. These two scenarios are called “high-side injection” and “low-side injection,” respectively.⁵ The choice of one over the other is governed by issues such as high-frequency design issues of the LO, the strength of the image-band interferers, and other system requirements.

5. These have also been called “superdyne” and “infradyne,” respectively.

Example 4.5

The designer of an IEEE802.11g receiver attempts to place the image frequency in the GPS band, which contains only low-level satellite transmissions and hence no strong interferers. Is this possible?

Solution:

The two bands are shown in Fig. 4.11. The LO frequency must cover a range of 80 MHz but, unfortunately, the GPS band spans only 20 MHz. For example, if the lowest LO frequency is chosen so as to make 1.565 GHz the image of 2.4 GHz, then 802.11g channels above 2.42 GHz have images beyond the GPS band.

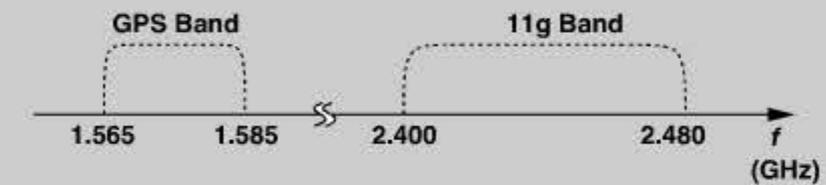


Figure 4.11 Attempt to make the GPS band the image of an 11g receiver.

Example 4.6

A dual-mode receiver is designed for both 802.11g and 802.11a. Can this receiver operate with a single LO?

Solution:

Figure 4.12(a) depicts the two bands. We choose the LO frequency halfway between the two so that a single LO covers the 11g band by high-side injection and the 11a band by low-side injection [Fig. 4.12(b)]. This greatly simplifies the design of the system but makes

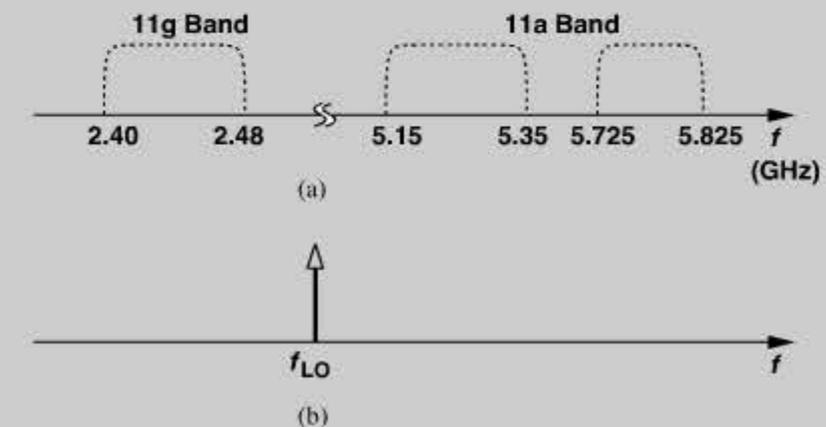


Figure 4.12 (a) 11g and 11a bands, (b) choice of f_{LO} .

(Continues)

Example 4.6 (Continued)

each band the image of the other. For example, if the receiver is in the 11a mode while an 11g transmitter operates in close proximity, the reception may be heavily corrupted. Note that also the IF in this case is quite high, an issue revisited later.

Image Rejection If the choice of the LO frequency leads to an image frequency in a high-interference band, the receiver must incorporate a means of suppressing the image. The most common approach is to precede the mixer with an “image-reject filter.” As shown in Fig. 4.13, the filter exhibits a relatively small loss in the desired band and a large attenuation in the image band, two requirements that can be simultaneously met if $2\omega_{IF}$ is sufficiently large.

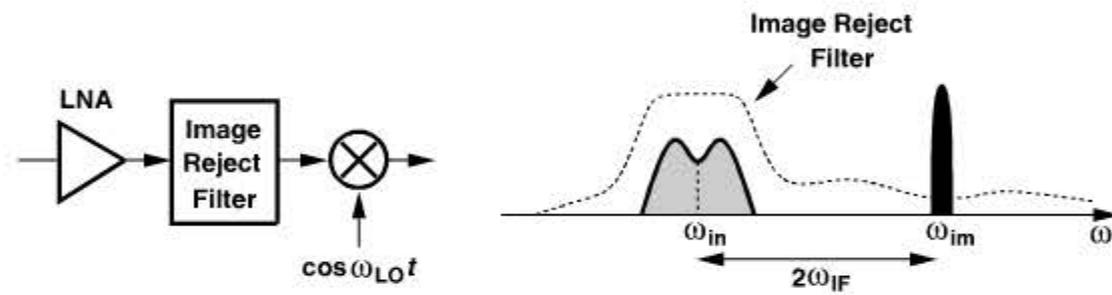


Figure 4.13 Image rejection by filtering.

Can the filter be placed before the LNA? More generally, can the front-end band-select filter provide image rejection? Yes, but since this filter’s in-band loss proves critical, its selectivity and hence out-of-band attenuation are inadequate.⁶ Thus, a filter with high image rejection typically appears between the LNA and the mixer so that the gain of the LNA lowers the filter’s contribution to the receiver noise figure.

The linearity and selectivity required of the image-reject filter have dictated passive, off-chip implementations. Operating at high frequencies, the filters are designed to provide 50- Ω input and output impedances. The LNA must therefore drive a load impedance of 50 Ω , a difficult and power-hungry task.

Image Rejection versus Channel Selection As noted in Fig. 4.13, the desired channel and the image have a frequency difference of $2\omega_{IF}$. Thus, to maximize image rejection, it is desirable to choose a large value for ω_{IF} , i.e., a large difference between ω_{in} and ω_{LO} . How large can $2\omega_{IF}$ be? Recall that the premise in a heterodyne architecture is to translate the center frequency to a sufficiently *low* value so that channel selection by means

6. As mentioned earlier, passive filters suffer from a trade-off between the in-band loss and the out-of-band attenuation.

of practical filters becomes feasible. However, as $2\omega_{IF}$ increases, so does the center of the downconverted channel (ω_{IF}), necessitating a higher Q in the IF filter.

Shown in Fig. 4.14 are two cases corresponding to high and low values of IF so as to illustrate the trade-off. A high IF [Fig. 4.14(a)] allows substantial rejection of the image whereas a low IF [Fig. 4.14(b)] helps with the suppression of in-band interferers. We thus say heterodyne receivers suffer from a trade-off between image rejection and channel selection.

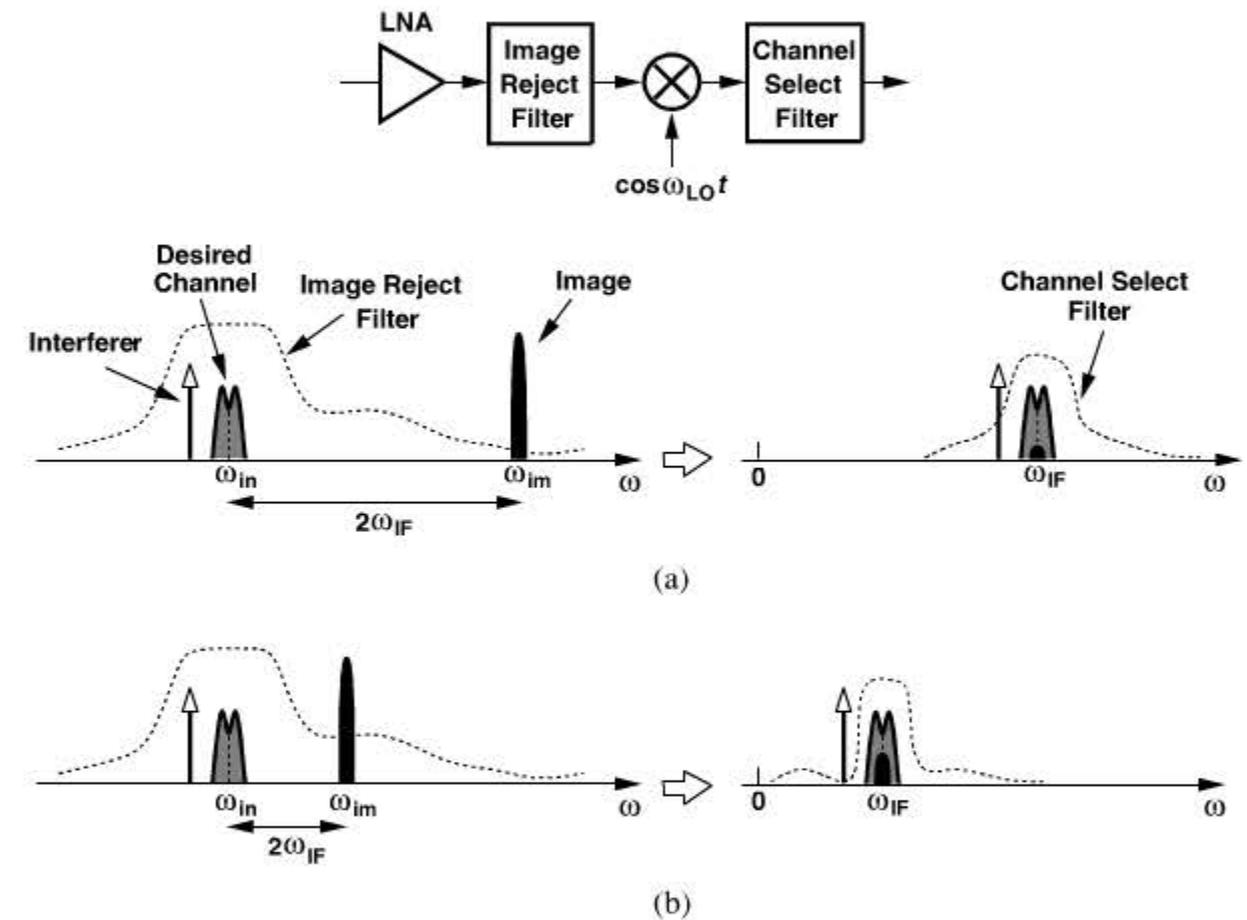


Figure 4.14 Trade-off between image rejection and channel selection for (a) high IF and (b) low IF.

Example 4.7

An engineer is to design a receiver for space applications with no concern for interferers. The engineer constructs the heterodyne front end shown in Fig. 4.15(a), avoiding band-select and image-select filters. Explain why this design suffers from a relatively high noise figure.

(Continues)

Example 4.7 (Continued)

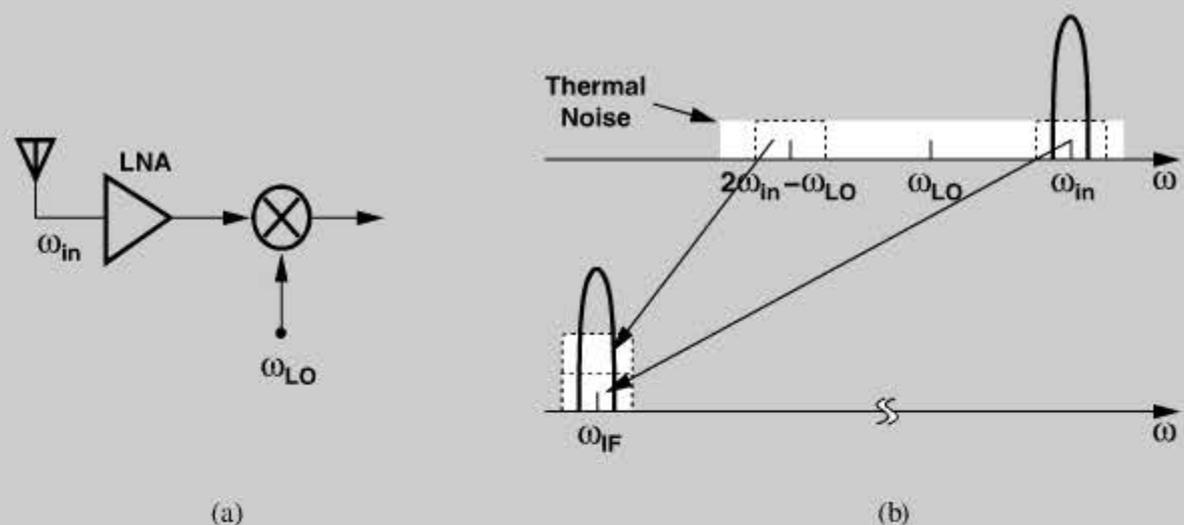


Figure 4.15 (a) Receiver for space applications, (b) effect of noise in image band.

Solution:

Even in the absence of interferers, the thermal noise produced by the antenna and the LNA in the image band arrives at the input of the mixer. Thus, the desired signal, the thermal noise in the desired channel, and the thermal noise in the image band are downconverted to IF [Fig. 4.15(b)], leading to a higher noise figure for the receiver (unless the LNA has such a limited bandwidth that it suppresses the noise in the image band). An image-reject filter would remove the noise in the image band. We return to this effect in Chapter 6.

Dual Downconversion The trade-off between image rejection and channel selection in the simple heterodyne architecture of Fig. 4.14 often proves quite severe: if the IF is high, the image can be suppressed but complete channel selection is difficult, and vice versa. To resolve this issue, the concept of heterodyning can be extended to multiple downconversions, each followed by filtering and amplification. Illustrated in Fig. 4.16, this technique performs *partial* channel selection at progressively lower center frequencies, thereby relaxing the Q required of each filter. Note that the second downconversion may also entail an image called the “secondary image” here.

Figure 4.16 also shows the spectra at different points along the cascade. The front-end filter selects the band while providing some image rejection as well. After amplification and image-reject filtering, the spectrum at point C is obtained. A sufficiently linear mixer then translates the desired channel and the adjacent interferers to the first IF (point D). Partial channel selection in BPF_3 permits the use of a second mixer with reasonable linearity. Next, the spectrum is translated to the second IF, and BPF_4 suppresses the interferers to acceptably low levels (point G). We call MX_1 and MX_2 the “RF mixer” and the “IF mixer,” respectively.

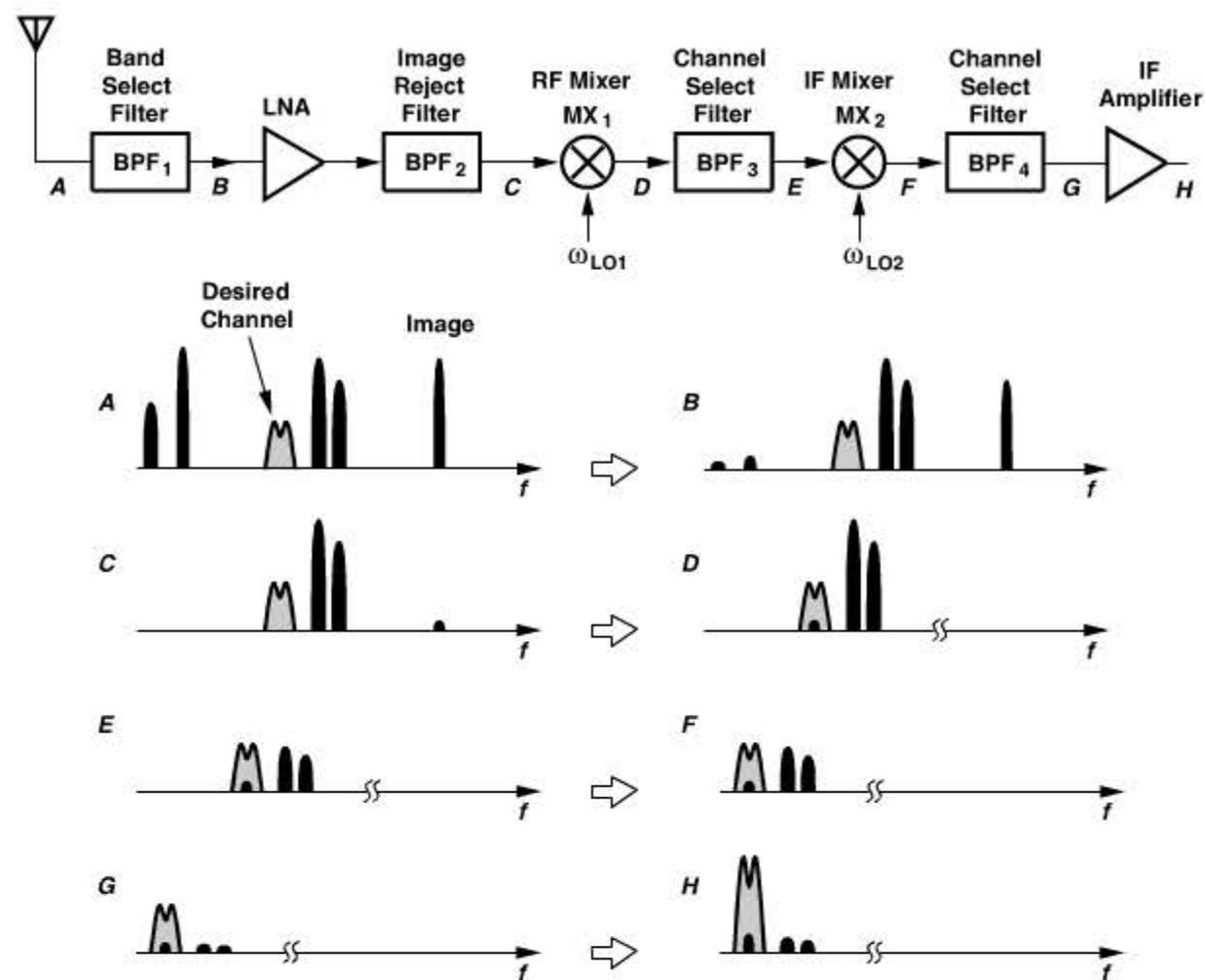


Figure 4.16 Dual-IF receiver.

Recall from Chapter 2 that in a cascade of gain stages, the noise figure is most critical in the front end and the linearity in the back end. Thus, an optimum design scales both the noise figure and the IP_3 of each stage according to the total gain preceding that stage. Now suppose the receiver of Fig. 4.16 exhibits a total gain of, say, 40 dB from A to G . If the two IF filters provided *no* channel selection, then the IP_3 of the IF amplifier would need to be about 40 dB higher than that of the LNA, e.g., in the vicinity of +30 dBm. It is difficult to achieve such high linearity with reasonable noise, power dissipation, and gain, especially if the circuit must operate from a low supply voltage. If each IF filter attenuates the in-band interferers to some extent, then the linearity required of the subsequent stages is relaxed proportionally. This is sometimes loosely stated as “every dB of gain requires 1 dB of prefiltering,” or “every dB of prefiltering relaxes the IP_3 by 1 dB.”

Example 4.8

Assuming low-side injection for both downconversion mixers in Fig. 4.16, determine the image frequencies.

(Continues)

Example 4.8 (Continued)

Solution:

As shown in Fig. 4.17, the first image lies at $2\omega_{LO1} - \omega_{in}$. The second image is located at $2\omega_{LO2} - (\omega_{in} - \omega_{LO1})$.

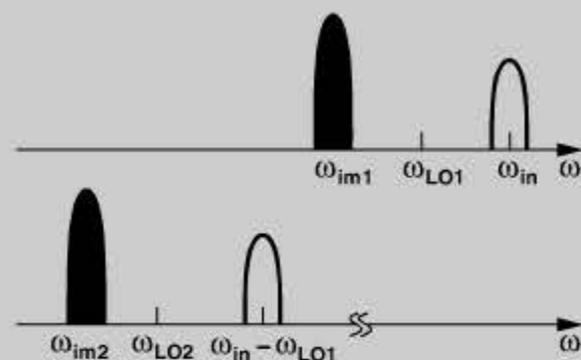


Figure 4.17 Secondary image in a heterodyne RX.

Mixing Spurs In the heterodyne receiver of Fig. 4.16, we have assumed ideal RF and IF mixers. In practice, mixers depart from simple analog multipliers, introducing undesirable effects in the receive path. Specifically, as exemplified by the switching mixer studied in Chapter 2, mixers in fact multiply the RF input by a *square-wave* LO even if the LO signal applied to the mixer is a sinusoid. As explained in Chapter 6, this internal sinusoid/square-wave conversion⁷ is inevitable in mixer design. We must therefore view mixing as multiplication of the RF input by all harmonics of the LO.⁸ In other words, the RF mixer in Fig. 4.16 produces components at $\omega_{in} \pm m\omega_{LO1}$ and the IF mixer, $\omega_{in} \pm m\omega_{LO1} \pm n\omega_{LO2}$, where m and n are integers. For the desired signal, of course, only $\omega_{in} - \omega_{LO1} - \omega_{LO2}$ is of interest. But if an interferer, ω_{int} , is downconverted to the same IF, it corrupts the signal; this occurs if

$$\omega_{int} \pm m\omega_{LO1} \pm n\omega_{LO2} = \omega_{in} - \omega_{LO1} - \omega_{LO2}. \quad (4.7)$$

Called “mixing spurs,” such interferers require careful attention to the choice of the LO frequencies.

Example 4.9

Figure 4.18(a) shows a 2.4-GHz dual downconversion receiver, where the first LO frequency is chosen so as to place the (primary) image in the GPS band for some of the channels. Determine a few mixing spurs.

7. Also called “limiting.”

8. Or only the odd harmonics of the LO if the LO and the mixer are perfectly symmetric (Chapter 6).

Example 4.9 (Continued)

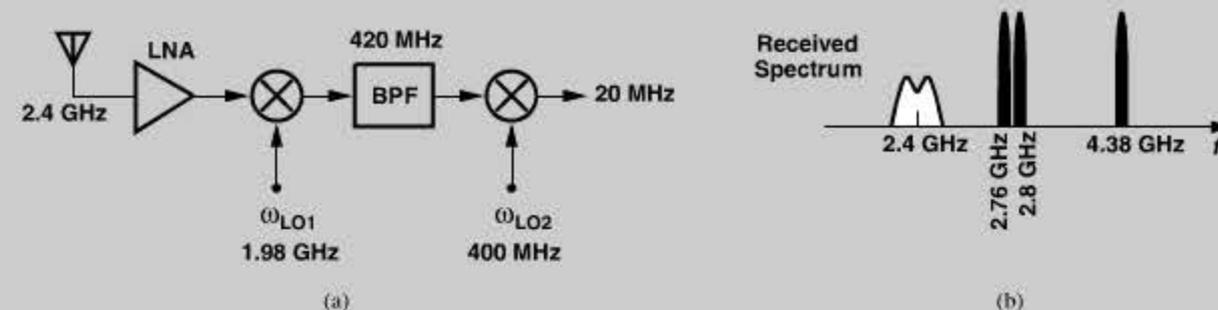


Figure 4.18 (a) Heterodyne RX for 2.4-GHz band, (b) mixing spurs.

Solution:

Let us consider the second harmonic of LO_2 , 800 MHz. If an interferer appears at the first IF at 820 MHz or 780 MHz, then it coincides with the desired signal at the second IF. In the RF band, the former corresponds to $820 \text{ MHz} + 1980 \text{ MHz} = 2.8 \text{ GHz}$ and the latter arises from $780 \text{ MHz} + 1980 \text{ MHz} = 2.76 \text{ GHz}$. We can also identify the image corresponding to the second harmonic of LO_1 by writing $f_{in} - 2f_{LO1} - f_{LO2} = 20 \text{ MHz}$ and hence $f_{in} = 4.38 \text{ GHz}$. Figure 4.18(b) summarizes these results. We observe that numerous spurs can be identified by considering other combinations of LO harmonics.

The architecture of Fig. 4.16 consists of two downconversion steps. Is it possible to use more? Yes, but the additional IF filters and LO further complicate the design and, more importantly, the mixing spurs arising from additional downconversion mixers become difficult to manage. For these reasons, most heterodyne receivers employ only two downconversion steps.

4.2.2 Modern Heterodyne Receivers

The receiver of Fig. 4.16 employs several bulky, passive (off-chip) filters and two local oscillators; it has thus become obsolete. Today’s architecture and circuit design omits all of the off-chip filters except for the front-end band-select device.

With the omission of a highly-selective filter at the first IF, no channel selection occurs at this point, thereby dictating a high linearity for the second mixer. Fortunately, CMOS mixers achieve high linearities. But the lack of a selective filter also means that the secondary image—that associated with ω_{LO2} —may become serious.

Zero Second IF To avoid the secondary image, most modern heterodyne receivers employ a *zero* second IF. Illustrated in Fig. 4.19, the idea is to place ω_{LO2} at the center of the first IF signal so that the output of the second mixer contains the desired channel with a zero center frequency. In this case, the image is the signal itself, i.e., the left part of the signal spectrum is the image of the right part and vice versa. As explained below, this effect can be handled properly. The key point here is that no interferer at other frequencies can be downconverted as an image to a zero center frequency if $\omega_{LO2} = \omega_{IF1}$.

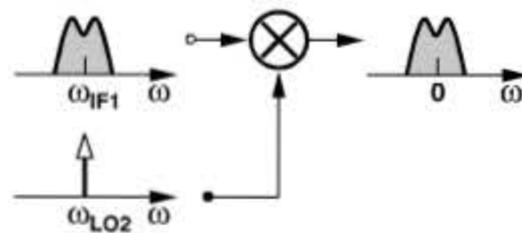


Figure 4.19 Choice of second LO frequency to avoid secondary image.

Example 4.10

Suppose the desired signal in Fig. 4.19 is accompanied by an interferer in the adjacent channel. Plot the spectrum at the second IF if $\omega_{LO2} = \omega_{IF1}$.

Solution:

Let us consider the spectra at the first IF carefully. As shown in Fig. 4.20, the desired channel appears at $\pm\omega_{IF1}$ and is accompanied by the interferer.⁹ Upon mixing in the time domain, the spectrum at negative frequencies is convolved with the LO impulse at $+\omega_{LO2}$, sliding to a zero center frequency for the desired channel. Similarly, the spectrum at positive frequencies is convolved with the impulse at $-\omega_{LO2}$ and shifted down to zero. The output thus consists of two copies of the desired channel surrounded by the interferer spectrum at both positive and negative frequencies.

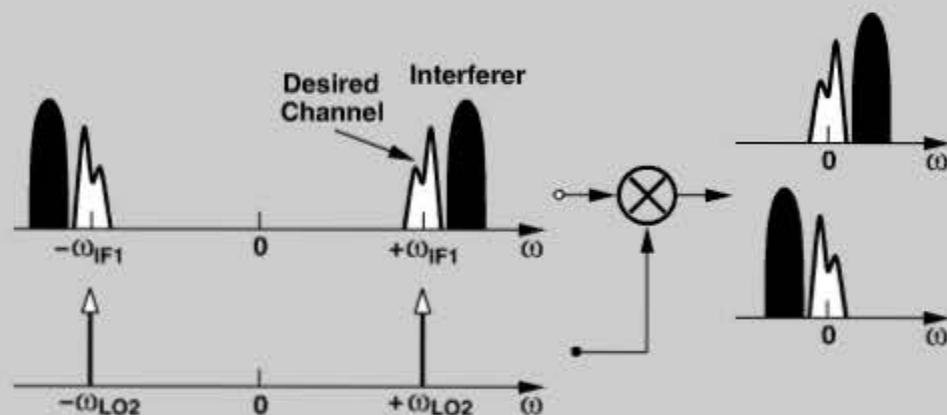


Figure 4.20 Downconversion of a desired signal and an interferer in the adjacent channel.

What happens if the signal becomes its own image? To understand this effect, we must distinguish between “symmetrically-modulated” and “asymmetrically-modulated” signals. First, consider the generation of an AM signal, Fig. 4.21(a), where a real baseband signal having a symmetric spectrum $S_a(f)$ is mixed with a carrier, thereby producing an output spectrum that remains symmetric with respect to f_{LO} . We say AM signals are symmetric because their *modulated* spectra carry exactly the same information on both sides of the carrier.¹⁰

9. The spectrum of a real signal is symmetric with respect to the origin.

10. In fact, it is possible to remove one side without losing information.

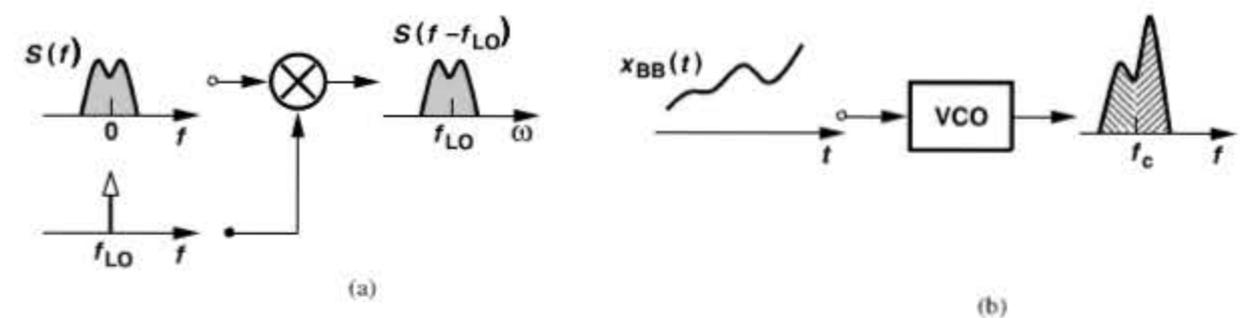


Figure 4.21 (a) AM signal generation, (b) FM signal generation.

Now, consider an FM signal generated by a voltage-controlled oscillator [Fig. 4.21(b)] (Chapter 8). We note that as the baseband voltage becomes more positive, the output frequency, say, increases, and vice versa. That is, the information in the sideband below the carrier is different from that in the sideband above the carrier. We say FM signals have an asymmetric spectrum. Most of today’s modulation schemes, e.g., FSK, QPSK, GMSK, and QAM, exhibit asymmetric spectra around their carrier frequencies. While the conceptual diagram in Fig. 4.21(b) shows the asymmetry in the *magnitude*, some modulation schemes may exhibit asymmetry in only their phase.

As exemplified by the spectra in Fig. 4.20, downconversion to a zero IF superimposes two copies of the signal, thereby causing corruption if the signal spectrum is asymmetric. Figure 4.22 depicts this effect more explicitly.

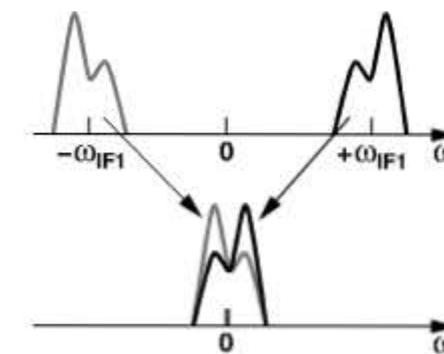


Figure 4.22 Overlap of signal sidebands after second downconversion.

Example 4.11

Downconversion to what minimum intermediate frequency avoids self-corruption of asymmetric signals?

Solution:

To avoid self-corruption, the downconverted spectra must not overlap each other. Thus, as shown in Fig. 4.23, the signal can be downconverted to an IF equal to *half* of the signal bandwidth. Of course, an interferer may now become the image.

(Continues)

Example 4.11 (Continued)

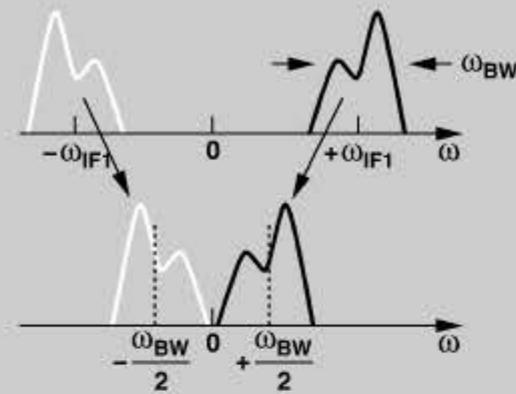


Figure 4.23 Downconversion without overlap of signal sidebands.

How can downconversion to a zero IF avoid self-corruption? This is accomplished by creating *two* versions of the downconverted signal that have a phase difference of 90° . Illustrated in Fig. 4.24, “quadrature downconversion” is performed by mixing $x_{IF}(t)$ with the quadrature phases of the second LO ($\omega_{LO2} = \omega_{IF1}$). The resulting outputs, $x_{BB,I}(t)$ and $x_{BB,Q}(t)$, are called the “quadrature baseband signals.” Though exhibiting identical spectra, $x_{BB,I}(t)$ and $x_{BB,Q}(t)$ are separated in phase and together can reconstruct the original information. In Problem 4.8, we show that even an AM signal of the form $A(t) \cos \omega_c t$ may require quadrature downconversion.

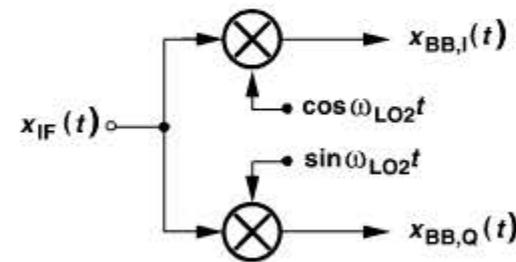


Figure 4.24 Quadrature downconversion.

Figure 4.25 shows a heterodyne receiver constructed after the above principles. In the absence of an (external) image-reject filter, the LNA need not drive a $50\text{-}\Omega$ load, and the LNA/mixer interface can be optimized for gain, noise, and linearity with little concern for the interface impedance values. However, the lack of an image-reject filter requires careful attention to the interferers in the image band, and dictates a narrow-band LNA design so that the thermal noise of the antenna and the LNA in the image band is heavily suppressed (Example 4.7). Moreover, no channel-select filter is shown at the first IF, but some “mild” on-chip band-pass filtering is usually inserted here to suppress out-of-band interferers. For example, the RF mixer may incorporate an LC load, providing some filtration.

Sliding-IF Receivers Modern heterodyne receivers entail another important departure from their older counterparts: they employ only one oscillator. This is because the design of oscillators and frequency synthesizers proves difficult and, more importantly, oscillators

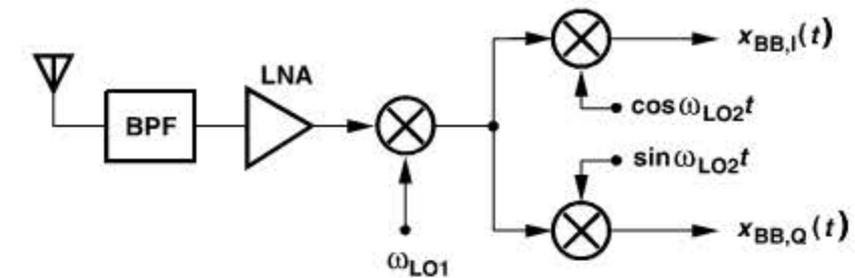


Figure 4.25 Heterodyne RX with quadrature downconversion.

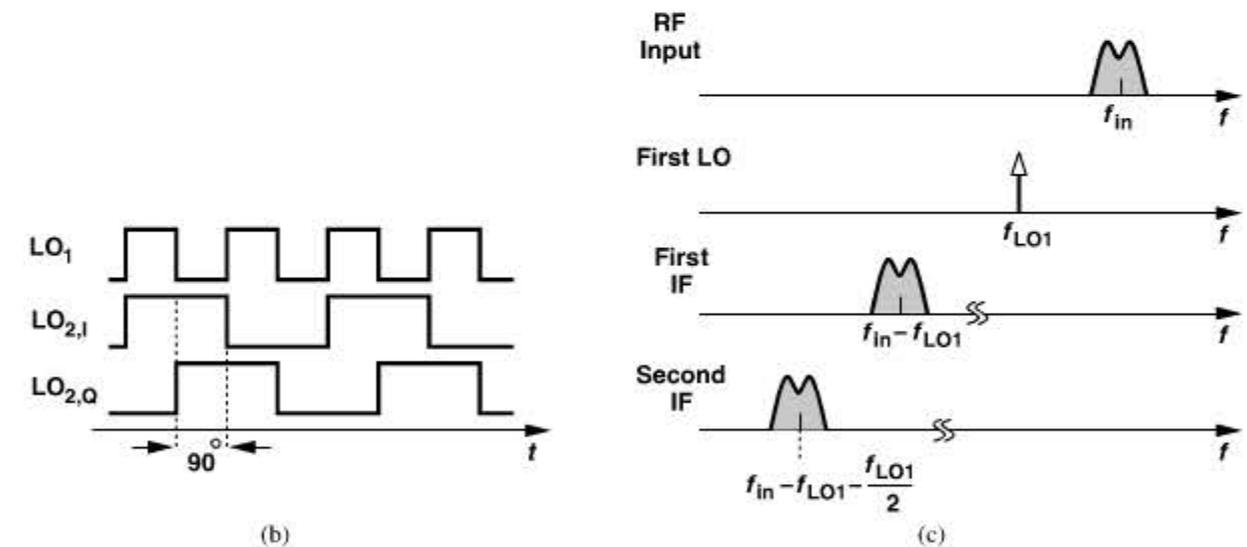
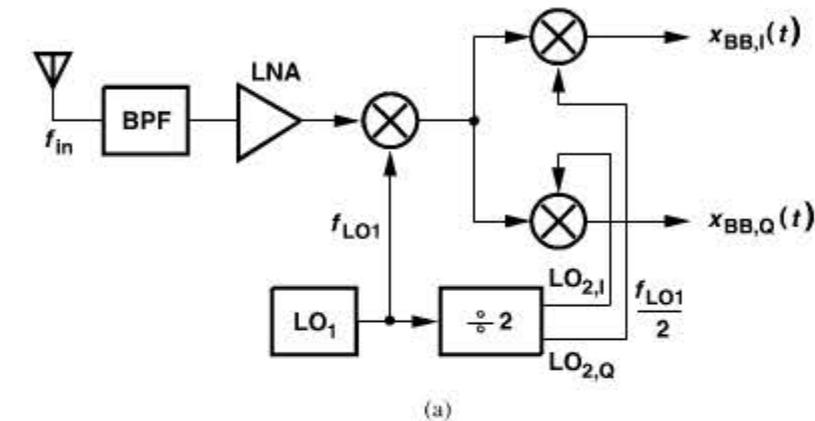


Figure 4.26 (a) Sliding-IF heterodyne RX, (b) divide-by-2 circuit waveforms, (c) resulting spectra.

fabricated on the same chip suffer from unwanted coupling. The second LO frequency is therefore *derived* from the first by “frequency division.”¹¹ Shown in Fig. 4.26(a) is an example, where the first LO is followed by a $\div 2$ circuit to generate the second LO waveforms at a frequency of $f_{LO1}/2$. As depicted in Fig. 4.26(b) and explained in Chapter 10, certain $\div 2$ topologies can produce quadrature outputs. Figure 4.26(c) shows the spectra at different points in the receiver.

11. Frequency division can be performed by a counter: for M input cycles, the counter produces one output cycle.

The receiver architecture of Fig. 4.26(a) has a number of interesting properties. To translate an input frequency of f_{in} to a second IF of zero, we must have

$$f_{LO1} + \frac{1}{2}f_{LO1} = f_{in} \quad (4.8)$$

and hence

$$f_{LO1} = \frac{2}{3}f_{in}. \quad (4.9)$$

That is, for an input band spanning the range $[f_1, f_2]$, the LO must cover a range of $[(2/3)f_1, (2/3)f_2]$ (Fig. 4.27). Moreover, the first IF in this architecture is *not* constant because

$$f_{IF1} = f_{in} - f_{LO} \quad (4.10)$$

$$= \frac{1}{3}f_{in}. \quad (4.11)$$

Thus, as f_{in} varies from f_1 to f_2 , f_{IF1} goes from $f_1/3$ to $f_2/3$ (Fig. 4.27). For this reason, this topology is called the “sliding-IF architecture.” Unlike the conventional heterodyne receiver of Fig. 4.16, where the first IF filter must exhibit a *narrow* bandwidth to perform some channel selection, this sliding IF topology requires a fractional (or normalized) IF bandwidth¹² equal to the RF input fractional bandwidth. This is because the former is given by

$$BW_{IF,frac} = \frac{\frac{1}{3}f_2 - \frac{1}{3}f_1}{\left(\frac{1}{3}f_2 + \frac{1}{3}f_1\right)/2}, \quad (4.12)$$

and the latter,

$$BW_{RF,frac} = \frac{f_2 - f_1}{(f_2 + f_1)/2}. \quad (4.13)$$

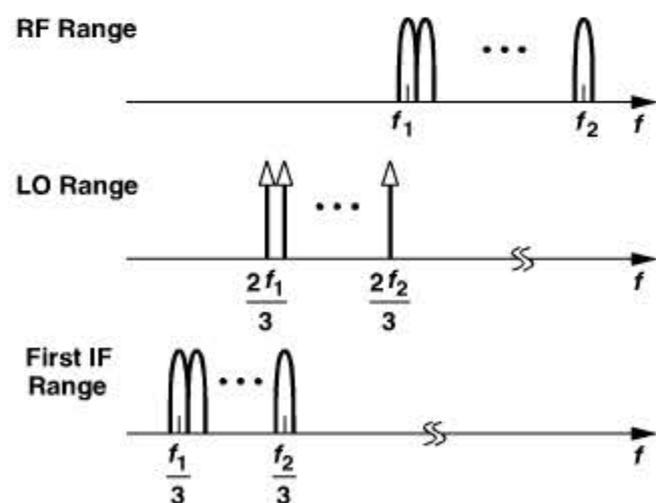


Figure 4.27 LO and IF ranges in the sliding-IF RX.

12. Fractional bandwidth is defined as the bandwidth of interest divided by the center frequency of the band.

Example 4.12

Suppose the input band is partitioned into N channels, each having a bandwidth of $(f_2 - f_1)/N = \Delta f$. How does the LO frequency vary as the receiver translates each channel to a zero second IF?

Solution:

The first channel is located between f_1 and $f_1 + \Delta f$. Thus the first LO frequency is chosen equal to two-thirds of the *center* of the channel: $f_{LO} = (2/3)(f_1 + \Delta f/2)$. Similarly, for the second channel, located between $f_1 + \Delta f$ and $f_1 + 2\Delta f$, the LO frequency must be equal to $(2/3)(f_1 + 3\Delta f/2)$. In other words, the LO increments in steps of $(2/3)\Delta f$.

Example 4.13

With the aid of the frequency bands shown in Fig. 4.27, determine the image band for the architecture of Fig. 4.26(a).

Solution:

For an LO frequency of $(2/3)f_1$, the image lies at $2f_{LO} - f_{in} = f_1/3$. Similarly, if $f_{LO1} = (2/3)f_2$, then the image is located at $f_2/3$. Thus, the image band spans the range $[f_1/3, f_2/3]$ (Fig. 4.28). Interestingly, the image band is *narrower* than the input band.

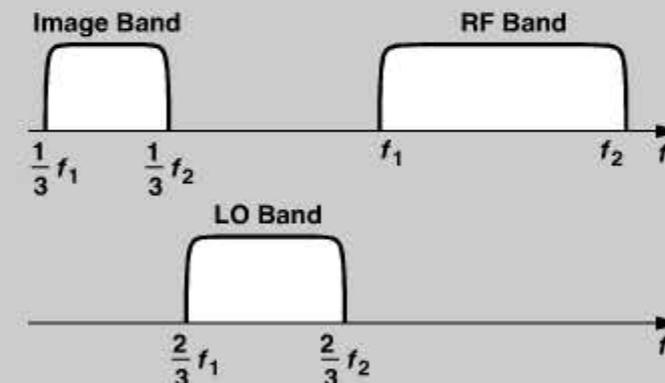


Figure 4.28 Image band in the sliding-IF RX.

Does this mean that the image for each channel is also narrower? No, recall from the above example that the LO increments by $(2/3)\Delta f$ as we go from one channel to the next. Thus, consecutive image channels have an overlap of $\Delta f/3$.

The sliding-IF architecture may incorporate greater divide ratios in the generation of the second LO from the first. For example, a $\div 4$ circuit produces quadrature outputs at $f_{LO1}/4$, leading to the following relationship

$$f_{LO1} + \frac{1}{4}f_{LO1} = f_{in} \quad (4.14)$$

and hence

$$f_{LO1} = \frac{4}{5}f_{in}. \quad (4.15)$$

The detailed spectra of such an architecture are studied in Problem 4.1. But we must make two observations here. (1) With a $\div 4$ circuit, the second LO frequency is equal to $f_{in}/5$, slightly lower than that of the first sliding-IF architecture. This is desirable because generation of LO quadrature phases at lower frequencies incurs smaller mismatches. (2) Unfortunately, the use of a $\div 4$ circuit reduces the frequency difference between the image and the signal, making it more difficult to reject the image and even the thermal noise of the antenna and the LNA in the image band. In other words, the choice of the divide ratio is governed by the trade-off between quadrature accuracy and image rejection.

Example 4.14

We wish to select a sliding-IF architecture for an 802.11g receiver. Determine the pros and cons of a $\div 2$ or a $\div 4$ circuit in the LO path.

Solution:

With a $\div 2$ circuit, the 11g band (2.40–2.48 GHz) requires an LO range of 1.600–1.653 GHz and hence an image range of 800–827 MHz [Fig. 4.29(a)]. Unfortunately, since the CDMA transmit band begins at 824 MHz, such a sliding-IF receiver may experience a large image in the range of 824–827 MHz.

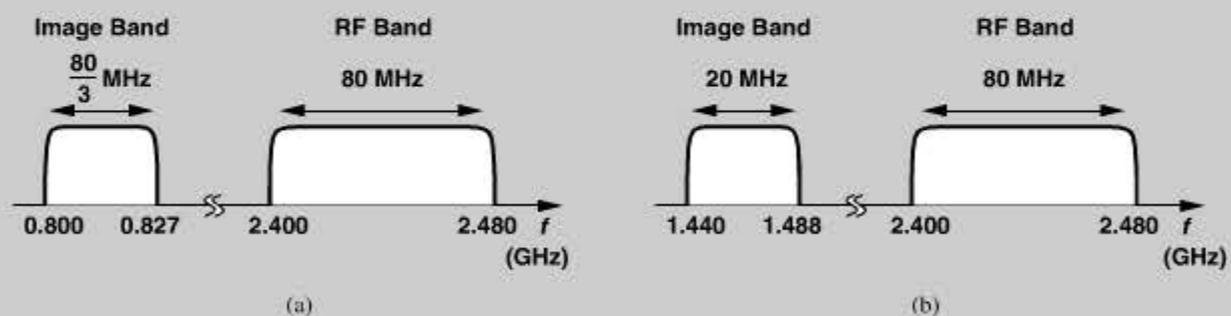


Figure 4.29 Image band in an 11g RX with a (a) divide-by-2 circuit, (b) divide-by-4 circuit.

With a $\div 4$ circuit, the LO range is 1.920–1.984 GHz and the image range, 1.440–1.488 GHz [Fig. 4.29(b)]. This image band is relatively quiet. (Only Japan has allocated a band around 1.4 GHz to WCDMA.) Thus, the choice of the $\div 4$ ratio proves advantageous here if the LNA selectivity can suppress the thermal noise in the image band. The first IF is lower in the second case and may be beneficial in some implementations.

The baseband signals produced by the heterodyne architecture of Fig. 4.26(a) suffer from a number of critical imperfections, we study these effects in the context of direct-conversion architectures.

4.2.3 Direct-Conversion Receivers

In our study of heterodyne receivers, the reader may have wondered why the RF spectrum is not simply translated to the baseband in the first downconversion. Called the “direct-conversion,” “zero-IF,” or “homodyne” architecture,¹³ this type of receiver entails its own issues but has become popular in the past decade. As explained in Section 4.2.2 and illustrated in Fig. 4.22, downconversion of an asymmetrically-modulated signal to a zero IF leads to self-corruption unless the baseband signals are separated by their phases. The direct-conversion receiver (DCR) therefore emerges as shown in Fig. 4.30, where $\omega_{LO} = \omega_{in}$.

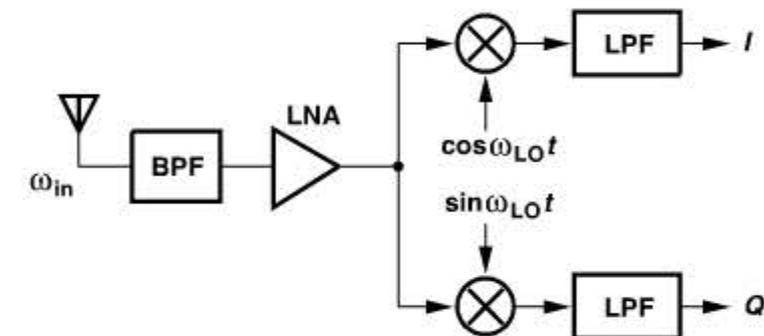


Figure 4.30 Direct-conversion receiver.

Three aspects of direct conversion make it a superior choice with respect to heterodyning. First, the absence of an image greatly simplifies the design process. Second, channel selection is performed by *low-pass* filters, which can be realized on-chip as active circuit topologies with relatively sharp cut-off characteristics. Third, mixing spurs are considerably reduced in number and hence simpler to handle.

The architecture of Fig. 4.30 appears to easily lend itself to integration. Except for the front-end band-select filter, the cascade of stages need not connect to external components, and the LNA/mixer interface can be optimized for gain, noise, and linearity without requiring a 50- Ω impedance. The simplicity of the architecture motivated many attempts in the history of RF design, but it was only in the 1990s and 2000s that integration and sophisticated signal processing made direct conversion a viable choice. We now describe the issues that DCRs face and introduce methods of resolving them. Many of these issues also appear in *heterodyne* receivers having a zero second IF.

LO Leakage A direct-conversion receiver *emits* a fraction of its LO power from its antenna. To understand this effect, consider the simplified topology shown in Fig. 4.31, where the LO couples to the antenna through two paths: (1) device capacitances between the LO and RF ports of the mixer and device capacitances or resistances between the output and input of the LNA; (2) the substrate to the input pad, especially because the LO employs large on-chip spiral inductors. The LO emission is undesirable because it may desensitize other receivers operating in the same band. Typical acceptable values range from -50 to -70 dBm (measured at the antenna).

13. The term *homodyne* originates from *homo* (same) and *dyne* (mixing) and has been historically used for only “coherent” reception.

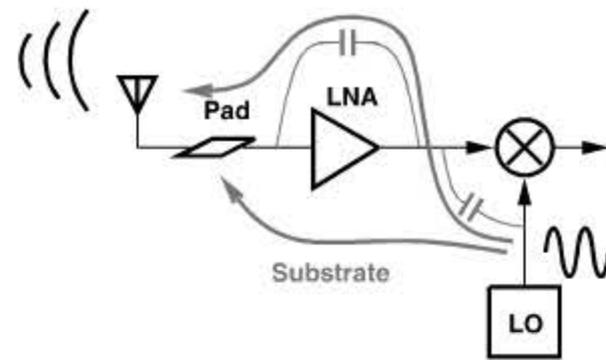


Figure 4.31 LO leakage.

Example 4.15

Determine the LO leakage from the output to the input of a cascode LNA.

Solution:

As depicted in Fig. 4.32(a), we apply a test voltage to the output and measure the voltage delivered to the antenna, R_{ant} . Considering only r_{O2} and C_{GD1} as the leakage path, we

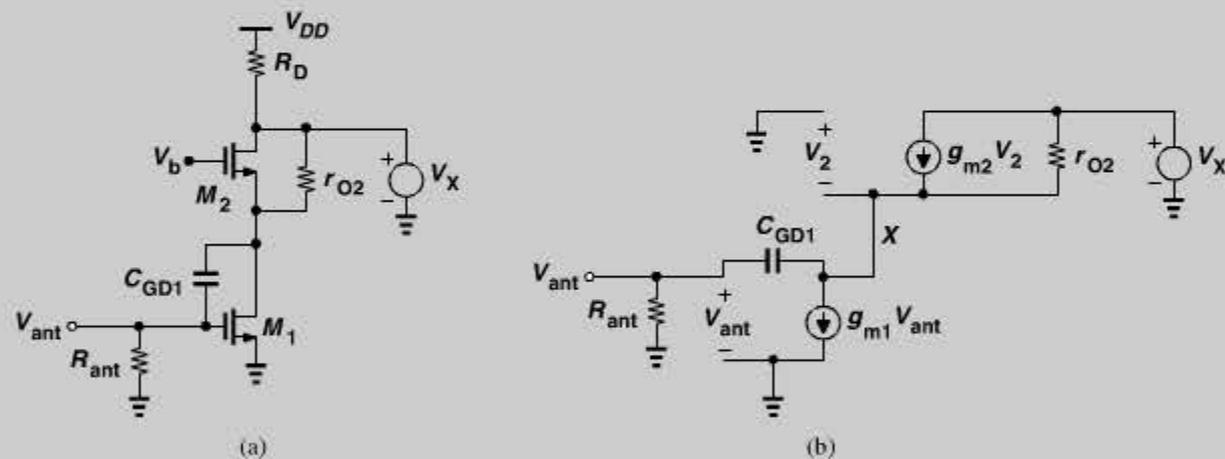


Figure 4.32 LO leakage in a cascode LNA.

construct the equivalent circuit shown in Fig. 4.32(b), note that the current flowing through R_{ant} and C_{GD1} is given by V_{ant}/R_{ant} , and write $V_2 = -[V_{ant} + V_{ant}/(R_{ant}C_{GD1}s)]$. Thus, a KCL at node X yields

$$\left(V_{ant} + \frac{V_{ant}}{R_{ant}C_{GD1}s}\right)g_{m2} + \frac{V_{ant}}{R_{ant}} + g_{m1}V_{ant} = \frac{1}{r_{O2}} \left[V_X - \left(V_{ant} + \frac{V_{ant}}{R_{ant}C_{GD1}s}\right)\right]. \quad (4.16)$$

If $g_{m2} \gg 1/r_{O2}$,

$$\frac{V_{ant}}{V_X} \approx \frac{C_{GD1}s}{(g_{m1}R_{ant} + g_{m2}R_{ant} + 1)C_{GD1}s + g_{m2}} \cdot \frac{R_{ant}}{r_{O2}}. \quad (4.17)$$

Example 4.15 (Continued)

This quantity is called the “reverse isolation” of the LNA. In a typical design, the denominator is approximately equal to g_{m2} , yielding a value of $R_{ant}C_{GD1}\omega/(g_{m2}r_{O2})$ for V_{out}/V_X .

Does LO leakage occur in heterodyne receivers? Yes, but since the LO frequency falls *outside* the band, it is suppressed by the front-end band-select filters in both the emitting receiver and the victim receiver.

LO leakage can be minimized through symmetric layout of the oscillator and the RF signal path. For example, as shown in Fig. 4.33, if the LO produces differential outputs and the leakage paths from the LO to the input pad remain symmetric, then no LO is emitted from the antenna. In other words, LO leakage arises primarily from random or deterministic *asymmetries* in the circuits and the LO waveform.

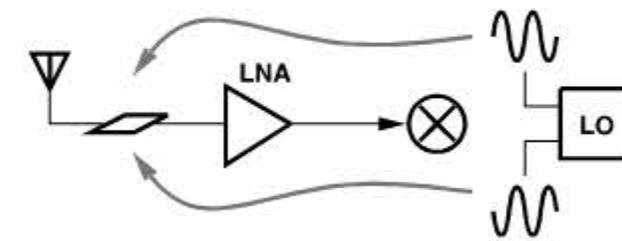


Figure 4.33 Cancellation of LO leakage by symmetry.

DC Offsets The LO leakage phenomenon studied above also gives rise to relatively large dc offsets in the baseband, thus creating certain difficulties in the design. Let us first see how the dc offset is generated. Consider the simplified receiver in Fig. 4.34, where a finite amount of in-band LO leakage, kV_{LO} , appears at the LNA input. Along with the desired signal, V_{RF} , this component is amplified and mixed with the LO. Called “LO self-mixing,” this effect produces a dc component in the baseband because multiplying a sinusoid by itself results in a dc term.

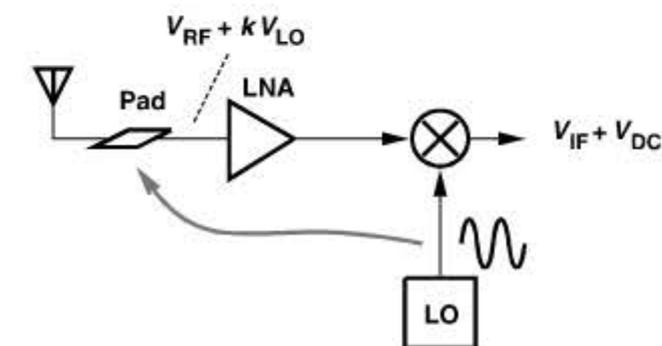


Figure 4.34 DC offset in a direct-conversion RX.

Why is a dc component troublesome? It appears that, if constant, a dc term does not corrupt the desired signal. However, such a component makes the processing of the baseband signal difficult. To appreciate the issue, we make three observations: (1) the cascade

of RF and baseband stages in a receiver must amplify the antenna signal by typically 70 to 100 dB; (as a rule of thumb, the signal at the end of the baseband chain should reach roughly 0 dBm.) (2) the received signal and the LO leakage are amplified and processed alongside each other; (3) for an RF signal level of, say, -80 dBm at the antenna, the receiver must provide a gain of about 80 dB, which, applied to an LO leakage of, say, -60 dBm, yields a very large dc offset in the baseband stages. Such an offset saturates the baseband circuits, simply prohibiting signal detection.

Example 4.16

A direct-conversion receiver incorporates a voltage gain of 30 dB from the LNA input to each mixer output and another gain of 40 dB in the baseband stages following the mixer (Fig. 4.35). If the LO leakage at the LNA input is equal to -60 dBm, determine the offset voltage at the output of the mixer and at the output of the baseband chain.

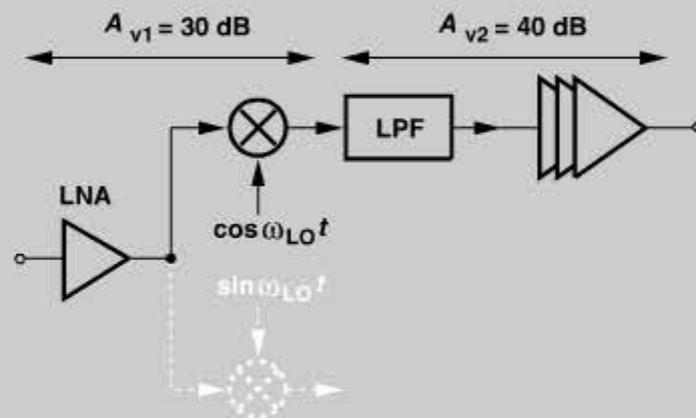


Figure 4.35 Effect of dc offset in baseband chain.

Solution:

What does $A_{V1} = 30$ dB mean? If a sinusoid $V_0 \cos \omega_{in} t$ is applied to the LNA input, then the baseband signal at the mixer output, $V_{bb} \cos(\omega_{in} - \omega_{LO})t$, has an amplitude given by

$$V_{bb} = A_{V1} \cdot V_0. \quad (4.18)$$

Thus, for an input $V_{leak} \cos \omega_{LO} t$, the dc value at the mixer output is equal to

$$V_{dc} = A_{V1} \cdot V_{leak}. \quad (4.19)$$

Since $A_{V1} = 31.6$ and $V_{leak} = (632/2) \mu\text{V}$, we have $V_{dc} = 10$ mV. Amplified by another 40 dB, this offset reaches 1 V at the baseband output!

Example 4.17

The dc offsets measured in the baseband I and Q outputs are often *unequal*. Explain why.

Example 4.17 (Continued)

Solution:

Suppose, in the presence of the quadrature phases of the LO, the net LO leakage at the input of the LNA is expressed as $V_{leak} \cos(\omega_{LO} t + \phi_{leak})$, where ϕ_{leak} arises from the phase shift through the path(s) from each LO phase to the LNA input and also the summation of the leakages $V_{LO} \cos \omega_{LO} t$ and $V_{LO} \sin \omega_{LO} t$ (Fig. 4.36). The LO leakage travels through the LNA and each mixer, experiencing an additional phase shift, ϕ_{ckt} , and is multiplied by $V_{LO} \cos \omega_{LO} t$ and $V_{LO} \sin \omega_{LO} t$. The dc components are therefore given by

$$V_{dc,I} = \alpha V_{leak} V_{LO} \cos(\phi_{leak} + \phi_{ckt}) \quad (4.20)$$

$$V_{dc,Q} = -\alpha V_{leak} V_{LO} \sin(\phi_{leak} + \phi_{ckt}). \quad (4.21)$$

Thus, the two dc offsets are generally unequal.

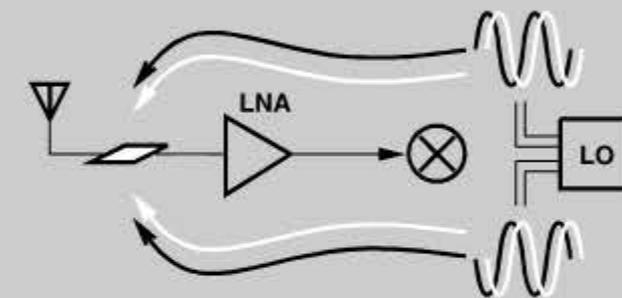


Figure 4.36 Leakage of quadrature phases of LO.

Does the problem of dc offsets occur in heterodyne receivers having a zero second IF [Fig. 4.26(a)]? Yes, the leakage of the second LO to the input of the IF mixers produces dc offsets in the baseband. Since the second LO frequency is equal to $f_{in}/3$ in Fig. 4.26(a), the leakage is smaller than that in direct-conversion receivers,¹⁴ but the dc offset is still large enough to saturate the baseband stages or at least create substantial nonlinearity.

The foregoing study implies that receivers having a final zero IF must incorporate some means of offset cancellation in each of the baseband I and Q paths. A natural candidate is a high-pass filter (ac coupling) as shown in Fig. 4.37(a), where C_1 blocks the dc offset and R_1 establishes proper bias, V_b , for the input of A_1 . However, as depicted in Fig. 4.37(b), such a network also removes a fraction of the signal's spectrum near zero frequency, thereby introducing intersymbol interference. As a rule of thumb, the corner frequency of the high-pass filter, $f_1 = (2\pi R_1 C_1)^{-1}$, must be less than one-thousandth of the symbol rate for negligible ISI. In practice, careful simulations are necessary to determine the maximum tolerable value of f_1 for a given modulation scheme.

The feasibility of on-chip ac coupling depends on both the symbol rate and the type of modulation. For example, the bit rate of 271 kb/s in GSM necessitates a corner frequency of roughly 20–30 Hz and hence extremely large capacitors and/or resistors. Note

14. Also because the LO in direct-conversion receivers employs inductors, which couple the LO waveform into the substrate, whereas the second LO in heterodyne architectures is produced by an inductor-less divider.

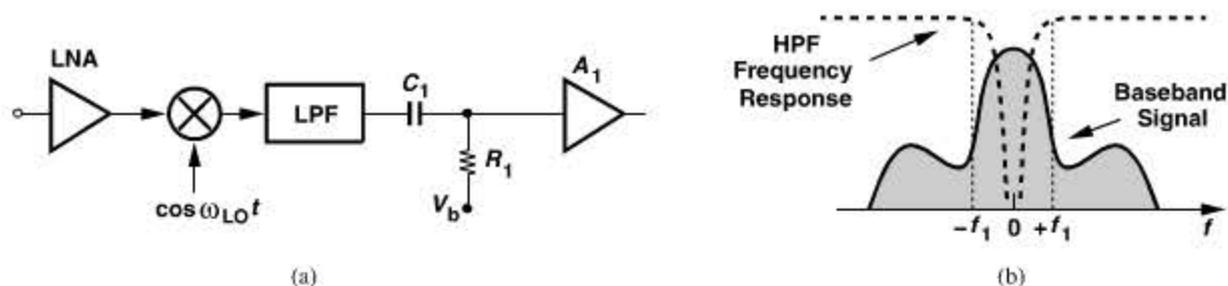


Figure 4.37 (a) Use of a high-pass filter to remove dc offset, (b) effect on signal spectrum.

that the quadrature mixers require *four* high-pass networks in their differential outputs. On the other hand, 802.11b at a maximum bit rate of 20 Mb/s can operate with a high-pass corner frequency of 20 kHz, a barely feasible value for on-chip integration.

Modulation schemes that contain little energy around the carrier better lend themselves to ac coupling in the baseband. Figure 4.38 depicts two cases for FSK signals: for a small modulation index, the spectrum still contains substantial energy around the carrier frequency, f_c , but for a large modulation index, the two frequencies generated by ONES and ZEROS become distinctly different, leaving a deep notch at f_c . If downconverted to baseband, the latter can be high-pass filtered more easily.

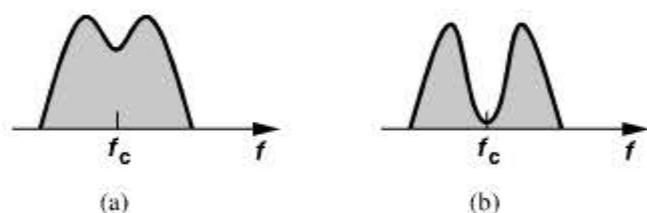


Figure 4.38 FSK spectrum with (a) small and (b) large frequency deviation.

A drawback of ac coupling stems from its slow response to transient inputs. With a very low $f_l = (2\pi R_1 C_1)^{-1}$, the circuit inevitably suffers from a long time constant, failing to block the offset if the offset suddenly changes. This change occurs if (a) the LO frequency is switched to another channel, hence changing the LO leakage, or (b) the *gain* of the LNA is switched to a different value, thus changing the reverse isolation of the LNA. (LNA gain switching is necessary to accommodate varying levels of the received signal.) For these reasons, and due to the relatively large size of the required capacitors, ac coupling is rarely used in today's direct-conversion receivers.

Example 4.18

Figure 4.39(a) shows another method of suppressing dc offsets in the baseband. Here, the main signal path consists of G_{m1} (a transconductance amplifier), R_D , and A_1 , providing a total voltage gain of $G_{m1} R_D A_1$. The negative feedback branch comprising R_1 , C_1 and $-G_{mF}$ returns a low-frequency current to node X so as to drive the dc content of V_{out} toward zero. Note that this topology suppresses the dc offsets of all of the stages in the baseband. Calculate the corner frequency of the circuit.

Example 4.18 (Continued)

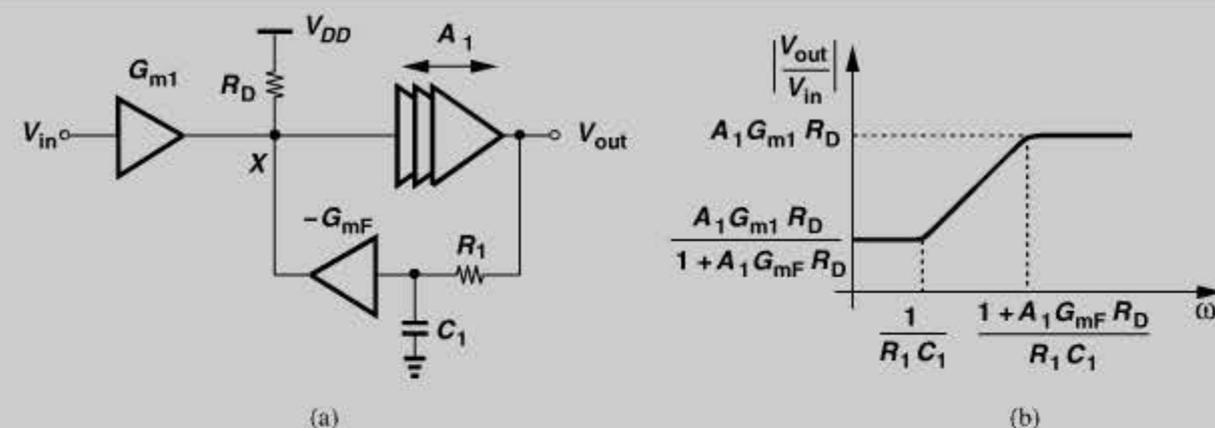


Figure 4.39 (a) Offset cancellation by feedback, (b) resulting frequency response.

Solution:

Recognizing that the current returned by $-G_{mF}$ to node X is equal to $-G_{mF} V_{out} / (R_1 C_1 s + 1)$ and the current produced by G_{m1} is given by $G_{m1} V_{in}$, we sum the two at node X , multiply the sum by R_D and A_1 , and equate the result to V_{out} .

$$\left(\frac{-G_{mF} V_{out}}{R_1 C_1 s + 1} + G_{m1} V_{in} \right) R_D A_1 = V_{out}. \quad (4.22)$$

It follows that

$$\frac{V_{out}}{V_{in}} = \frac{G_{m1} R_D A_1 (R_1 C_1 s + 1)}{R_1 C_1 s + G_{mF} R_D A_1 + 1}. \quad (4.23)$$

The circuit thus exhibits a pole at $-(1 + G_{mF} R_D A_1) / (R_1 C_1)$ and a zero at $-1 / (R_1 C_1)$ [Fig. 4.39(b)]. The input offset is amplified by a factor of $G_{m1} R_D A_1 / (1 + G_{mF} R_D A_1) \approx G_{m1} / G_{mF}$ if $G_{mF} R_D A_1 \gg 1$. This gain must remain below unity, i.e., G_{mF} is typically chosen larger than G_{m1} . Unfortunately, the high-pass corner frequency is given by

$$f_l \approx \frac{G_{mF} R_D A_1}{2\pi (R_1 C_1)}, \quad (4.24)$$

a factor of $G_{mF} R_D A_1$ higher than that of the passive circuit in Fig. 4.37(a). This “active feedback” circuit therefore requires greater values for R_1 and C_1 to provide a low f_l . The advantage is that C_1 can be realized by a MOSFET [while that in Fig. 4.37(a) cannot].

The most common approach to offset cancellation employs digital-to-analog converters (DACs) to draw a corrective current in the same manner as the G_{mF} stage in Fig. 4.39(a). Let us first consider the cascade shown in Fig. 4.40(a), where I_1 is drawn from node X

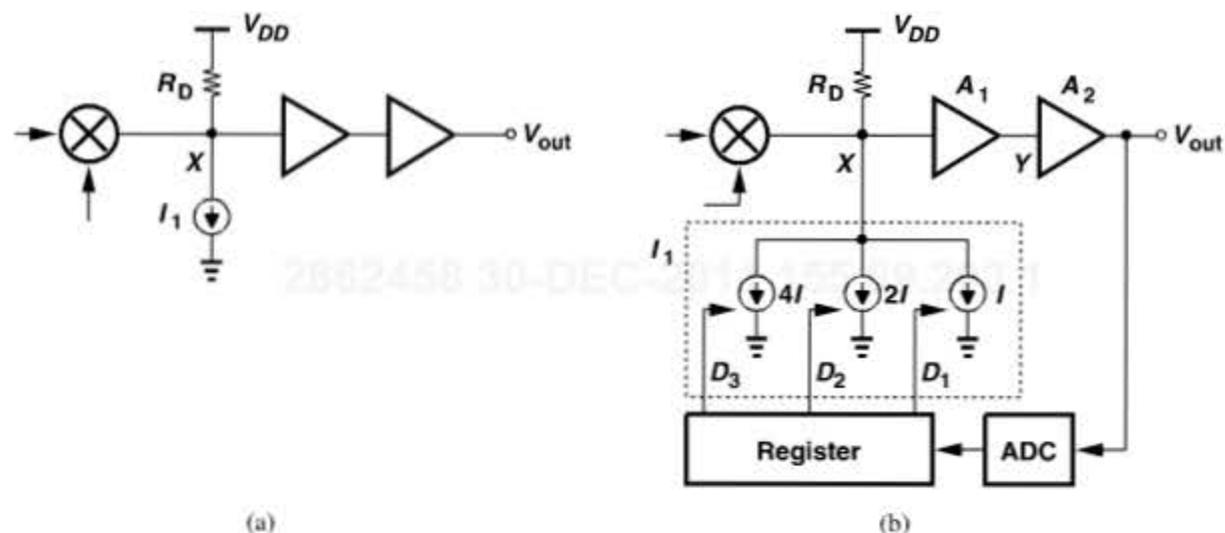


Figure 4.40 (a) Offset cancellation by means of a current source, (b) actual implementation.

and its value is adjusted so as to drive the dc content in V_{out} to zero.¹⁵ For example, if the mixer produces an offset of ΔV at X and the subsequent stages exhibit no offset, then $I_1 = \Delta V/R_D$ with proper polarity. In Fig. 4.39(a), the corrective current provided by G_{mF} is continuously adjusted (even in the presence of signal), leading to the high-pass behavior; we thus seek a method of “freezing” the value of I_1 so that it does not affect the baseband frequency response. This requires that I_1 be controlled by a register and hence vary in discrete steps. As illustrated in Fig. 4.40(b), I_1 is decomposed into units that are turned on or off according to the values stored in the register. For example, a binary word $D_3D_2D_1$ controls “binary-weighted” current sources $4I$, $2I$, and I . These current sources form a DAC.

How is the correct value of the register determined? When the receiver is turned on, an analog-to-digital converter (ADC) digitizes the baseband output (in the absence of signals) and drives the register. The entire negative-feedback loop thus converges such that V_{out} is minimized. The resulting values are then stored in the register and remain frozen during the actual operation of the receiver.

The arrangement of Fig. 4.40(b) appears rather complex, but, with the scaling of CMOS technology, the area occupied by the DAC and the register is in fact considerably smaller than that of the capacitors in Figs. 4.37(a) and 4.39(a). Moreover, the ADC is also used during signal reception.

The digital storage of offset affords other capabilities as well. For example, since the offset may vary with the LO frequency or gain settings before or after the mixer, at power-up the receiver is cycled through all possible combinations of LO and gain settings, and the required values of I_1 are stored in a small memory. During reception, for the given LO and gain settings, the value of I_1 is recalled from the memory and loaded into the register.

The principal drawback of digital storage originates from the finite resolution with which the offset is cancelled. For example, with the 3-bit DAC in Fig. 4.40(b), an offset of, say, 10 mV at node X , can be reduced to about 1.2 mV after the overall loop settles. Thus, for an A_1A_2 of, say, 40 dB, V_{out} still suffers from 120 mV of offset. To alleviate this issue,

a higher resolution must be realized or multiple DACs must be tied to different nodes (e.g., Y and V_{out}) in the cascade to limit the maximum offset.

Example 4.19

In the arrangement of Fig. 4.40(b), another 3-bit DAC is tied to node Y . If the mixer produces an offset of 10 mV and $A_1A_2 = 40$ dB, what is the minimum offset that can be achieved in V_{out} ? Assume A_1 and A_2 have no offset.

Solution:

The second DAC lowers the output offset by another factor of 8, yielding a minimum of about $10 \text{ mV} \times 100/64 \approx 16 \text{ mV}$.

Even-Order Distortion Our study of nonlinearity in Chapter 2 indicates that third-order distortion results in compression and intermodulation. Direct-conversion receivers are additionally sensitive to even-order nonlinearity in the RF path, and so are heterodyne architectures having a second zero IF.

Suppose, as shown in Fig. 4.41, two strong interferers at ω_1 and ω_2 experience a nonlinearity such as $y(t) = \alpha_1x(t) + \alpha_2x^2(t)$ in the LNA. The second-order term yields the product of these two interferers and hence a low-frequency “beat” at $\omega_2 - \omega_1$. What is the effect of this component? Upon multiplication by $\cos \omega_{LO}t$ in an ideal mixer, such a term is translated to high frequencies and hence becomes unimportant. In reality, however, asymmetries in the mixer or in the LO waveform allow a fraction of the RF input of the mixer to appear at the output *without* frequency translation. As a result, a fraction of the low-frequency beat appears in the baseband, thereby corrupting the downconverted signal. Of course, the beat generated by the LNA can be removed by ac coupling, making the input transistor of the mixer the dominant source of even-order distortion.

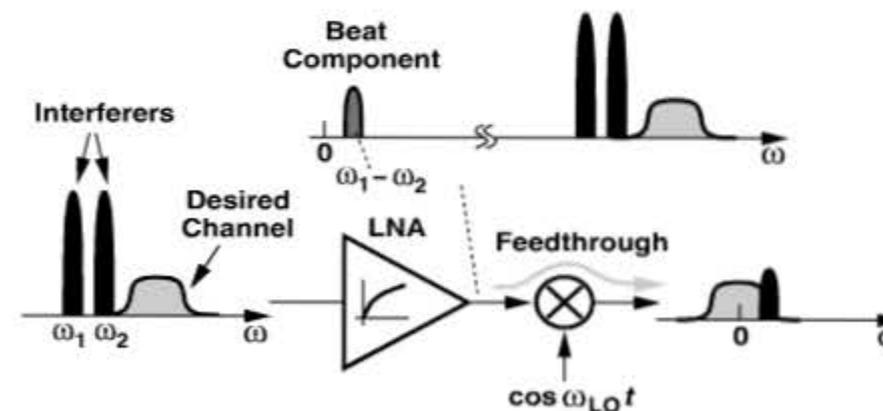


Figure 4.41 Effect of even-order distortion on direct conversion.

To understand how asymmetries give rise to direct “feedthrough” in a mixer, first consider the circuit shown in Fig. 4.42(a). As explained in Chapter 2, the output can be written as the product of V_{in} and an ideal LO, i.e., a square-wave toggling between 0 and 1 with

15. We assume that the mixer generates an output current.

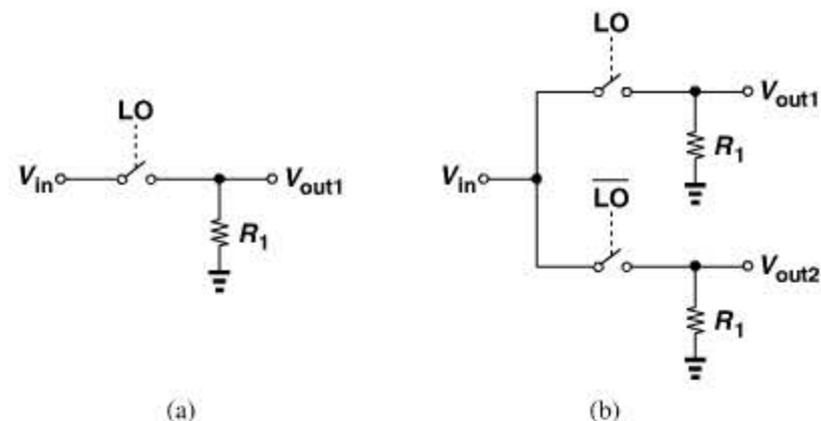


Figure 4.42 (a) Simple mixer, (b) mixer with differential output.

50% duty cycle, $S(t)$:

$$V_{out}(t) = V_{in}(t) \cdot S(t) \quad (4.25)$$

$$= V_{in}(t) \left[S(t) - \frac{1}{2} \right] + V_{in}(t) \cdot \frac{1}{2}. \quad (4.26)$$

We recognize that $S(t) - 1/2$ represents a “dc-free” square wave consisting of only odd harmonics. Thus, $V_{in}(t) \cdot [S(t) - 1/2]$ contains the product of V_{in} and the odd harmonics of the LO. The second term in (4.26), $V_{in}(t) \times 1/2$, denotes the RF feedthrough to the output (with no frequency translation).

Next, consider the topology depicted in Fig. 4.42(b), where a second branch driven by \overline{LO} (the complement of LO) produces a second output. Expressing \overline{LO} as $1 - S(t)$, we have

$$V_{out1}(t) = V_{in}(t)S(t) \quad (4.27)$$

$$V_{out2}(t) = V_{in}(t)[1 - S(t)]. \quad (4.28)$$

As with $V_{out1}(t)$, the second output $V_{out2}(t)$ contains an RF feedthrough equal to $V_{in}(t) \times 1/2$ because $1 - S(t)$ exhibits a dc content of $1/2$. If the output is sensed *differentially*, the RF feedthroughs in $V_{out1}(t)$ and $V_{out2}(t)$ are cancelled while the signal components add. It is this cancellation that is sensitive to asymmetries; for example, if the switches exhibit a mismatch between their on-resistances, then a net RF feedthrough arises in the differential output.

The problem of even-order distortion is critical enough to merit a quantitative measure. Called the “second intercept point” (IP_2), such a measure is defined according to a two-tone test similar to that for IP_3 except that the output of interest is the beat component rather than the intermodulation product. If $V_{in}(t) = A \cos \omega_1 t + A \cos \omega_2 t$, then the LNA output is given by

$$V_{out}(t) = \alpha_1 V_{in}(t) + \alpha_2 V_{in}^2(t) \quad (4.29)$$

$$= \alpha_1 A (\cos \omega_1 t + \cos \omega_2 t) + \alpha_2 A^2 \cos(\omega_1 + \omega_2)t + \alpha_2 A^2 \cos(\omega_1 - \omega_2)t + \dots, \quad (4.30)$$

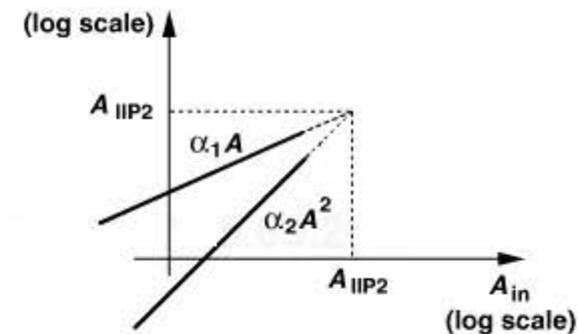


Figure 4.43 Plot illustrating IP_2 .

Revealing that the beat amplitude grows with the *square* of the amplitude of the input tones. Thus, as shown in Fig. 4.43, the beat amplitude rises with a slope of 2 on a log scale. Since the net feedthrough of the beat depends on the mixer and LO asymmetries, the beat amplitude measured in the baseband depends on the device dimensions and the layout and is therefore difficult to formulate.

Example 4.20

Suppose the attenuation factor experienced by the beat as it travels through the mixer is equal to k , whereas the gain seen by each tone as it is downconverted to the baseband is equal to unity. Calculate the IP_2 .

Solution:

From Eq. (4.30), the value of A that makes the output beat amplitude, $k\alpha_2 A^2$, equal to the main tone amplitude, $\alpha_1 A$, is given by

$$k\alpha_2 A_{IP2}^2 = \alpha_1 A_{IP2} \quad (4.31)$$

and hence

$$A_{IP2} = \frac{1}{k} \cdot \frac{\alpha_1}{\alpha_2}. \quad (4.32)$$

Even-order distortion may manifest itself even in the absence of interferers. Suppose in addition to frequency and phase modulation, the received signal also exhibits *amplitude modulation*. For example, as explained in Chapter 3, QAM, OFDM, or simple QPSK with baseband pulse shaping produce variable-envelope waveforms. We express the signal as $x_{in}(t) = [A_0 + a(t)] \cos[\omega_c t + \phi(t)]$, where $a(t)$ denotes the envelope and typically varies slowly, i.e., it is a low-pass signal. Upon experiencing second-order distortion, the signal appears as

$$\alpha_2 x_{in}^2(t) = \alpha_2 \left[A_0^2 + 2A_0 a(t) + a^2(t) \right] \frac{1 + \cos[2\omega_c t + 2\phi(t)]}{2}. \quad (4.33)$$

Both of the terms $\alpha_2 A_0 a(t)$ and $\alpha_2 a^2(t)/2$ are *low-pass* signals and, like the beat component shown in Fig. 4.41, pass through the mixer with finite attenuation, corrupting the down-converted signal. We say even-order distortion demodulates AM because the amplitude information appears as $\alpha_2 A_0 a(t)$. This effect may corrupt the signal by its own envelope or by the envelope of a large interferer. We consider both cases below.

Example 4.21

Quantify the self-corruption expressed by Eq. (4.33) in terms of the IP_2 .

Solution:

Assume, as in Example 4.20, that the low-pass components, $\alpha_2 A_0 a(t) + \alpha_2 a^2(t)/2$, experience an attenuation factor of k and the desired signal, $\alpha_1 A_0$, sees a gain of unity. Also, typically $a(t)$ is several times smaller than A_0 and hence the baseband corruption can be approximated as $k\alpha_2 A_0 a(t)$. Thus, the signal-to-noise ratio arising from self-corruption is given by

$$SNR = \frac{\alpha_1 A_0 / \sqrt{2}}{k\alpha_2 A_0 a_{rms}} \quad (4.34)$$

$$= \frac{A_{IP2}}{\sqrt{2} a_{rms}}, \quad (4.35)$$

where $A_0/\sqrt{2}$ denotes the rms signal amplitude and a_{rms} the rms value of $a(t)$.

How serious is the above phenomenon? Equation (4.35) predicts that the SNR falls to dangerously low levels as the envelope variation becomes comparable with the input IP_2 . In reality, this is unlikely to occur. For example, if $a_{rms} = -20$ dBm, then A_0 is perhaps on the order of -10 to -15 dBm, large enough to saturate the receiver chain. For such high input levels, the gain of the LNA and perhaps the mixer is switched to much lower values to avoid saturation, automatically minimizing the above self-corruption effect.

The foregoing study nonetheless points to another, much more difficult, situation. If the desired channel is accompanied by a large amplitude-modulated interferer, then even-order distortion demodulates the AM component of the interferer, and mixer feedthrough allows it to appear in the baseband. In this case, Eq. (4.34) still applies but the numerator must represent the desired signal, $\alpha_1 A_{sig}/\sqrt{2}$, and the denominator, the interferer $k\alpha_2 A_{int} a_{rms}$:

$$SNR = \frac{\alpha_1 A_{sig}/\sqrt{2}}{k\alpha_2 A_{int} a_{rms}} \quad (4.36)$$

$$= \frac{A_{IP2} A_{sig}/\sqrt{2}}{A_{int} a_{rms}}, \quad (4.37)$$

Example 4.22

A desired signal at -100 dBm is received along with an interferer $[A_{int} + a(t)] \cos[\omega_c t + \phi(t)]$, where $A_{int} = 5$ mV and $a_{rms} = 1$ mV. What IP_2 is required to ensure $SNR \geq 20$ dB?

Solution:

Since -100 dBm is equivalent to a peak amplitude of $A_{sig} = 3.16 \mu\text{V}$, we have

$$A_{IP2} = SNR \frac{A_{int} a_{rms}}{A_{sig}/\sqrt{2}} \quad (4.38)$$

$$= 22.4 \text{ V} \quad (4.39)$$

$$= +37 \text{ dBm}. \quad (4.40)$$

Note that the interferer level ($A_{int} = -36$ dBm) falls well below the compression point of typical receivers, but it can still corrupt the signal if the IP_2 is not as high as $+37$ dBm.

This study reveals the relatively high IP_2 values required in direct-conversion receivers. We deal with methods of achieving a high IP_2 in Chapter 6.

Flicker Noise Since linearity requirements typically limit the gain of the LNA/mixer cascade to about 30 dB, the downconverted signal in a direct-conversion receiver is still relatively small and hence susceptible to noise in the baseband circuits. Furthermore, since the signal is centered around zero frequency, it can be substantially corrupted by flicker noise. As explained in Chapter 6, the mixers themselves may also generate flicker noise at their output.

In order to quantify the effect of flicker noise, let us assume the downconverted spectrum shown in Fig. 4.44, where f_{BW} is half of the RF channel bandwidth. The flicker noise is denoted by $S_{1/f}$ and the thermal noise at the end of the baseband by S_{th} . The frequency at which the two profiles meet is called f_c . We wish to determine the penalty due to flicker

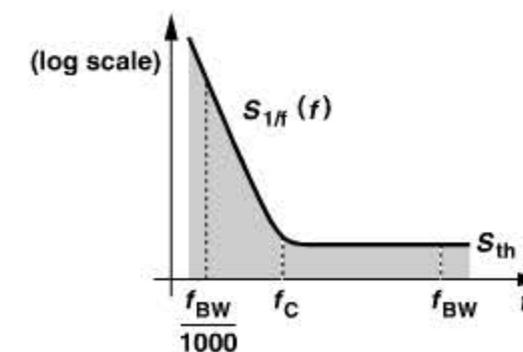


Figure 4.44 Spectrum for calculation of flicker noise.

noise, i.e., the additional noise power contributed by $S_{1/f}$. To this end, we note that if $S_{1/f} = \alpha/f$, then at f_c ,

$$\frac{\alpha}{f_c} = S_{th}. \quad (4.41)$$

That is, $\alpha = f_c \cdot S_{th}$. Also, we assume noise components below roughly $f_{BW}/1000$ are unimportant because they vary so slowly that they negligibly affect the baseband symbols.¹⁶ The total noise power from $f_{BW}/1000$ to f_{BW} is equal to

$$P_{n1} = \int_{f_{BW}/1000}^{f_c} \frac{\alpha}{f} df + (f_{BW} - f_c)S_{th} \quad (4.42)$$

$$= \alpha \ln \frac{1000f_c}{f_{BW}} + (f_{BW} - f_c)S_{th} \quad (4.43)$$

$$= \left(6.9 + \ln \frac{f_c}{f_{BW}}\right) f_c S_{th} + (f_{BW} - f_c)S_{th} \quad (4.44)$$

$$= \left(5.9 + \ln \frac{f_c}{f_{BW}}\right) f_c S_{th} + f_{BW} S_{th}. \quad (4.45)$$

In the absence of flicker noise, the total noise power from $f_{BW}/1000$ to f_{BW} is given by

$$P_{n2} \approx f_{BW} S_{th}. \quad (4.46)$$

The ratio of P_{n1} and P_{n2} can serve as a measure of the flicker noise penalty:

$$\frac{P_{n1}}{P_{n2}} = 1 + \left(5.9 + \ln \frac{f_c}{f_{BW}}\right) \frac{f_c}{f_{BW}}. \quad (4.47)$$

Example 4.23

An 802.11g receiver exhibits a baseband flicker noise corner frequency of 200 kHz. Determine the flicker noise penalty.

Solution:

We have $f_{BW} = 10$ MHz, $f_c = 200$ kHz, and hence

$$\frac{P_{n1}}{P_{n2}} = 1.04. \quad (4.48)$$

How do the above results depend on the gain of the LNA/mixer cascade? In a good design, the thermal noise at the end of the baseband chain arises mostly from the noise of

16. As an extreme example, a noise component with a period of one day varies so slowly that it has negligible effect on a 20-minute phone conversation.

the antenna, the LNA, and the mixer. Thus, a higher front-end gain directly raises S_{th} in Fig. 4.44, thereby lowering the value of f_c and hence the flicker noise penalty.

Example 4.24

A GSM receiver exhibits a baseband flicker noise corner frequency of 200 kHz. Determine the flicker noise penalty.

Solution:

Figure 4.45 plots the baseband spectra, implying that the noise must be integrated up to

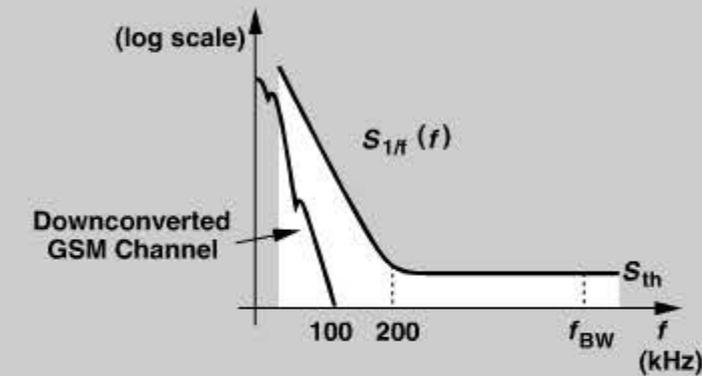


Figure 4.45 Effect of flicker noise on a GSM channel.

100 kHz. Assuming a lower end equal to about 1/1000 of the bit rate, we write the total noise as

$$P_{n1} = \int_{27 \text{ Hz}}^{100 \text{ kHz}} \frac{\alpha}{f} df \quad (4.49)$$

$$= f_c \cdot S_{th} \ln \frac{100 \text{ kHz}}{27 \text{ Hz}} \quad (4.50)$$

$$= 8.2 f_c S_{th}. \quad (4.51)$$

Without flicker noise,

$$P_{n2} \approx (100 \text{ kHz}) S_{th}. \quad (4.52)$$

That is, the penalty reaches

$$\frac{P_{n1}}{P_{n2}} = \frac{8.2 f_c}{100 \text{ kHz}} \quad (4.53)$$

$$= 16.4. \quad (4.54)$$

As expected, the penalty is much more severe in this case than in the 802.11g receiver of Example 4.23.

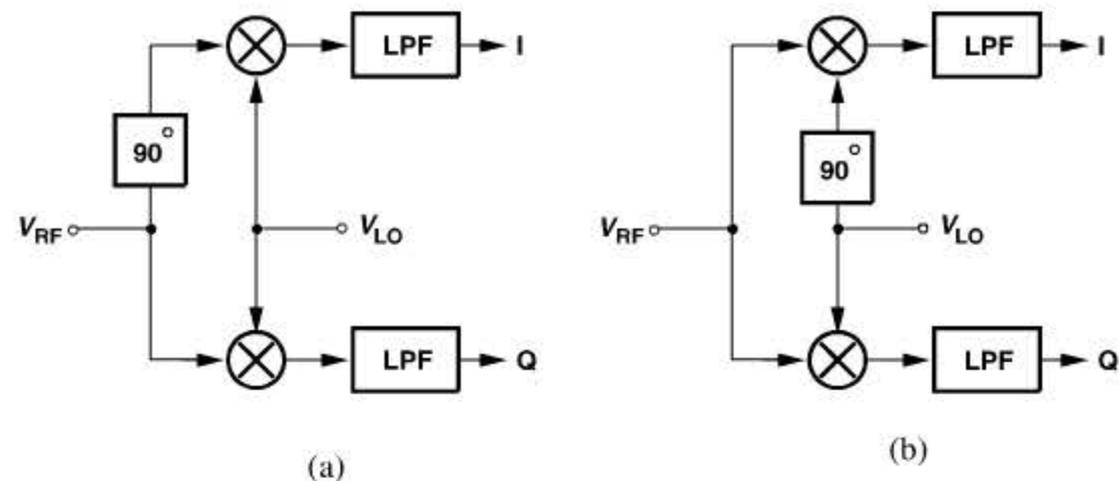


Figure 4.46 Shift of (a) RF signal, or (b) LO waveform by 90° .

As evident from the above example, the problem of flicker noise makes it difficult to employ direct conversion for standards that have a narrow channel bandwidth. In such cases, the “low-IF” architecture proves a more viable choice (Section 4.2.5).

I/Q Mismatch As explained in Section 4.2.2, downconversion of an asymmetrically-modulated signal to a zero IF requires separation into quadrature phases. This can be accomplished by shifting either the RF signal or the LO waveform by 90° (Fig. 4.46). Since shifting the RF signal generally entails severe noise-power-gain trade-offs, the approach in Fig. 4.46(b) is preferred. In either case, as illustrated in Fig. 4.47, errors in the 90° phase shift circuit and mismatches between the quadrature mixers result in imbalances in the amplitudes and phases of the baseband I and Q outputs. The baseband stages themselves may also contribute significant gain and phase mismatches.¹⁷

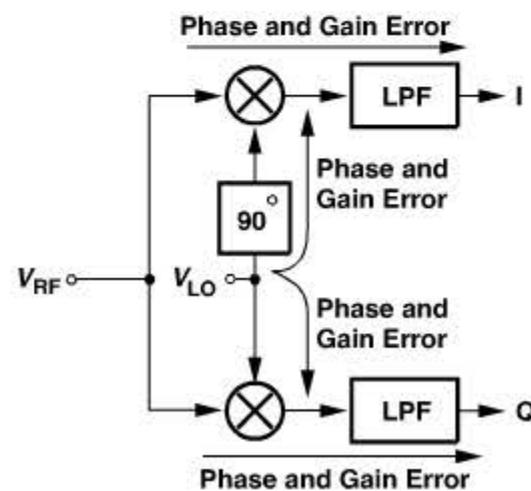


Figure 4.47 Sources of I and Q mismatch.

17. We use the terms “amplitude mismatch” and “gain mismatch” interchangeably.

Quadrature mismatches tend to be larger in direct-conversion receivers than in heterodyne topologies. This occurs because (1) the propagation of a higher frequency (f_{in}) through quadrature mixers experiences greater mismatches; for example, a delay mismatch of 10 ps between the two mixers translates to a phase mismatch of 18° at 5 GHz [Fig. 4.48(a)] and 3.6° at 1 GHz [Fig. 4.48(b)]; or (2) the quadrature phases of the LO itself suffer from greater mismatches at higher frequencies; for example, since device dimensions are reduced to achieve higher speeds, the mismatches between transistors increase.

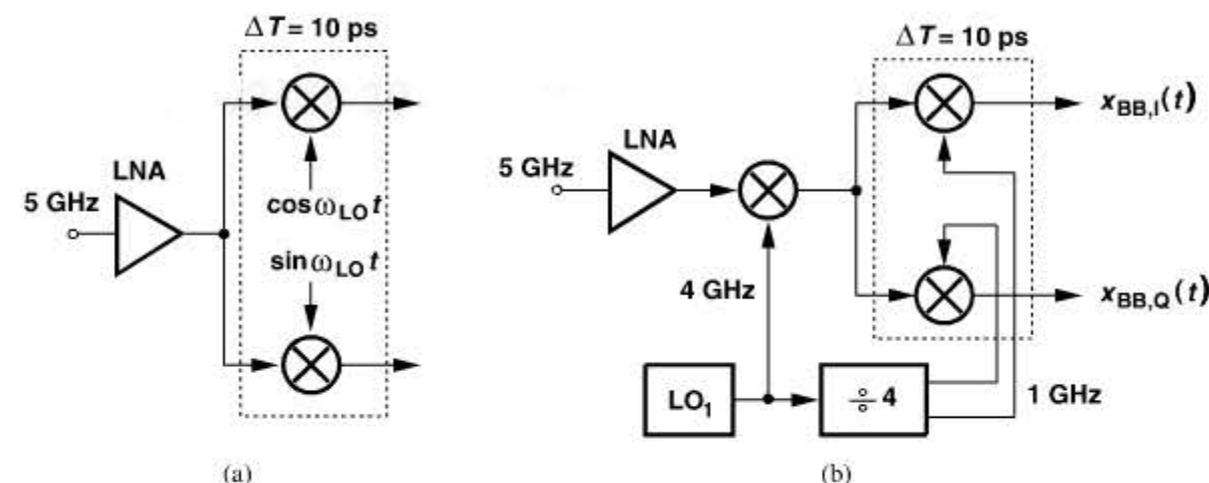


Figure 4.48 Effect of a 10-ps propagation mismatch on a (a) direct-conversion and (b) heterodyne receiver.

To gain insight into the effect of I/Q imbalance, consider a QPSK signal, $x_{in}(t) = a \cos \omega_c t + b \sin \omega_c t$, where a and b are either -1 or $+1$. Now let us lump all of the gain and phase mismatches shown in Fig. 4.47 in the LO path (Fig. 4.49)

$$x_{LO,I}(t) = 2 \left(1 + \frac{\epsilon}{2} \right) \cos \left(\omega_c t + \frac{\theta}{2} \right) \quad (4.55)$$

$$x_{LO,Q}(t) = 2 \left(1 - \frac{\epsilon}{2} \right) \sin \left(\omega_c t - \frac{\theta}{2} \right), \quad (4.56)$$

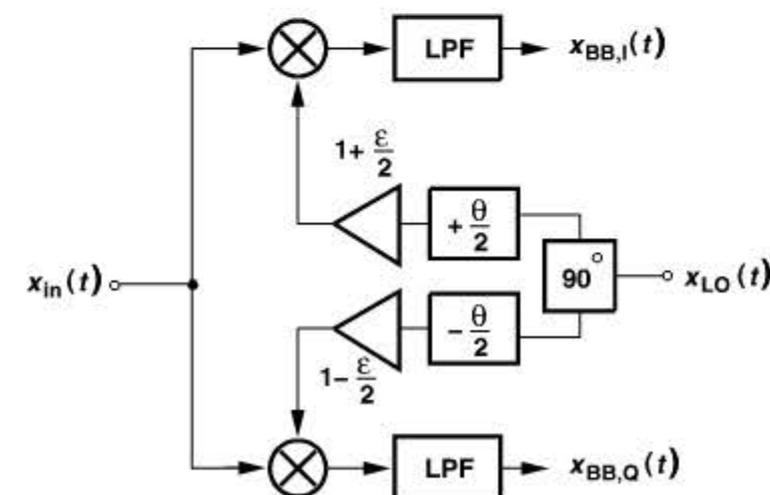


Figure 4.49 Mismatches lumped in LO path.

where the factor of 2 is included to simplify the results and ϵ and θ represent the amplitude and phase mismatches, respectively. Multiplying $x_{in}(t)$ by the quadrature LO waveforms and low-pass filtering the results, we obtain the following baseband signals:

$$x_{BB,I}(t) = a \left(1 + \frac{\epsilon}{2}\right) \cos \frac{\theta}{2} - b \left(1 + \frac{\epsilon}{2}\right) \sin \frac{\theta}{2} \quad (4.57)$$

$$x_{BB,Q}(t) = -a \left(1 - \frac{\epsilon}{2}\right) \sin \frac{\theta}{2} + b \left(1 - \frac{\epsilon}{2}\right) \cos \frac{\theta}{2}. \quad (4.58)$$

We now examine the results for two special cases: $\epsilon \neq 0, \theta = 0$ and $\epsilon = 0, \theta \neq 0$. In the former case, $x_{BB,I}(t) = a(1 + \epsilon/2)$ and $x_{BB,Q}(t) = b(1 - \epsilon/2)$, implying that the quadrature baseband symbols are scaled differently in amplitude [Fig. 4.50(a)]. More importantly, the points in the constellation are displaced [Fig. 4.50(b)].

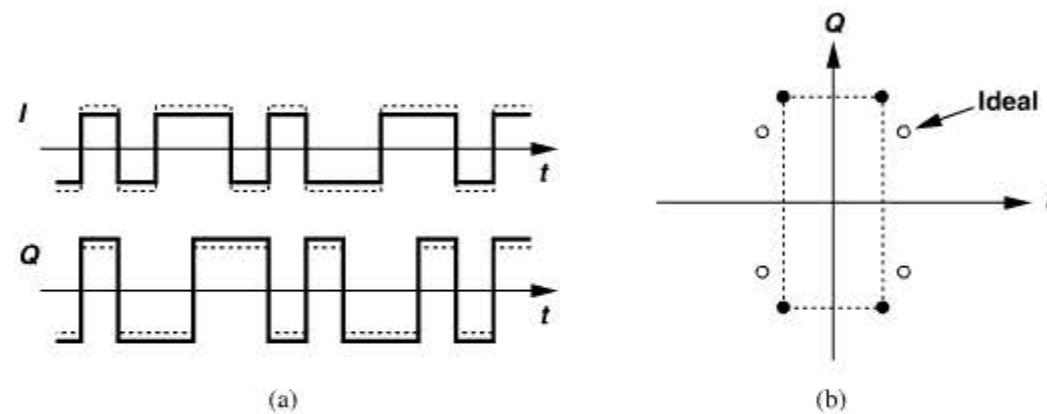


Figure 4.50 Effect of gain mismatch on (a) time-domain waveforms and (b) constellation of a QPSK signal.

With $\epsilon = 0, \theta \neq 0$, we have $x_{BB,I}(t) = a \cos(\theta/2) - b \sin(\theta/2)$ and $x_{BB,Q}(t) = -a \sin(\theta/2) + b \cos(\theta/2)$. That is, each baseband output is corrupted by a fraction of the data symbols in the *other* output [Fig. 4.51(a)]. Also, the constellation is compressed along one diagonal and stretched along the other [Fig. 4.51(b)].

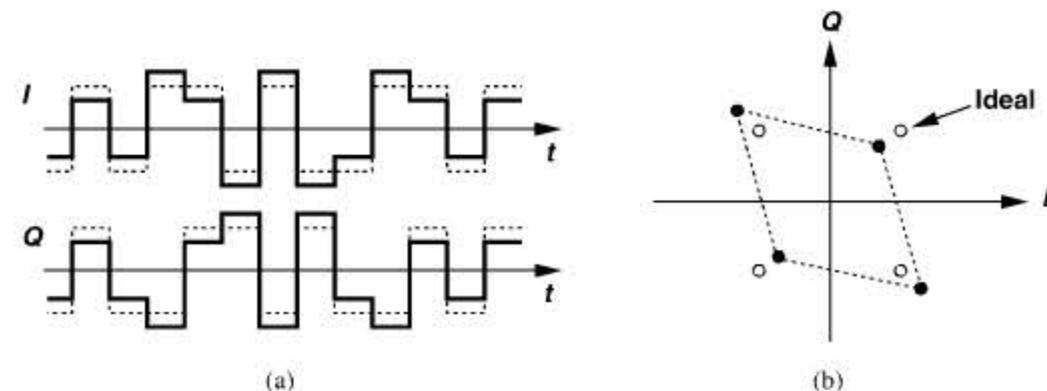


Figure 4.51 Effect of phase mismatch on (a) time-domain waveforms and (b) constellation of a QPSK signal.

Example 4.25

An FSK signal is applied to a direct-conversion receiver. Plot the baseband waveforms and determine the effect of I/Q mismatch.

Solution:

We express the FSK signal as $x_{FSK}(t) = A_0 \cos[(\omega_c + a\omega_1)t]$, where $a = \pm 1$ represents the binary information; i.e., the frequency of the carrier swings by $+\omega_1$ or $-\omega_1$. Upon multiplication by the quadrature phases of the LO, the signal produces the following baseband components:

$$x_{BB,I}(t) = -A_1 \cos a\omega_1 t \quad (4.59)$$

$$x_{BB,Q}(t) = +A_1 \sin a\omega_1 t. \quad (4.60)$$

Figure 4.52(a) illustrates the results: if the carrier frequency is equal to $\omega_c + \omega_1$ (i.e., $a = +1$), then the rising edges of $x_{BB,I}(t)$ coincide with the positive peaks of $x_{BB,Q}(t)$.

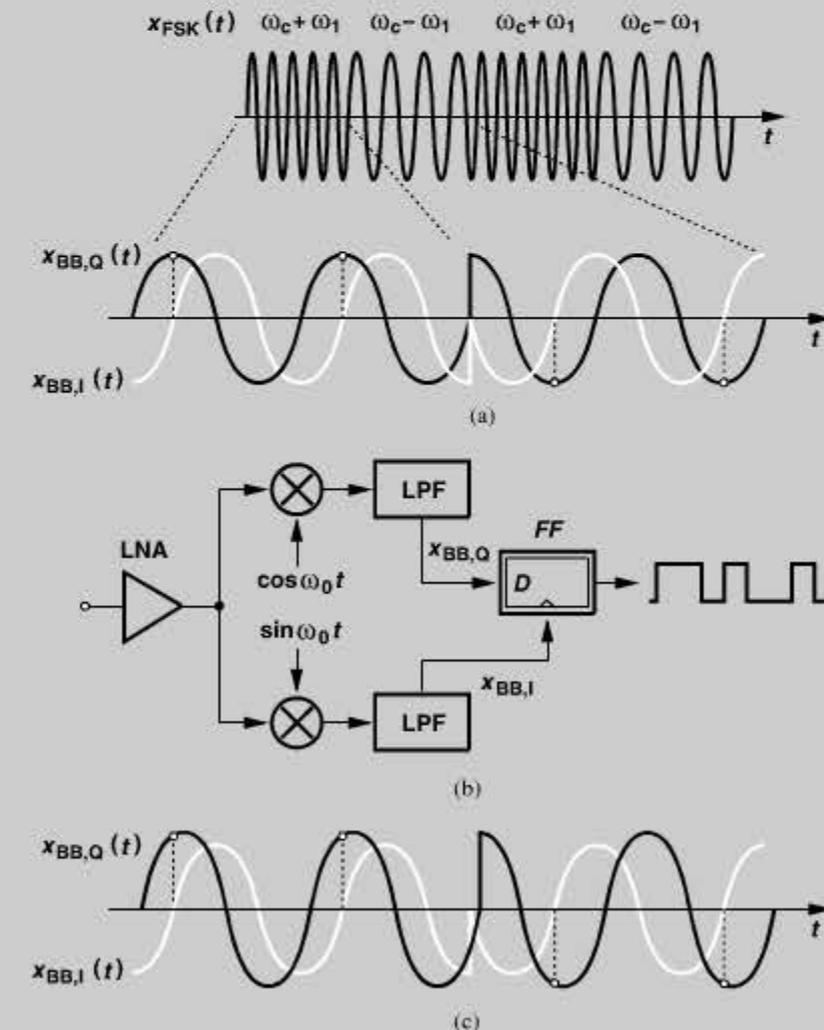


Figure 4.52 (a) Baseband waveforms for an FSK signal, (b) FSK detection by a D flipflop, (c) effect of phase and gain mismatches.

(Continues)

Example 4.25 (Continued)

Conversely, if the carrier frequency is equal to $\omega_c - \omega_1$, then the rising edges of $x_{BB,I}(t)$ coincide with the *negative* peaks of $x_{BB,Q}(t)$. Thus, the binary information is detected if $x_{BB,I}(t)$ simply *samples* $x_{BB,Q}(t)$, e.g., by means of a D flipflop [Fig. 4.52(b)].

The waveforms of Fig. 4.52(a) and the detection method of Fig. 4.52(b) suggest that FSK can tolerate large I/Q mismatches [Fig. 4.52(c)]: amplitude mismatch proves benign so long as the smaller output does not suffer from degraded SNR, and phase mismatch is tolerable so long as $x_{BB,I}(t)$ samples the correct *polarity* of $x_{BB,Q}(t)$. Of course, as the phase mismatch approaches 90° , the additive noise in the receive chain introduces errors.

In the design of an RF receiver, the maximum tolerable I/Q mismatch must be known so that the architecture and the building blocks are chosen accordingly. For complex signal waveforms such as OFDM with QAM, this maximum can be obtained by simulations: the bit error rate is plotted for different combinations of gain and phase mismatches, providing the maximum mismatch values that affect the performance negligibly. (The EVM can also reflect the effect of these mismatches.) As an example, Fig. 4.53 plots the BER curves for a system employing OFDM with 128 subchannels and QPSK modulation in each subchannel [1]. We observe that gain/phase mismatches below $-0.6 \text{ dB}/6^\circ$ have negligible effect.

In standards such as 802.11a/g, the required phase and gain mismatches are so small that the “raw” matching of the devices and the layout may not suffice. Consequently, in

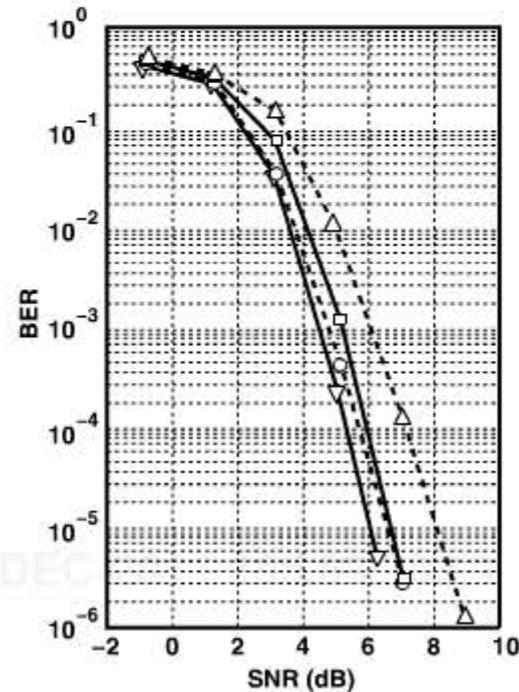


Figure 4.53 Effect of I/Q mismatch on an OFDM signal with QPSK modulation. (∇ : no imbalance; \square : $\theta = 6^\circ$, $\epsilon = 0.6 \text{ dB}$; \triangle : $\theta = 16^\circ$, $\epsilon = 1.4 \text{ dB}$.)

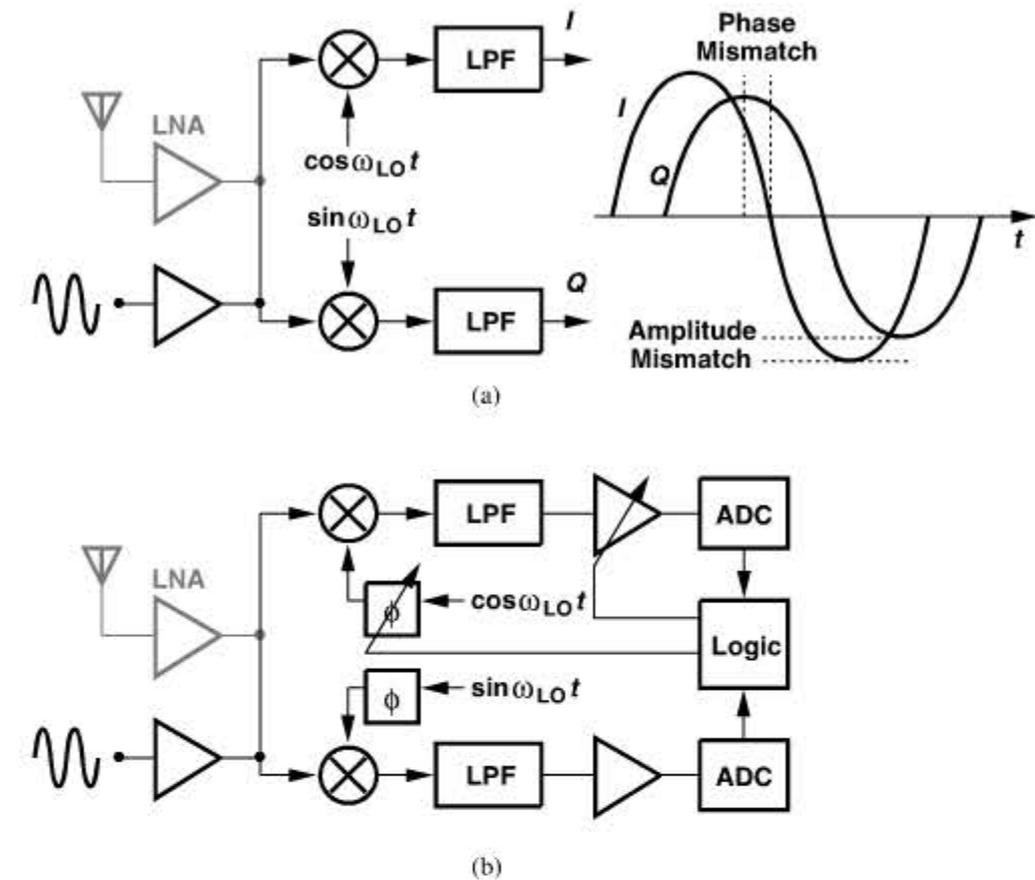


Figure 4.54 (a) Computation and (b) correction of I/Q mismatch in a direct-conversion receiver.

many high-performance systems, the quadrature phase and gain must be calibrated—either at power-up or continuously. As illustrated in Fig. 4.54(a), calibration at power-up can be performed by applying an RF tone at the input of the quadrature mixers and observing the baseband sinusoids in the analog or digital domain [2]. Since these sinusoids can be produced at arbitrarily low frequencies, their amplitude and phase mismatches can be measured accurately. With the mismatches known, the received signal constellation is corrected before detection. Alternatively, as depicted in Fig. 4.54(b), a variable-phase stage, ϕ , and a variable-gain stage can be inserted in the LO and baseband paths, respectively, and adjusted until the mismatches are sufficiently small. Note that the adjustment controls must be stored digitally during the actual operation of the receiver.

Mixing Spurs Unlike heterodyne systems, direct-conversion receivers rarely encounter corruption by mixing spurs. This is because, for an input frequency f_1 to fall in the baseband after experiencing mixing with nf_{LO} , we must have $f_1 \approx nf_{LO}$. Since f_{LO} is equal to the desired channel frequency, f_1 lies far from the band of interest and is greatly suppressed by the selectivity of the antenna, the band-select filter, and the LNA.

The issue of LO harmonics does manifest itself if the receiver is designed for a wide frequency band (greater than two octaves). Examples include TV tuners, “software-defined radios,” and “cognitive radios.”

4.2.4 Image-Reject Receivers

Our study of heterodyne and direct-conversion receivers has revealed various pros and cons. For example, heterodyning must deal with the image and mixing spurs and direct conversion, with even-order distortion and flicker noise. “Image-reject” architectures are another class of receivers that suppress the image without filtering, thereby avoiding the trade-off between image rejection and channel selection.

90° Phase Shift Before studying these architectures, we must define a “shift-by-90°” operation. First, let us consider a tone, $A \cos \omega_c t = (A/2)[\exp(+j\omega_c t) + \exp(-j\omega_c t)]$. The two exponentials respectively correspond to impulses at $+\omega_c$ and $-\omega_c$ in the frequency domain. We now shift the waveform by 90°:

$$A \cos(\omega_c t - 90^\circ) = A \frac{e^{+j(\omega_c t - 90^\circ)} + e^{-j(\omega_c t - 90^\circ)}}{2} \quad (4.61)$$

$$= -\frac{A}{2} j e^{+j\omega_c t} + \frac{A}{2} j e^{-j\omega_c t} \quad (4.62)$$

$$= A \sin \omega_c t. \quad (4.63)$$

Equivalently, the impulse at $+\omega_c$ is multiplied by $-j$ and that at $-\omega_c$, by $+j$. We illustrate this transformation in the three-dimensional diagram of Fig. 4.55(a), recognizing that the impulse at $+\omega_c$ is rotated clockwise and that at $-\omega_c$ counterclockwise.

Similarly, for a narrowband modulated signal, $x(t) = A(t) \cos[\omega_c t + \phi(t)]$, we perform a 90° phase shift as

$$A(t) \cos[\omega_c t + \phi(t) - 90^\circ] = A(t) \frac{e^{+j[\omega_c t + \phi(t) - 90^\circ]} + e^{-j[\omega_c t + \phi(t) - 90^\circ]}}{2} \quad (4.64)$$

$$= A(t) \frac{-j e^{+j[\omega_c t + \phi(t)]} + j e^{-j[\omega_c t + \phi(t)]}}{2} \quad (4.65)$$

$$= A(t) \sin[\omega_c t + \phi(t)]. \quad (4.66)$$

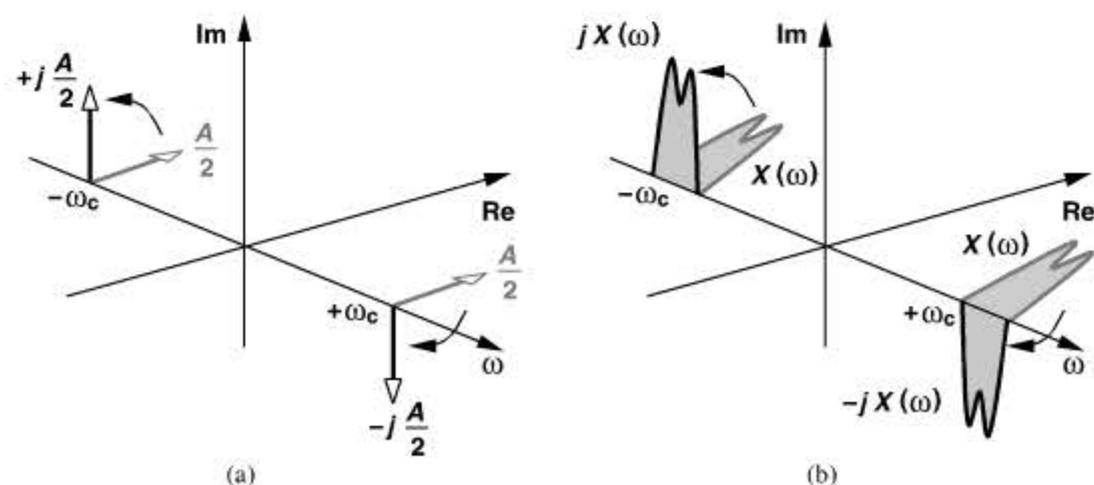


Figure 4.55 Illustration of 90° phase shift for (a) a cosine and (b) a modulated signal.

As depicted in Fig. 4.55(b), the positive-frequency contents are multiplied by $-j$ and the negative-frequency contents by $+j$ (if ω_c is positive). Alternatively, we write in the frequency domain:

$$X_{90^\circ}(\omega) = X(\omega)[-j \operatorname{sgn}(\omega)], \quad (4.67)$$

where $\operatorname{sgn}(\omega)$ denotes the signum (sign) function. The shift-by-90° operation is also called the “Hilbert transform.” The reader can prove that the Hilbert transform of the Hilbert transform (i.e., the cascade of two 90° phase shifts) simply negates the original signal.

Example 4.26

In phasor diagrams, we simply multiply a phasor by $-j$ to rotate it by 90° clockwise. Is that inconsistent with the Hilbert transform?

Solution:

No, it is not. A phasor is a representation of $A \exp(j\omega_c t)$, i.e., only the positive frequency content. That is, we implicitly assume that if $A \exp(j\omega_c t)$ is multiplied by $-j$, then $A \exp(-j\omega_c t)$ is also multiplied by $+j$.

The Hilbert transform, as expressed by Eq. (4.67), *distinguishes* between negative and positive frequencies. This distinction is the key to image rejection.

Example 4.27

Plot the spectrum of $A \cos \omega_c t + jA \sin \omega_c t$.

Solution:

Multiplication of the spectrum of $A \sin \omega_c t$ by j rotates both impulses by 90° counterclockwise [Fig. 4.56(a)]. Upon adding this spectrum to that $A \cos \omega_c t$, we obtain the

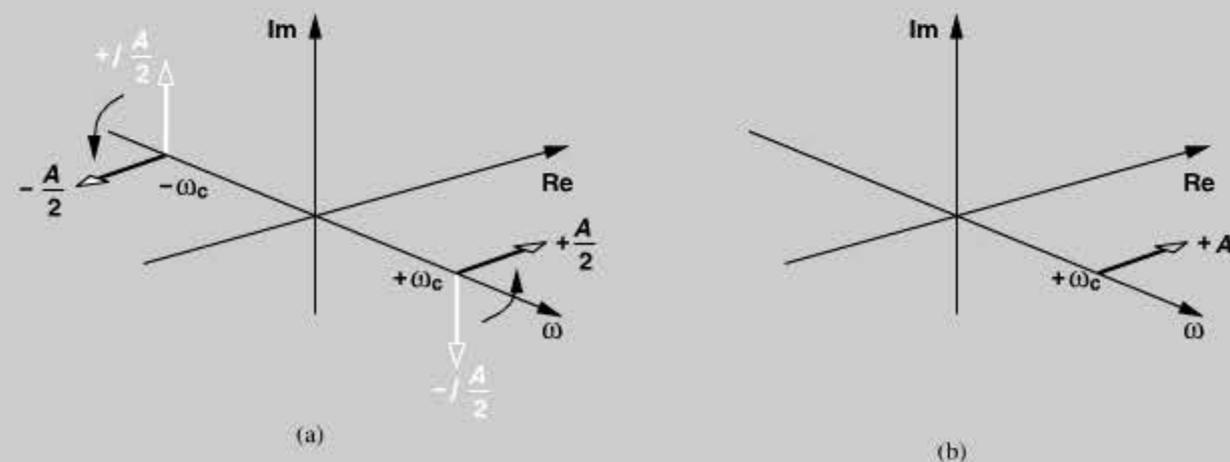


Figure 4.56 (a) A sine subjected to 90° phase shift, (b) spectrum of $A \cos \omega_c t + j \sin \omega_c t$.

(Continues)

Example 4.27 (Continued)

one-sided spectrum shown in Fig. 4.56(b). This is, of course, to be expected because $A \cos \omega_c t + jA \sin \omega_c t = A \exp(-j\omega_c t)$, whose Fourier transform is a single impulse located at $\omega = +\omega_c$.

Example 4.28

A narrowband signal $I(t)$ with a real spectrum is shifted by 90° to produce $Q(t)$. Plot the spectrum of $I(t) + jQ(t)$.¹⁸

Solution:

We first multiply $I(\omega)$ by $-j\text{sgn}(\omega)$ [Fig. 4.57(a)] and then, in a manner similar to the previous example, multiply the result by j [Fig. 4.57(b)]. The spectrum of $jQ(t)$ therefore cancels that of $I(t)$ at negative frequencies and enhances it at positive frequencies [Fig. 4.57(c)]. The one-sided spectrum of $I(t) + jQ(t)$ proves useful in the analysis of transceivers.

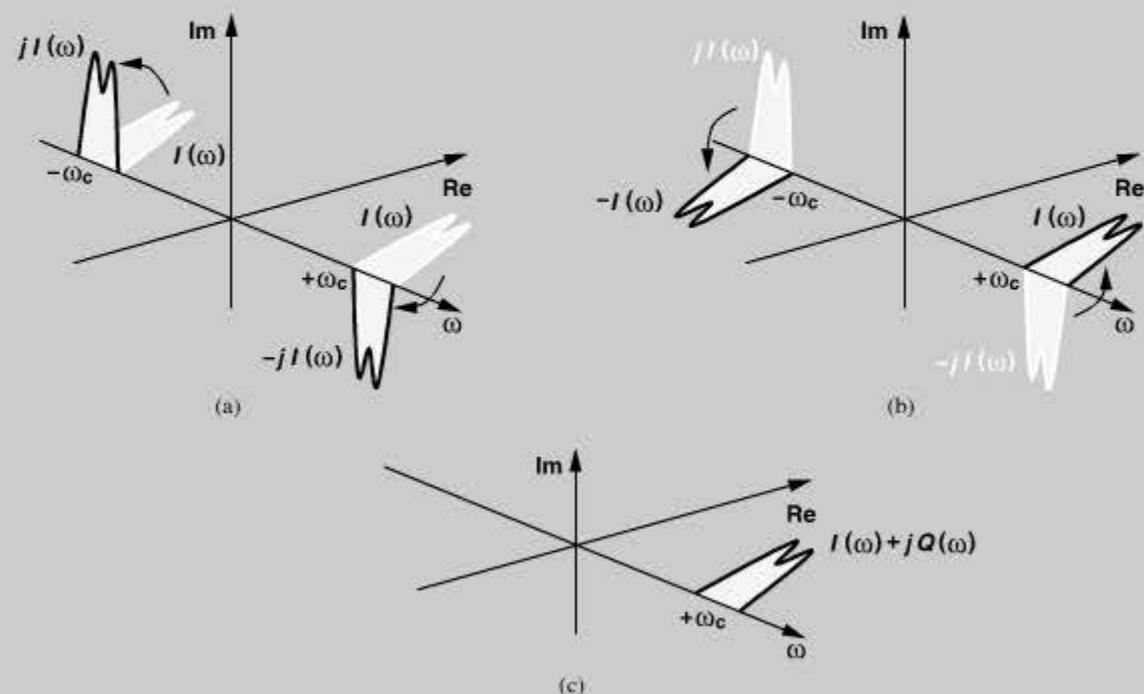


Figure 4.57 (a) 90° phase shift applied to I to produce Q , (b) multiplication of the result by j , (c) analytic signal.

18. This sum is called the “analytic signal” of $I(t)$.

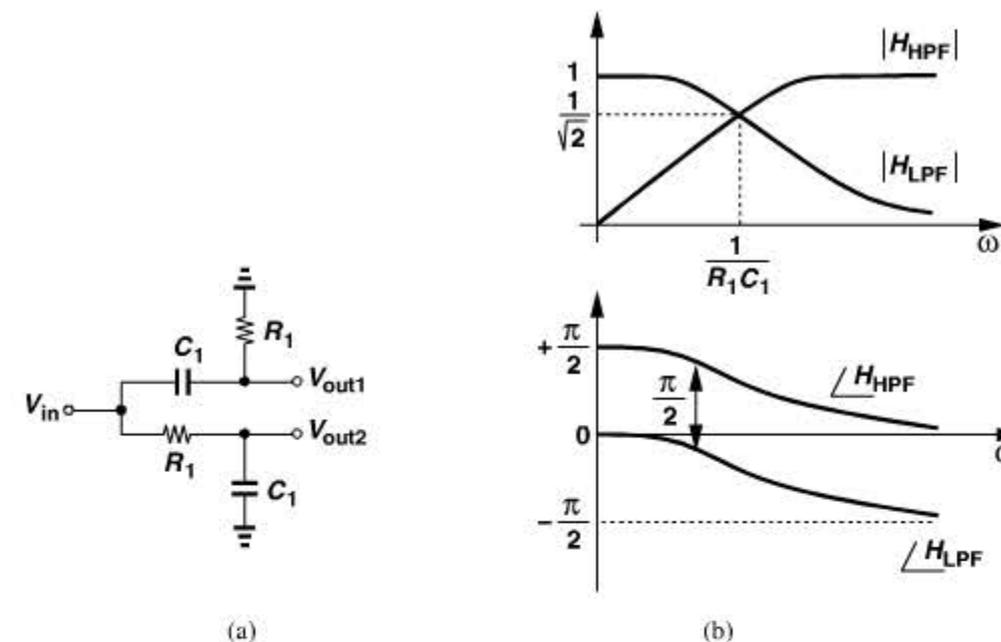


Figure 4.58 (a) Use of an RC-CR network to perform a 90° phase shift, (b) frequency response of the network.

How is the 90° phase shift implemented? Consider the RC-CR network shown in Fig. 4.58(a), where the high-pass and low-pass transfer functions are respectively given by

$$H_{HPF}(s) = \frac{V_{out1}}{V_{in}} = \frac{R_1 C_1 s}{R_1 C_1 s + 1} \quad (4.68)$$

$$H_{LPF}(s) = \frac{V_{out2}}{V_{in}} = \frac{1}{R_1 C_1 s + 1}. \quad (4.69)$$

The transfer functions exhibit a phase of $\angle H_{HPF} = \pi/2 - \tan^{-1}(R_1 C_1 \omega)$ and $\angle H_{LPF} = -\tan^{-1}(R_1 C_1 \omega)$. Thus, $\angle H_{HPF} - \angle H_{LPF} = \pi/2$ at all frequencies and any choice of R_1 and C_1 . Also, $|V_{out1}/V_{in}| = |V_{out2}/V_{in}| = 1/\sqrt{2}$ at $\omega = (R_1 C_1)^{-1}$ [Fig. 4.58(b)]. We can therefore consider V_{out2} as the Hilbert transform of V_{out1} at frequencies close to $(R_1 C_1)^{-1}$.

Another approach to realizing the 90° -phase-shift operation is illustrated in Fig. 4.59(a), where the RF input is mixed with the quadrature phases of the LO so as to translate the spectrum to a nonzero IF. As shown in Fig. 4.59(b), as a result of mixing with $\cos \omega_{LO} t$, the impulse at $-\omega_{LO}$ is convolved with the input spectrum around $+\omega_c$, generating that at $-\omega_{IF}$. Similarly, the impulse at $+\omega_{LO}$ produces the spectrum at $+\omega_{IF}$ from that at $-\omega_c$. Depicted in Fig. 4.59(c), mixing with $\sin \omega_{LO} t$ results in an IF spectrum at $-\omega_{IF}$ with a coefficient $+j/2$ and another at $+\omega_{IF}$ with a coefficient $-j/2$. We observe that, indeed, the IF spectrum emerging from the lower arm is the Hilbert transform of that from the upper arm.

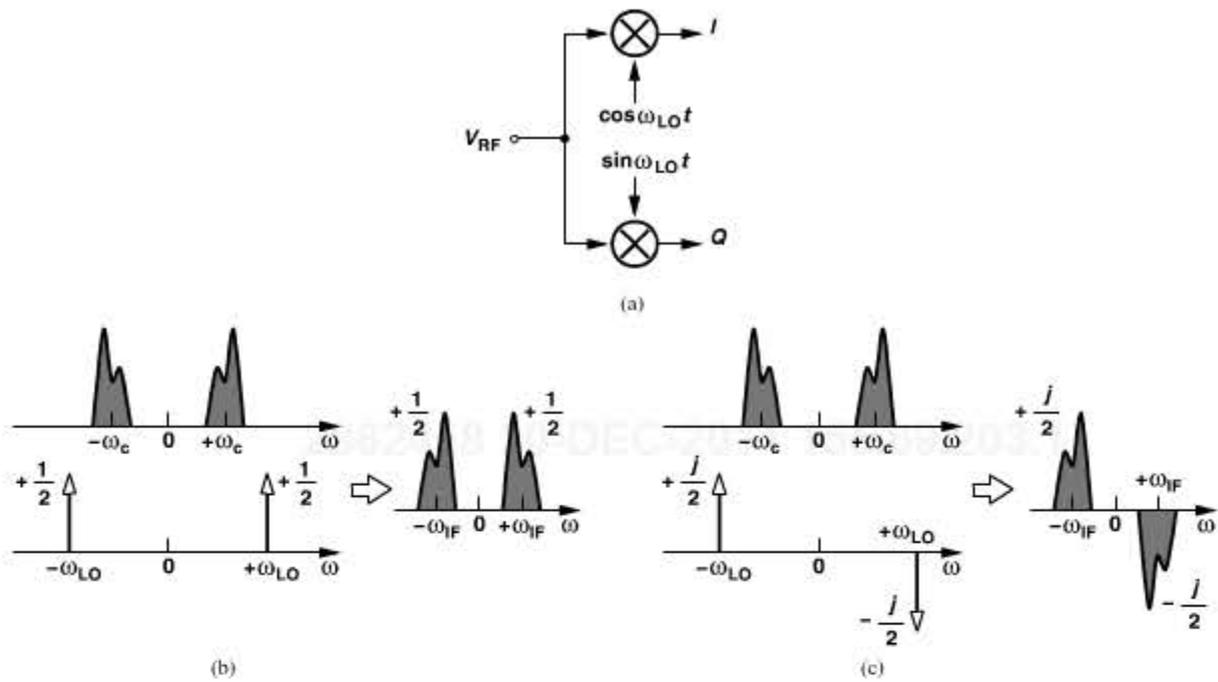


Figure 4.59 (a) Quadrature downconversion as a 90° phase shifter, (b) output spectrum resulting from multiplication by $\cos \omega_{LO}t$, (c) output spectrum resulting from multiplication by $\sin \omega_{LO}t$.

Example 4.29

The realization of Fig. 4.59(a) assumes high-side injection for the LO. Repeat the analysis for low-side injection.

Solution:

Figures 4.60(a) and (b) show the spectra for mixing with $\cos \omega_{LO}t$ and $\sin \omega_{LO}t$, respectively. In this case, the IF component in the lower arm is the *negative* of the Hilbert transform of that in the upper arm.

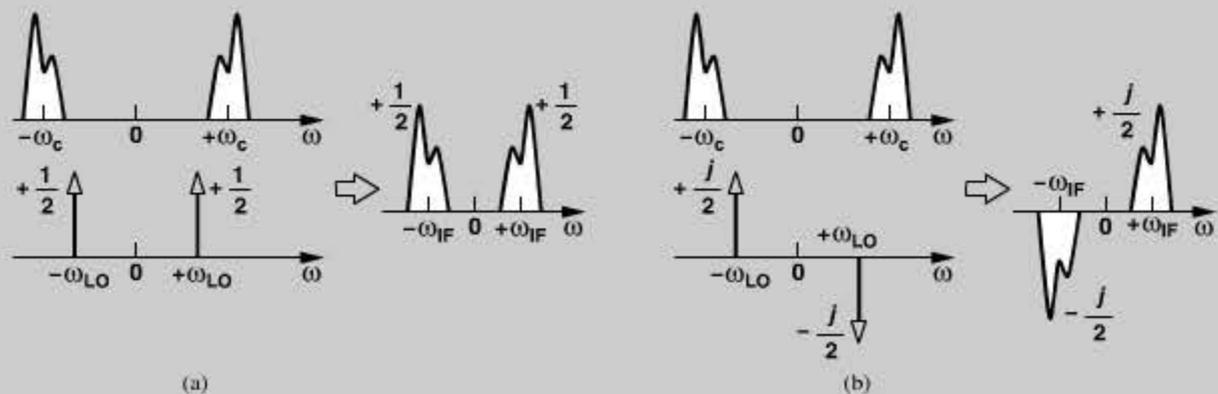


Figure 4.60 Low-side-injection mixing of an RF signal with (a) $\cos \omega_{LO}t$ and (b) $\sin \omega_{LO}t$.

Let us summarize our findings thus far. The quadrature converter¹⁹ of Fig. 4.59(a) produces at its output a signal and its Hilbert transform if $\omega_c > \omega_{LO}$ or a signal and the negative of its Hilbert transform if $\omega_c < \omega_{LO}$. This arrangement therefore distinguishes between the desired signal and its image. Figure 4.61 depicts the three-dimensional IF spectra if a signal and its image are applied at the input and $\omega_{LO} < \omega_c$.

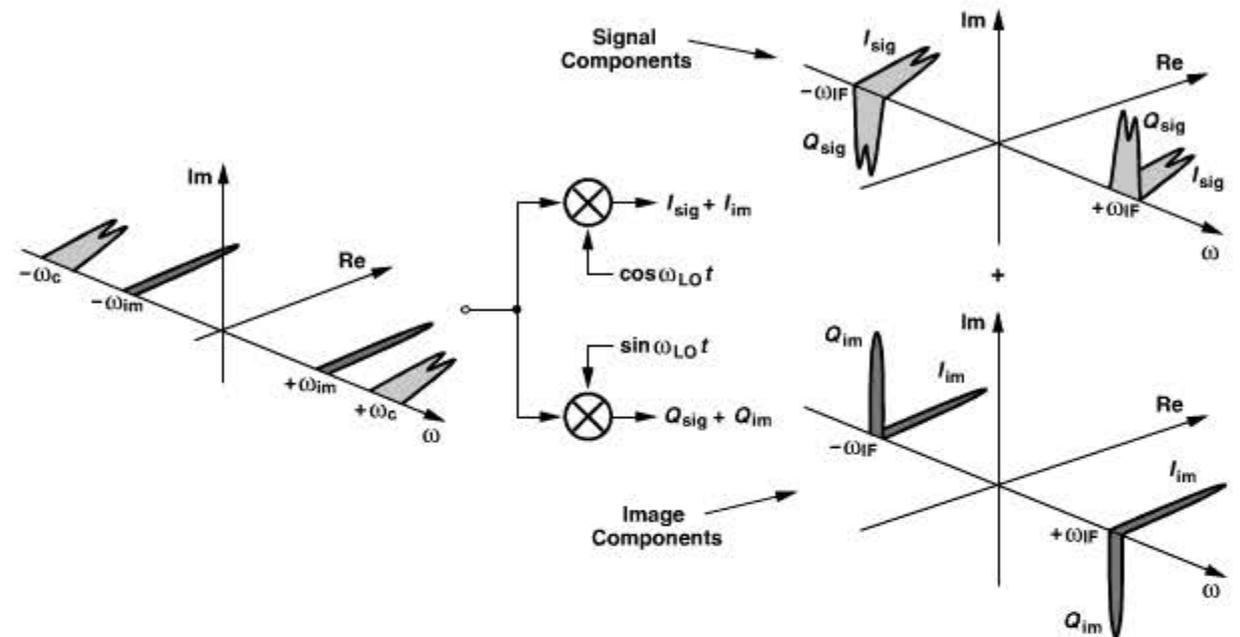


Figure 4.61 Input and output spectra in a quadrature downconverter with low-side injection.

Hartley Architecture How can the image components in Fig. 4.61 cancel each other? For example, is $I(t) + Q(t)$ free from the image? Since the image components in $Q(t)$ are 90° out of phase with respect to those in $I(t)$, this summation still contains the image. However, since the Hilbert transform of the Hilbert transform negates the signal, if we shift $I(t)$ or $Q(t)$ by another 90° before adding them, the image may be removed. This hypothesis forms the foundation for the Hartley architecture shown in Fig. 4.62. (The original idea proposed

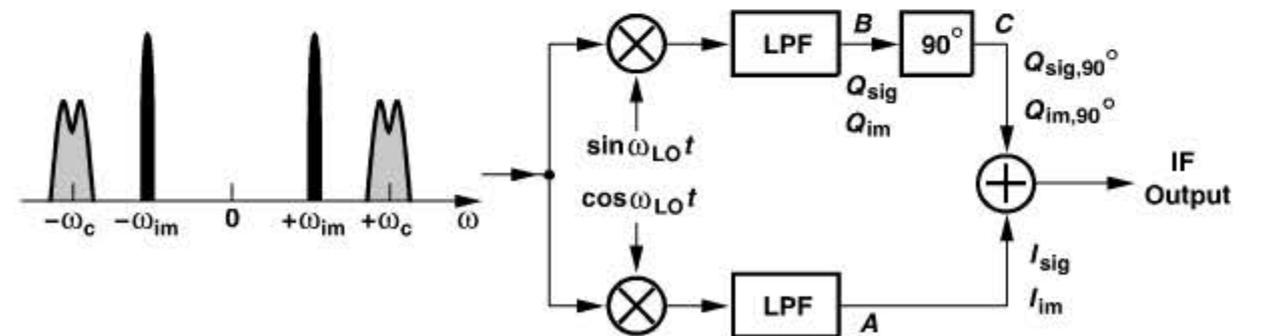


Figure 4.62 Hartley image-reject receiver.

19. We can also consider this a quadrature downconverter if $\omega_{IF} < \omega_c$. In Problem 4.14, we study the case $\omega_{IF} > \omega_c$.

by Hartley relates to single-sideband transmitters [4].) The low-pass filters are inserted to remove the unwanted high-frequency components generated by the mixers.

To understand the operation of Hartley's architecture, we assume low-side injection and apply a 90° phase shift to the Hilbert transforms of the signal and the image (the Q arm) in Fig. 4.61, obtaining $Q_{sig,90^\circ}$ and $Q_{im,90^\circ}$ as shown in Fig. 4.63. Multiplication of Q_{sig} by $-j\text{sgn}(\omega)$ rotates and superimposes the spectrum of Q_{sig} on that of I_{sig} (from Fig. 4.61), doubling the signal amplitude. On the other hand, multiplication of Q_{im} by $-j\text{sgn}(\omega)$ creates the opposite of I_{im} , cancelling the image.

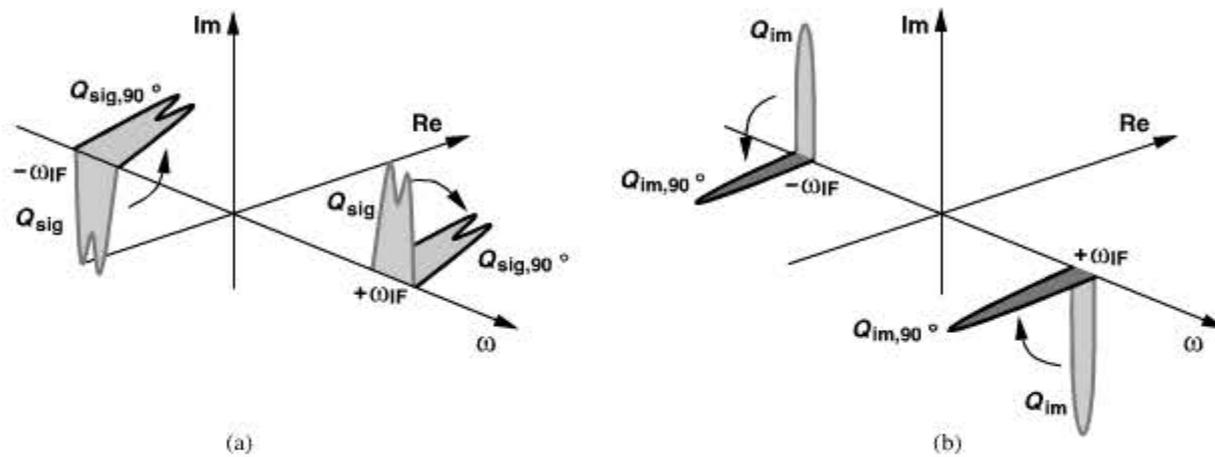


Figure 4.63 Spectra at points B and C in Hartley receiver.

In summary, the Hartley architecture first takes the negative Hilbert transform of the signal and the Hilbert transform of the image (or vice versa) by means of quadrature mixing, subsequently takes the Hilbert transform of one of the downconverted outputs, and sums the results. That is, the signal spectrum is multiplied by $[+j\text{sgn}(\omega)][-j\text{sgn}(\omega)] = +1$, whereas the image spectrum is multiplied by $[-j\text{sgn}(\omega)][-j\text{sgn}(\omega)] = -1$.

Example 4.30

An eager student constructs the Hartley architecture but with high-side injection. Explain what happens.

Solution:

From Fig. 4.60, we note that the quadrature converter takes the Hilbert transform of the signal and the negative Hilbert transform of the image. Thus, with another 90° phase shift, the outputs C and A in Fig. 4.62 contain the signal with *opposite* polarities and the image with the same polarity. The circuit therefore operates as a “signal-reject” receiver! Of course, the design is salvaged if the addition is replaced with subtraction.

The behavior of the Hartley architecture can also be expressed analytically. Let us represent the received signal and image as $x(t) = A_{sig} \cos(\omega_c t + \phi_{sig}) + A_{im} \cos(\omega_{im} t + \phi_{im})$, where the amplitudes and phases are functions of time in the general case. Multiplying $x(t)$ by the LO phases and neglecting the high-frequency components, we obtain the signals at

points A and B in Fig. 4.62:

$$x_A(t) = \frac{A_{sig}}{2} \cos[(\omega_c - \omega_{LO})t + \phi_{sig}] + \frac{A_{im}}{2} \cos[(\omega_{im} - \omega_{LO})t + \phi_{im}] \quad (4.70)$$

$$x_B(t) = -\frac{A_{sig}}{2} \sin[(\omega_c - \omega_{LO})t + \phi_{sig}] - \frac{A_{im}}{2} \sin[(\omega_{im} - \omega_{LO})t + \phi_{im}], \quad (4.71)$$

where a unity LO amplitude is assumed for simplicity. Now, $x_B(t)$ must be shifted by 90° . With low-side injection, the first sine has a positive frequency and becomes negative of a cosine after the 90° shift (why?). The second sine, on the other hand, has a negative frequency. We therefore write $-(A_{im}/2) \sin[(\omega_{im} - \omega_{LO})t + \phi_{im}] = (A_{im}/2) \sin[(\omega_{LO} - \omega_{im})t - \phi_{im}]$ so as to obtain a positive frequency and shift the result by 90° , arriving at $-(A_{im}/2) \cos[(\omega_{LO} - \omega_{im})t - \phi_{im}] = -(A_{im}/2) \cos[(\omega_{im} - \omega_{LO})t + \phi_{im}]$. It follows that

$$x_C(t) = \frac{A_{sig}}{2} \cos[(\omega_c - \omega_{LO})t + \phi_{sig}] - \frac{A_{im}}{2} \cos[(\omega_{im} - \omega_{LO})t + \phi_{im}]. \quad (4.72)$$

Upon addition of $x_A(t)$ and $x_C(t)$, we retain the signal and reject the image.

The 90° phase shift depicted in Fig. 4.62 is typically realized as a $+45^\circ$ shift in one path and -45° shift in the other (Fig. 4.64). This is because it is difficult to shift a single signal by 90° while circuit components vary with process and temperature.

The principal drawback of the Hartley architecture stems from its sensitivity to mismatches: the perfect image cancellation described above occurs only if the amplitude and phase of the negative of the image exactly match those of the image itself. If the LO phases are not in exact quadrature or the gains and phase shifts of the upper and lower arms in Fig. 4.64 are not identical, then a fraction of the image remains. To quantify this effect, we lump the mismatches of the receiver as a single amplitude error, ϵ , and phase error, $\Delta\theta$, in the LO path, i.e., one LO waveform is expressed as $\sin \omega_{LO} t$ and the other as $(1 + \epsilon) \cos(\omega_{LO} t + \Delta\theta)$. Expressing the received signal and image as $x(t) = A_{sig} \cos(\omega_c t + \phi_{sig}) + A_{im} \cos(\omega_{im} t + \phi_{im})$ and multiplying $x(t)$ by the LO waveforms, we write the downconverted signal at point A in Fig. 4.62 as

$$x_A(t) = \frac{A_{sig}}{2} (1 + \epsilon) \cos[(\omega_c - \omega_{LO})t + \phi_{sig} + \Delta\theta] + \frac{A_{im}}{2} (1 + \epsilon) \cos[(\omega_{im} - \omega_{LO})t + \phi_{im} + \Delta\theta]. \quad (4.73)$$

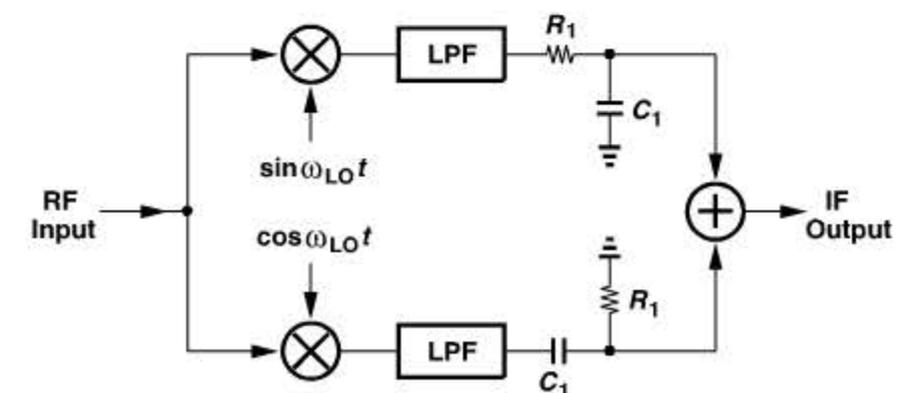


Figure 4.64 Realization of 90° phase shift in Hartley receiver.

The spectra at points B and C are still given by Eqs. (4.71) and (4.72), respectively. We now add $x_A(t)$ and $x_C(t)$ and group the signal and image components at the output:

$$x_{sig}(t) = \frac{A_{sig}}{2}(1 + \epsilon) \cos[(\omega_c - \omega_{LO})t + \phi_{sig} + \Delta\theta] + \frac{A_{sig}}{2} \cos[(\omega_c - \omega_{LO})t + \phi_{sig}] \quad (4.74)$$

$$x_{im}(t) = \frac{A_{im}}{2}(1 + \epsilon) \cos[(\omega_{im} - \omega_{LO})t + \phi_{im} + \Delta\theta] - \frac{A_{im}}{2} \cos[(\omega_{im} - \omega_{LO})t + \phi_{im}]. \quad (4.75)$$

To arrive at a meaningful measure of the image rejection, we divide the image-to-signal ratio at the input by the same ratio at the output.²⁰ The result is called the “image rejection ratio” (IRR). Noting that the average power of the vector sum $a \cos(\omega t + \alpha) + b \cos \omega t$ is given by $(a^2 + 2ab \cos \alpha + b^2)/2$, we write the output image-to-signal ratio as

$$\frac{P_{im}}{P_{sig}} \Big|_{out} = \frac{A_{im}^2 (1 + \epsilon)^2 - 2(1 + \epsilon) \cos \Delta\theta + 1}{A_{sig}^2 (1 + \epsilon)^2 + 2(1 + \epsilon) \cos \Delta\theta + 1}. \quad (4.76)$$

Since the image-to-signal ratio at the input is given by A_{im}^2/A_{sig}^2 , the IRR can be expressed as

$$IRR = \frac{(1 + \epsilon)^2 + 2(1 + \epsilon) \cos \Delta\theta + 1}{(1 + \epsilon)^2 - 2(1 + \epsilon) \cos \Delta\theta + 1}. \quad (4.77)$$

Note that ϵ denotes the *relative* gain error and $\Delta\theta$ is in radians. Also, to express IRR in dB, we must compute $10 \log IRR$ (rather than $20 \log IRR$).

Example 4.31

If $\epsilon \ll 1$ rad, simplify the expression for IRR.

Solution:

Since $\cos \Delta\theta \approx 1 - \Delta\theta^2/2$ for $\Delta\theta \ll 1$ rad, we can reduce (4.77) to

$$IRR \approx \frac{4 + 4\epsilon + \epsilon^2 - (1 + \epsilon)\Delta\theta^2}{\epsilon^2 + (1 + \epsilon)\Delta\theta^2}. \quad (4.78)$$

In the numerator, the first term dominates and in the denominator $\epsilon \ll 1$, yielding

$$IRR \approx \frac{4}{\epsilon^2 + \Delta\theta^2}. \quad (4.79)$$

20. Note that the ratio of the output image power and the input image power is not meaningful because it depends on the gain.

Example 4.31 (Continued)

For example, $\epsilon = 10\%$ (≈ 0.83 dB)²¹ limits the IRR to 26 dB. Similarly, $\Delta\theta = 10^\circ$ yields an IRR of 21 dB. While such mismatch values may be tolerable in direct-conversion receivers, they prove inadequate here.

With various mismatches arising in the LO and signal paths, the IRR typically falls below roughly 35 dB. This issue and a number of other drawbacks limit the utility of the Hartley architecture.

Another critical drawback, especially in CMOS technology, originates from the variation of the absolute values of R_1 and C_1 in Fig. 4.64. Recall from Fig. 4.58 that the phase shift produced by the RC - CR network remains equal to 90° even with such variations, but the output amplitudes are equal at only $\omega = (R_1 C_1)^{-1}$. Specifically, if R_1 and C_1 are nominally chosen for a certain IF, $(R_1 C_1)^{-1} = \omega_{IF}$, but respectively experience a small change of ΔR and ΔC with process or temperature, then the ratio of the output amplitudes of the high-pass and low-pass sections is given by

$$\left| \frac{H_{HPF}}{H_{LPF}} \right| = (R_1 + \Delta R)(C_1 + \Delta C)\omega_{IF} \quad (4.80)$$

$$\approx 1 + \frac{\Delta R}{R_1} + \frac{\Delta C}{C_1}. \quad (4.81)$$

Thus, the gain mismatch is equal to

$$\epsilon = \frac{\Delta R}{R_1} + \frac{\Delta C}{C_1}. \quad (4.82)$$

For example, $\Delta R/R_1 = 20\%$ limits the image rejection to only 20 dB. Note that these calculations have assumed perfect matching between the high-pass and low-pass sections. If the resistors or capacitors exhibit mismatches, the IRR degrades further.

Another drawback resulting from the RC - CR sections manifests itself if the signal translated to the IF has a wide bandwidth. Since the gains of the high-pass and low-pass sections depart from each other as the frequency departs from $\omega_{IF} = (R_1 C_1)^{-1}$ [Fig. 4.58(b)], the image rejection may degrade substantially near the edges of the channel. In Problem 4.17, the reader can prove that, at a frequency of $\omega_{IF} + \Delta\omega$, the IRR is given by

$$IRR = \left(\frac{\omega_{IF}}{\Delta\omega} \right)^2. \quad (4.83)$$

For example, a fractional bandwidth of $2\Delta\omega/\omega_{IF} = 5\%$ limits the IRR to 32 dB.

The limitation expressed by Eq. (4.83) implies that ω_{IF} cannot be zero, dictating a heterodyne approach. Figure 4.65 shows an example where the first IF is followed by another quadrature downconverter so as to produce the baseband signals. Unlike the sliding-IF

21. To calculate ϵ in dB, we write $20 \log(1 + 10\%) = 0.83$ dB.

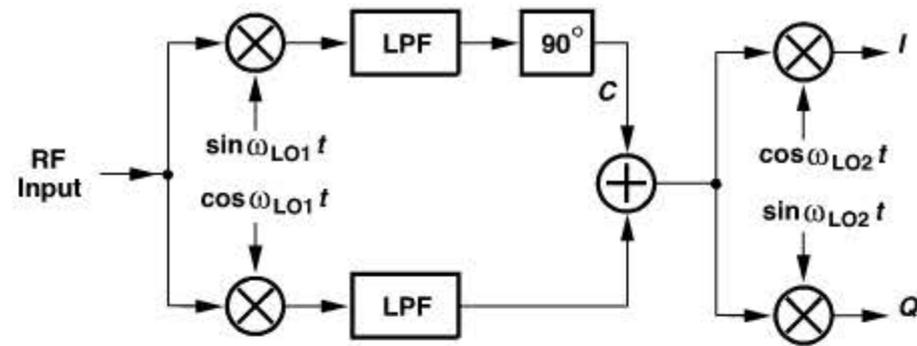


Figure 4.65 Downconversion of Hartley receiver output to baseband.

architecture of Fig. 4.26(a), this topology also requires the quadrature phases of the first LO, a critical disadvantage. The mixing spurs studied in Section 4.2.1 persist here as well.

The RC - CR sections used in Fig. 4.64 also introduce attenuation and noise. The 3-dB loss resulting from $|H_{HPF}| = |H_{LPF}| = 1/\sqrt{2}$ at $\omega = (R_1 C_1)^{-1}$ directly amplifies the noise of the following adder. Moreover, the input impedance of each section, $|R_1 + (C_1 s)^{-1}|$, reaches $\sqrt{2}R_1$ at $\omega = (R_1 C_1)^{-1}$, imposing a trade-off between the loading seen by the mixers and the thermal noise of the 90° shift circuit.

The voltage adder at the output of the Hartley architecture also poses difficulties as its noise and nonlinearity appear in the signal path. Illustrated in Fig. 4.66, the summation is typically realized by differential pairs, which convert the signal voltages to currents, sum the currents, and convert the result to a voltage.

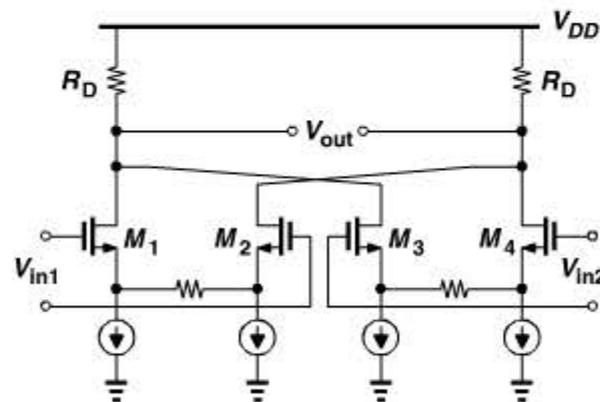


Figure 4.66 Summation of two voltages.

Weaver Architecture Our analysis of the Hartley architecture has revealed several issues that arise from the use of the RC - CR phase shift network. The Weaver receiver, derived from its transmitter counterpart [5], avoids these issues.

As recognized in Fig. 4.59, mixing a signal with quadrature phases of an LO takes the Hilbert transform. Depicted in Fig. 4.67, the Weaver architecture replaces the 90° phase shift network with quadrature mixing. To formulate the circuit's behavior, we begin with $x_A(t)$ and $x_B(t)$ as given by Eqs. (4.70) and (4.71), respectively, and perform the second

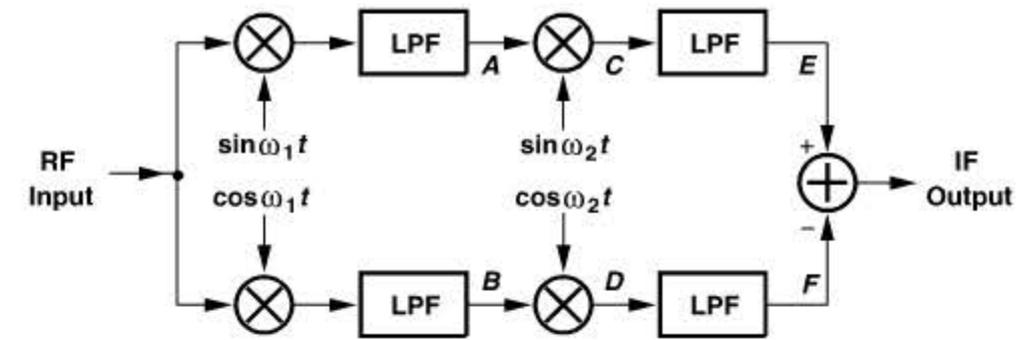


Figure 4.67 Weaver architecture.

quadrature mixing operation, arriving at

$$x_C(t) = \frac{A_{sig}}{4} \cos[(\omega_c - \omega_1 - \omega_2)t + \phi_{sig}] + \frac{A_{im}}{4} \cos[(\omega_{im} - \omega_1 - \omega_2)t + \phi_{im}] + \frac{A_{sig}}{4} \cos[(\omega_c - \omega_1 + \omega_2)t + \phi_{sig}] + \frac{A_{im}}{4} \cos[(\omega_{im} - \omega_1 + \omega_2)t + \phi_{im}] \quad (4.84)$$

$$x_D(t) = -\frac{A_{sig}}{4} \cos[(\omega_c - \omega_1 - \omega_2)t + \phi_{sig}] - \frac{A_{im}}{4} \cos[(\omega_{im} - \omega_1 - \omega_2)t + \phi_{im}] + \frac{A_{sig}}{4} \cos[(\omega_c - \omega_1 + \omega_2)t + \phi_{sig}] + \frac{A_{im}}{4} \cos[(\omega_{im} - \omega_1 + \omega_2)t + \phi_{im}]. \quad (4.85)$$

Should these results be added or subtracted? Let us assume low-side injection for both mixing stages. Thus, $\omega_{im} < \omega_1$ and $\omega_1 - \omega_{im} > \omega_2$ (Fig. 4.68). Also, $\omega_1 - \omega_{im} + \omega_2 > \omega_1 - \omega_{im} - \omega_2$. The low-pass filters following points C and D in Fig. 4.67 must therefore remove the components at $\omega_1 - \omega_{im} + \omega_2 (= \omega_c - \omega_1 + \omega_2)$, leaving only those at $\omega_1 - \omega_{im} - \omega_2 (= \omega_c - \omega_1 - \omega_2)$. That is, the *second* and *third* terms in Eqs. (4.84) and (4.85) are filtered. Upon subtracting $x_F(t)$ from $x_E(t)$, we obtain

$$x_E(t) - x_F(t) = \frac{A_{sig}}{2} \cos[(\omega_c - \omega_1 - \omega_2)t + \phi_{sig}]. \quad (4.86)$$

The image is therefore removed. In Problem 4.19, we consider the other three combinations of low-side and high-side injection so as to determine whether the outputs must be added or subtracted.

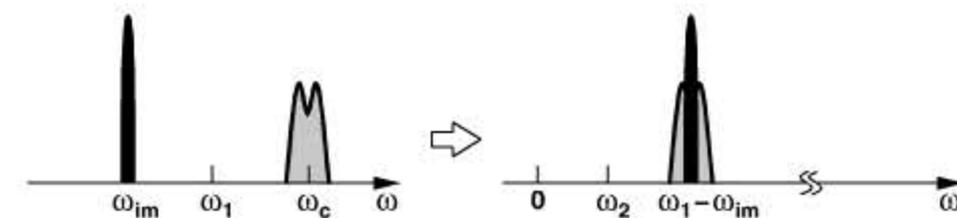


Figure 4.68 RF and IF spectra in Weaver architecture.

Example 4.32

Perform the above analysis graphically. Assume low-side injection for both mixing stages.

Solution:

Recall from Fig. 4.60(b) that low-side injection mixing with a sine multiplies the spectrum by $+(j/2)\text{sgn}(\omega)$. Beginning with the spectra of Fig. 4.61 and mixing them with $\sin \omega_2 t$ and $\cos \omega_2 t$, we arrive at the spectra shown in Fig. 4.69. Subtraction of $X_F(f)$ from $X_E(f)$ thus yields the signal and removes the image.

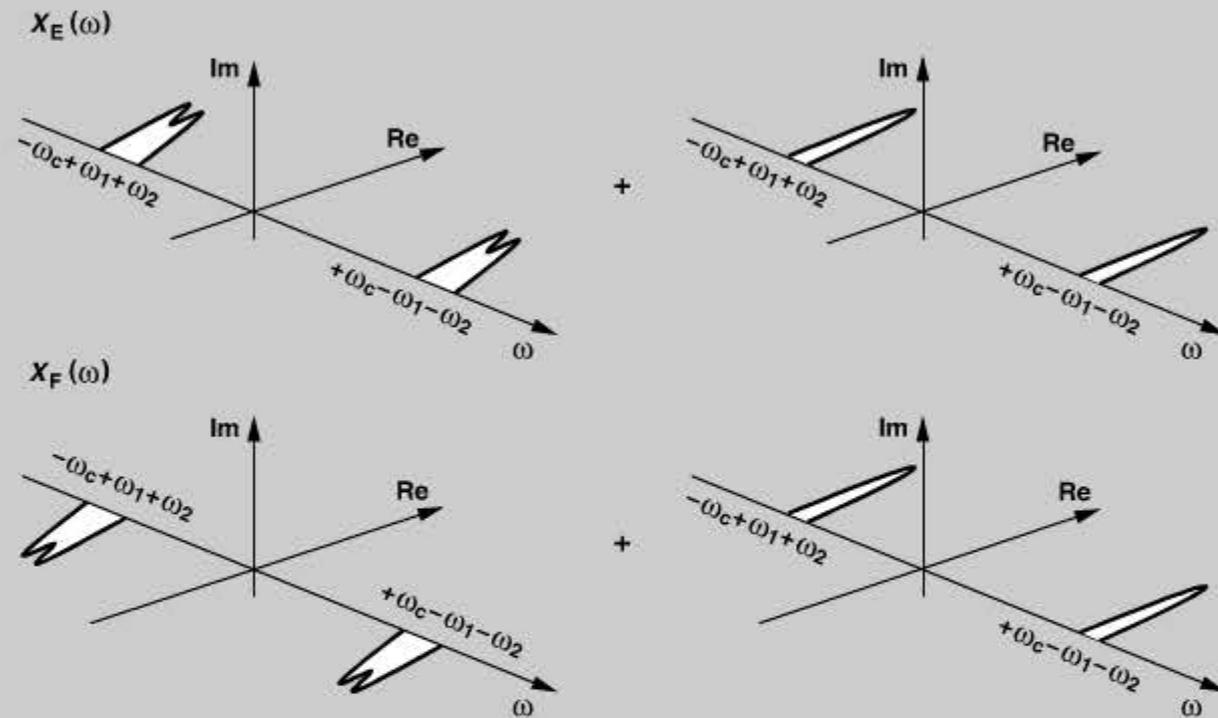


Figure 4.69 Signal and image spectra in Weaver architecture.

While employing two more mixers and one more LO than the Hartley architecture, the Weaver topology avoids the issues related to RC - CR networks: resistance and capacitance variations, degradation of IRR as the frequency departs from $1/(R_1 C_1)$, attenuation, and noise. Also, if the IF mixers are realized in active form (Chapter 6), their outputs are available in the current domain and can be summed directly. Nonetheless, the IRR is still limited by mismatches, typically falling below 40 dB.

The Weaver architecture must deal with a *secondary image* if the second IF is not zero. Illustrated in Fig. 4.70, this effect arises if a component at $2\omega_2 - \omega_{in} + 2\omega_1$ accompanies the RF signal. Downconversion to the first IF translates this component to $2\omega_2 - \omega_{in} + \omega_1$, i.e., image of the signal with respect to ω_2 , and mixing with ω_2 brings it to $\omega_2 - \omega_{in} + \omega_1$, the same IF at which the signal appears. For this reason, the second downconversion preferably

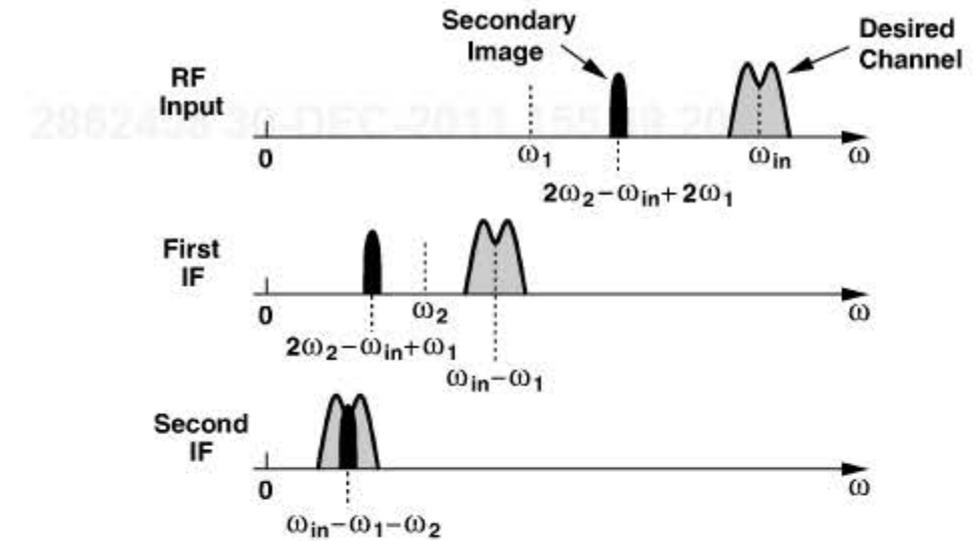


Figure 4.70 Secondary image in Weaver architecture.

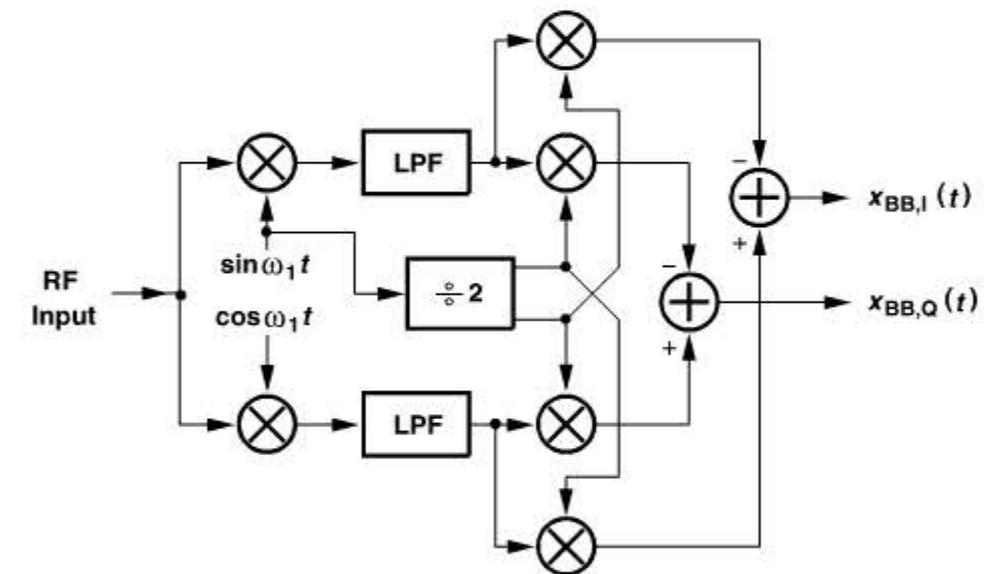


Figure 4.71 Double quadrature downconversion Weaver architecture to produce baseband outputs.

produces a zero IF, in which case it must perform quadrature separation as well. Figure 4.71 shows an example [6], where the second LO is derived from the first by frequency division.

The Weaver topology also suffers from mixing spurs in both downconversion steps. In particular, the harmonics of the second LO frequency may downconvert interferers from the first IF to baseband.

Calibration For image-rejection ratios well above 40 dB, the Hartley or Weaver architectures must incorporate calibration, i.e., a method of cancelling the gain and phase mismatches. A number of calibration techniques have been reported [7, 9].

4.2.5 Low-IF Receivers

In our study of heterodyne receivers, we noted that it is undesirable to place the image within the signal band because the image thermal noise of the antenna, the LNA, and the input stage of the RF mixer would raise the overall noise figure by approximately 3 dB.²² In “low-IF receivers,” the image indeed falls in the band but is suppressed by image rejection operations similar to those described in Section 4.2.4. To understand the motivation for the use of low-IF architectures, let us consider a GSM receiver as an example. As explained in Section 4.2.3, direct conversion of the 200-kHz desired channel to a zero IF may significantly corrupt the signal by $1/f$ noise. Furthermore, the removal of the dc offset by means of a high-pass filter proves difficult. Now suppose the LO frequency is placed at the *edge* of the desired (200-kHz) channel [Fig. 4.72(a)], thereby translating the RF signal to an IF of 100 kHz. With such an IF, and because the signal carries little information near the edge, the $1/f$ noise penalty is much less severe. Also, on-chip high-pass filtering of the signal becomes feasible. Called a “low-IF receiver,” this type of system is particularly attractive for narrow-channel standards.

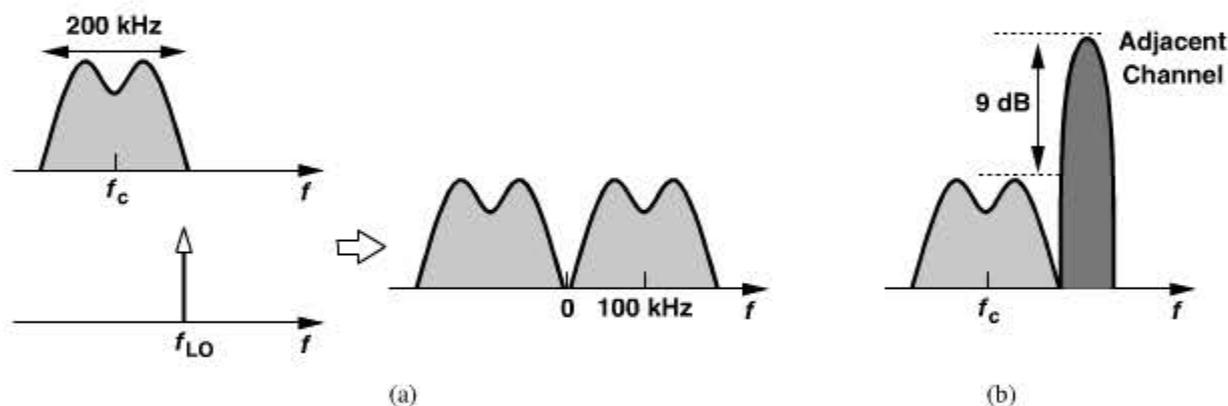


Figure 4.72 (a) Spectra in a low-IF receiver, (b) adjacent-channel specification in GSM.

The heterodyne downconversion nonetheless raises the issue of the image, which in this case falls in the adjacent channel. Fortunately, the GSM standard requires that receivers tolerate an adjacent channel only 9 dB above the desired channel (Chapter 3) [Fig. 4.72(b)]. Thus, an image-reject receiver with a moderate IRR can lower the image to well below the signal level. For example, if $IRR = 30$ dB, the image remains 21 dB below the signal.

Example 4.33

Repeat Example 4.24 for a low-IF receiver.

Solution:

Assuming that high-pass filtering of dc offsets also removes the flicker noise up to roughly 20 kHz, we integrate the noise from 20 kHz to 200 kHz (Fig. 4.73):

22. This occurs because the entire signal band must see a flat frequency response in the antenna/LNA/mixer chain.

Example 4.33 (Continued)

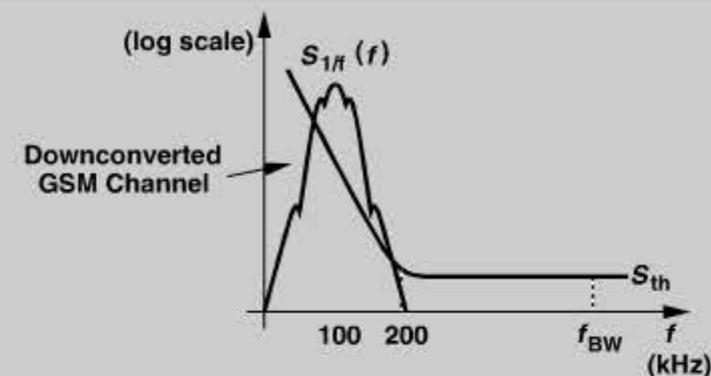


Figure 4.73 Effect of flicker noise in low-IF GSM receiver.

$$P_{n1} = \int_{20 \text{ kHz}}^{200 \text{ kHz}} \frac{\alpha}{f} df \quad (4.87)$$

$$= f_c \cdot S_{th} \ln 10 \quad (4.88)$$

$$= 2.3 f_c S_{th}. \quad (4.89)$$

Without flicker noise,

$$P_{n2} \approx (200 \text{ kHz}) S_{th}. \quad (4.90)$$

It follows that

$$\frac{P_{n1}}{P_{n2}} = 2.3 (= 3.62 \text{ dB}). \quad (4.91)$$

The flicker noise penalty is therefore much lower in this case.

How is image rejection realized in a low-IF receiver? The Hartley architecture employing the $RC-CR$ network (Fig. 4.64) appears to be a candidate, but the IF spectrum in a low-IF RX may extend to zero frequency, making it impossible to maintain a high IRR across the signal bandwidth. (The high-pass Section exhibits zero gain near frequency!) While avoiding this issue, the Weaver architecture must deal with the secondary image if the second IF is not zero or with flicker noise if it is.

One possible remedy is to move the 90° phase shift in the Hartley architecture from the IF path to the RF path. Illustrated in Fig. 4.74, the idea is to first create the quadrature phases of the RF signal and the image and subsequently perform another Hilbert transform by means of quadrature mixing. We also recognize some similarity between this topology and the Weaver architecture: both multiply quadrature components of the signal and the image by the quadrature phases of the LO and sum the results, possibly in the current domain. Here, the $RC-CR$ network is centered at a high frequency and can maintain a reasonable IRR across the band. For example, for the 25-MHz receive band of 900-MHz GSM, if $(2\pi R_1 C_1)^{-1}$ is chosen equal to the center frequency, then Eq. (4.83) implies an

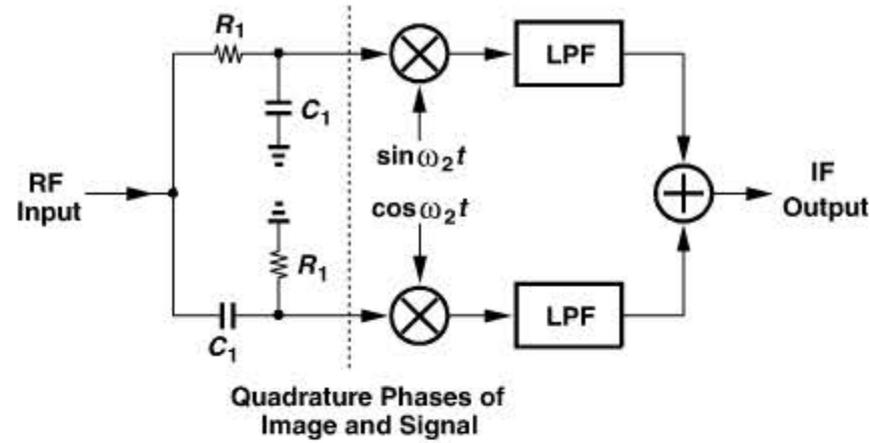


Figure 4.74 Quadrature phase separation in RF path of a Hartley receiver.

IRR of $20 \log(900 \text{ MHz}/12.5 \text{ MHz}) = 37 \text{ dB}$. However, the variation of R_1 and C_1 still limits the IRR to about 20 dB [Eq. (4.82)].

Another variant of the low-IF architecture is shown in Fig. 4.75. Here, the downconverted signals are applied to channel-select filters and amplifiers as in a direct-conversion receiver.²³ The results are then digitized and subjected to a Hilbert transform in the digital domain before summation. Avoiding the issues related to the analog 90° phase shift operation, this approach proves a viable choice. Note that the ADCs must accommodate a signal bandwidth twice that in a direct-conversion receiver, thus consuming higher power. This issue is unimportant in narrow-channel standards such as GSM because the ADC power dissipation is but a small fraction of that of the overall system.

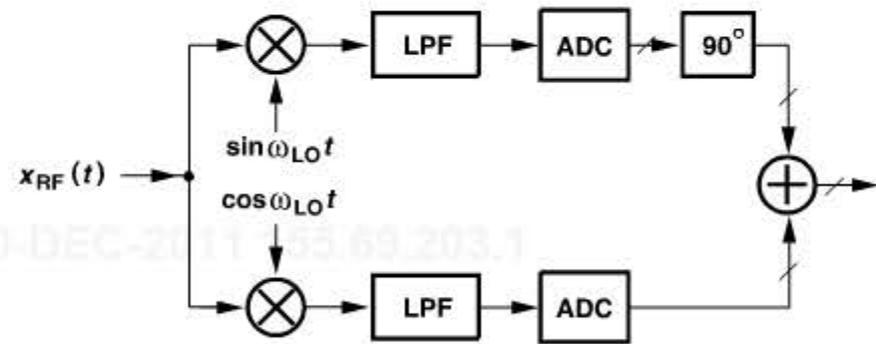


Figure 4.75 Low-IF receiver with 90° phase shift in digital domain.

Let us summarize our thoughts thus far. If a low-IF receiver places the image in the adjacent channel, then it cannot employ an RC - CR 90° phase shift after downconversion. Also, a 90° circuit in the RF path still suffers from RC variation. For these reasons, the concept of “low IF” can be extended to any downconversion that places the image within

23. The channel-select filters must, however, provide a bandwidth equal to the RF signal bandwidth rather than half of it [Fig. 4.72(a)].

the band so that the IF is significantly *higher* than the signal bandwidth, possibly allowing the use of an RC - CR network—but not so high as to unduly burden the ADC. Of course, since the image no longer lies in the adjacent channel, a substantially higher IRR may be required. Some research has therefore been expended on low-IF receivers with high image rejection. Such receivers often employ “polyphase filters” (PPFs) [10, 11].

Polyphase Filters Recall from Section 4.2.4 that heterodyne quadrature downconversion subjects the signal to low-side injection and the image to high-side injection, or vice versa, thus creating the Hilbert transform of one and the negative Hilbert transform of the other. Now let us consider the circuit shown in Fig. 4.76(a), where V_{out} can be viewed as a weighted sum of V_1 and V_2 :

$$V_{out} = \frac{V_1 + R_1 C_1 s V_2}{R_1 C_1 s + 1}. \quad (4.92)$$

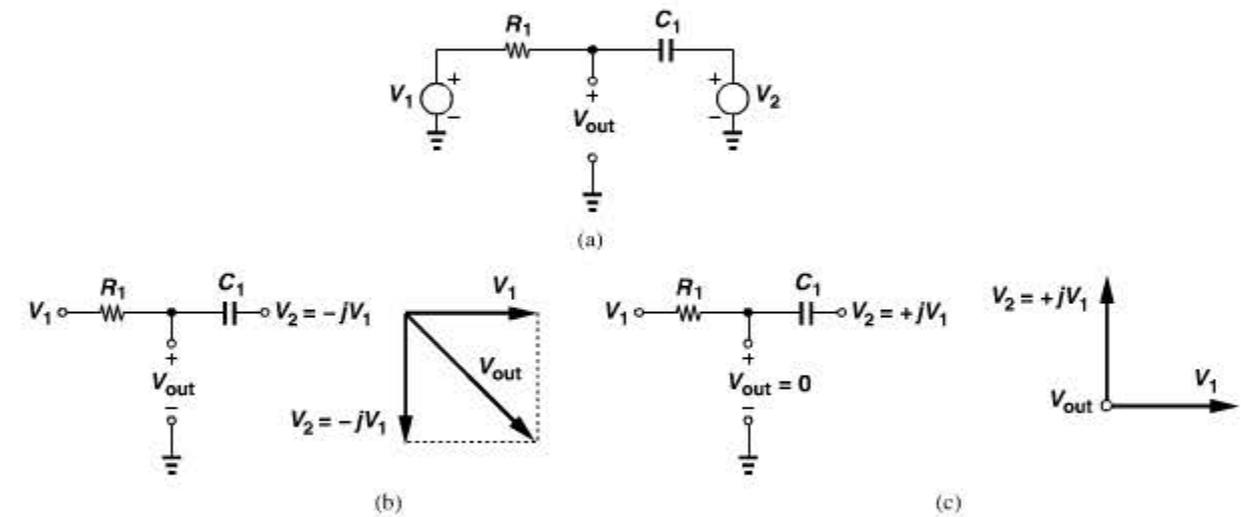


Figure 4.76 (a) Simple RC circuit, (b) output in response to V_1 and $-jV_1$, (c) output in response to V_1 and $+jV_1$.

We consider two special cases.

1. The voltage V_2 is the Hilbert transform of V_1 ; in phasor form, $V_2 = -jV_1$ (Example 4.26). Consequently, for $s = j\omega$,

$$V_{out} = V_1 \frac{R_1 C_1 \omega + 1}{jR_1 C_1 \omega + 1}. \quad (4.93)$$

If $\omega = (R_1 C_1)^{-1}$, then $V_{out} = 2V_1/(1 + j) = V_1(1 - j)$. That is, $|V_{out}| = \sqrt{2}V_1$ and $\angle V_{out} = \angle V_{in} - 45^\circ$ [Fig. 4.76(b)]. In this case, the circuit simply computes the vector summation of V_1 and $V_2 = -jV_1$. We say the circuit rotates by -45° the voltage sensed by the resistor.

2. The voltage V_2 is the *negative* Hilbert transform of V_1 , i.e., $V_2 = +jV_1$. For $s = j\omega$,

$$V_{out} = V_1 \frac{-R_1 C_1 \omega + 1}{jR_1 C_1 \omega + 1}. \quad (4.94)$$

Interestingly, if $\omega = (R_1 C_1)^{-1}$, then $V_{out} = 0$ [Fig. 4.76(c)]. Intuitively, we can say that C_1 rotates V_2 by another 90° so that the result cancels the effect of V_1 at the output node. The reader is encouraged to arrive at these conclusions using superposition.

In summary, the series branch of Fig. 4.76(a) rotates V_1 by -45° (to produce V_{out}) if $V_2 = -jV_1$ and rejects V_1 if $V_2 = +jV_1$. The circuit can therefore distinguish between the signal and the image if it follows a quadrature downconverter.

Example 4.34

Extend the topology of Fig. 4.76(a) if V_1 and $-jV_1$ are available in differential form and construct an image-reject receiver.

Solution:

Figure 4.77(a) shows the arrangement and the resulting phasors if $R_1 = R_2 = R$ and $C_1 = C_2 = C$. The connections to quadrature downconversion mixers are depicted in Fig. 4.77(b).

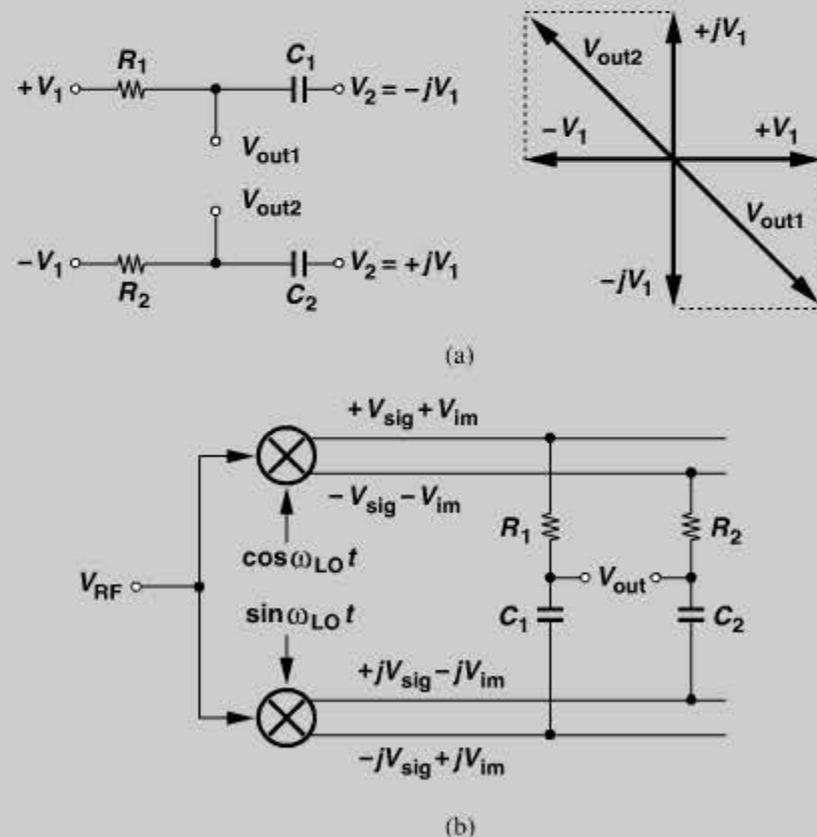


Figure 4.77 (a) RC circuit sensing differential inputs, (b) quadrature downconverter driving RC network of part (a).

In contrast to the Hartley architecture of Fig. 4.62, the circuit of Fig. 4.77(b) avoids an explicit voltage adder at the output. Nonetheless, this arrangement still suffers from RC variations and a narrow bandwidth. In fact, at an IF of $\omega = (R_1 C_1)^{-1} + \Delta\omega$, Eq. (4.94) yields the residual image as

$$|V_{out}| \approx |V_1| \frac{RC\Delta\omega}{\sqrt{2 + 2RC\Delta\omega}} \quad (4.95)$$

$$\approx |V_1| \frac{RC\Delta\omega}{\sqrt{2}}, \quad (4.96)$$

where it is assumed that $\Delta\omega \ll \omega$.

In the next step of our development of polyphase filters, let us now redraw the circuit of Fig. 4.77(a) and add two branches to it as shown in Fig. 4.78(a). Here, the capacitors are equal and so are the resistors. The top and bottom branches still produce differential outputs, but how about the left and right branches? Since R_3 and C_3 compute the weighted sum of $+jV_1$ and $+V_1$, we observe from Fig. 4.76(b) that V_{out3} is 45° more negative than $+jV_1$. By the same token, V_{out4} is 45° more negative than $-jV_1$. Figure 4.78(b) depicts the resulting phasors at $\omega = (R_1 C_1)^{-1}$, suggesting that the circuit produces quadrature outputs that are 45° out of phase with respect to the quadrature inputs.

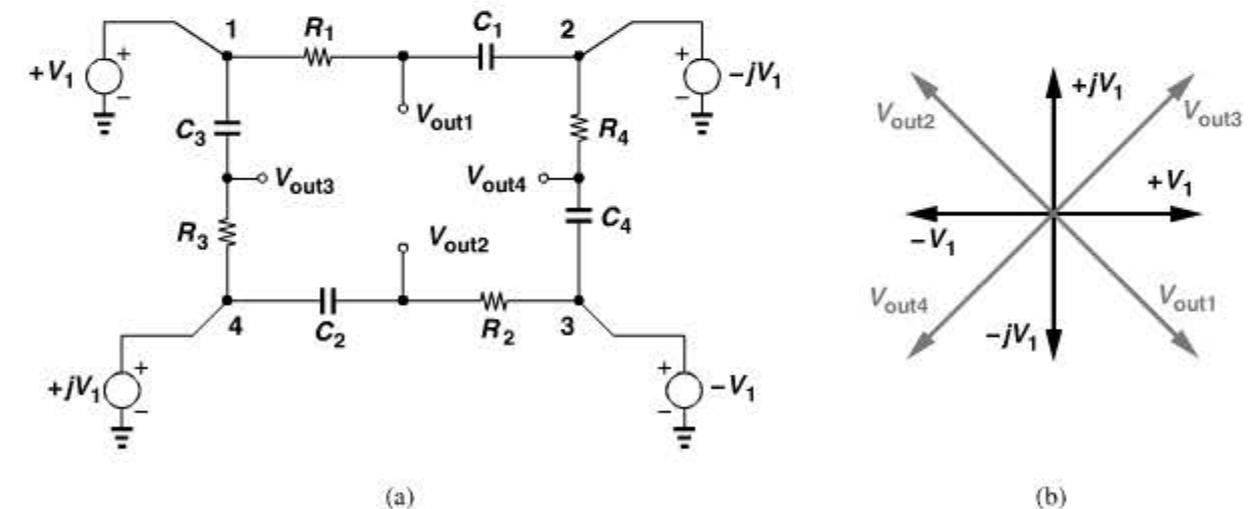


Figure 4.78 (a) RC network sensing differential quadrature phases, (b) resulting outputs.

Example 4.35

The outputs of a quadrature downconverter contain the signal, V_{sig} , and the image, V_{im} , and drive the circuit of Fig. 4.78(a) as shown in Fig. 4.79(a). Determine the outputs, assuming all capacitors are equal to C and all resistors equal to R .

(Continues)

Example 4.35 (Continued)

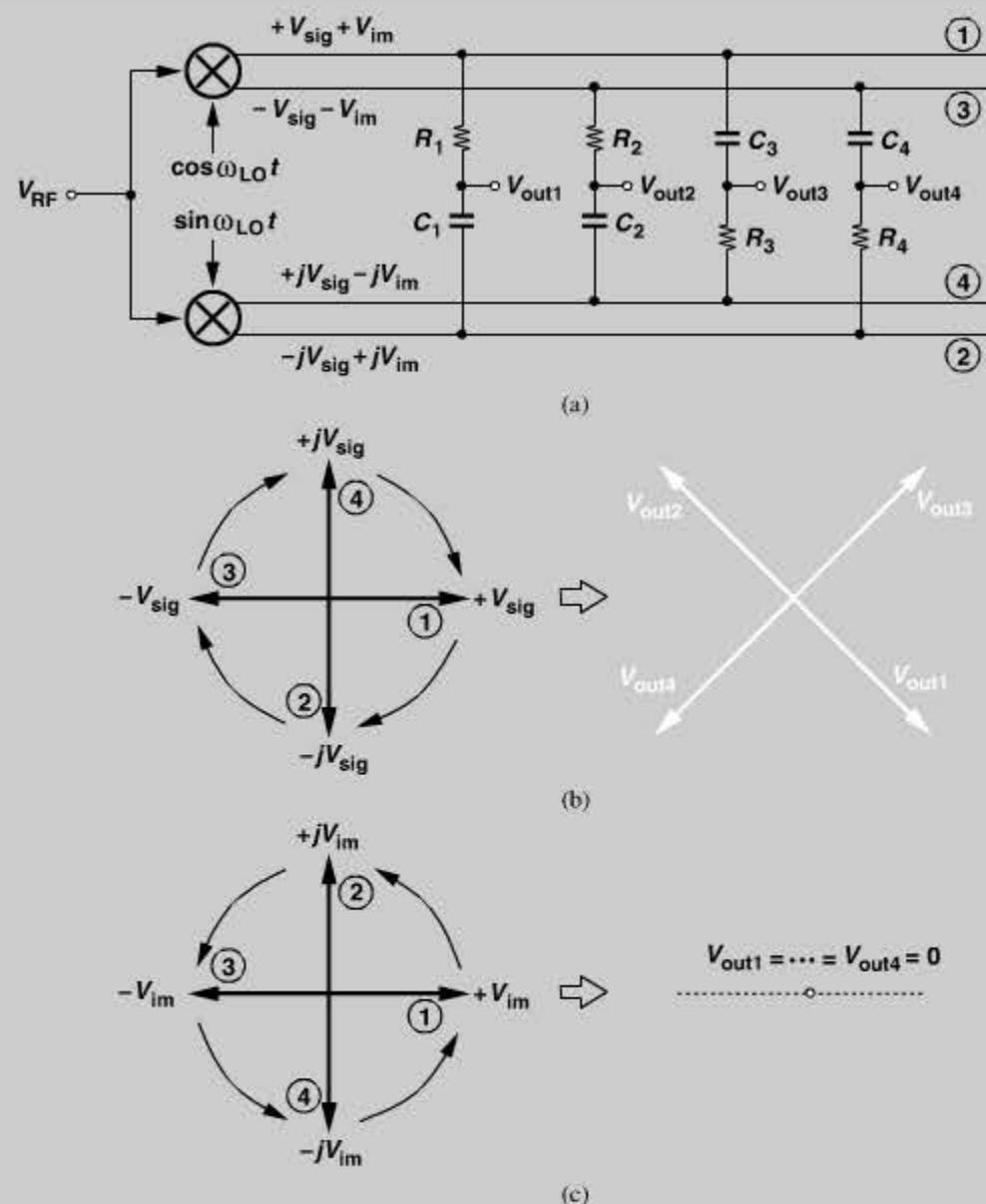


Figure 4.79 (a) Quadrature downconverter driving RC sections, (b) resulting signal output, (c) resulting image output.

Solution:

The quadrature downconverter produces $+V_{sig} + V_{im}$, $-V_{sig} - V_{im}$, $+jV_{sig} - jV_{im}$, and $-jV_{sig} + jV_{im}$. At $\omega = (RC)^{-1}$, the branch R_1C_1 rotates $+V_{sig}$ by -45° to generate V_{out1} . Similarly, R_2C_2 rotates $-V_{sig}$ by 45° to generate V_{out2} , etc. [Fig. 4.79(b)]. The image components, on the other hand, yield a zero output [Fig. 4.79(c)]. The key point here is that, if we consider the sequence of nodes 1, 2, 3, and 4, we observe that V_{sig} rotates *clockwise* by

Example 4.35 (Continued)

90° from one node to the next, whereas V_{im} rotates *counterclockwise* by 90° from one node to the next. The circuit therefore exhibits an asymmetry in its response to the “sequence” of the four inputs.

The multiphase circuit of Fig. 4.78(a) is called a “sequence-asymmetric polyphase filter” [8]. Since the signal and the image arrive at the inputs with different sequences, one is passed to the outputs while the other is suppressed. But what happens if $\omega \neq (RC)^{-1}$? Substituting $\omega = (R_1C_1)^{-1} + \Delta\omega$ in Eq. (4.93), we have

$$V_{out1} = V_{sig} \frac{2 + RC\Delta\omega}{1 + j(1 + RC\Delta\omega)}, \quad (4.97)$$

and hence,

$$|V_{out1}|^2 = |V_{sig}|^2 \frac{4 + 4RC\Delta\omega + R^2C^2\Delta\omega^2}{2 + 2RC\Delta\omega + R^2C^2\Delta\omega^2} \quad (4.98)$$

$$\approx 2|V_{sig}|^2 \left(1 + RC\Delta\omega + \frac{R^2C^2\Delta\omega^2}{4}\right) \left(1 - RC\Delta\omega - \frac{R^2C^2\Delta\omega^2}{2}\right) \quad (4.99)$$

$$\approx 2|V_{sig}|^2 \left(1 - \frac{5}{4}R^2C^2\Delta\omega^2\right). \quad (4.100)$$

That is,

$$|V_{out1}| \approx \sqrt{2}|V_{sig}| \left(1 - \frac{5}{8}R^2C^2\Delta\omega^2\right). \quad (4.101)$$

The phase of V_{out1} is obtained from (4.97) as

$$\angle V_{out1} = \angle V_{sig} - \tan^{-1}(1 + RC\Delta\omega). \quad (4.102)$$

Since $\tan^{-1}(1 + RC\Delta\omega) \approx \pi/4 + RC\Delta\omega/2$ for $RC\Delta\omega \ll 1$ rad,

$$\angle V_{out1} = \angle V_{sig} - \left(\frac{\pi}{4} + \frac{RC\Delta\omega}{2}\right). \quad (4.103)$$

Figure 4.80(a) illustrates the effect on all four phases of the signal, implying that the outputs retain their differential and quadrature relationship.

For the image, we return to Eq. (4.96) and note that the four outputs have a magnitude equal to $V_{im}RC\Delta\omega/\sqrt{2}$ and phases similar to those of the signal components in Fig. 4.80(a). The output image phasors thus appear as shown in Fig. 4.80(b). The reader is encouraged to prove that V_{out1} is at $-45^\circ - RC\Delta\omega/2$ and V_{out3} at $-135^\circ - RC\Delta\omega/2$.

An interesting observation in Fig. 4.80 is that the output signal and image components exhibit *opposite* sequences [10, 11]. We therefore expect that if this polyphase filter is

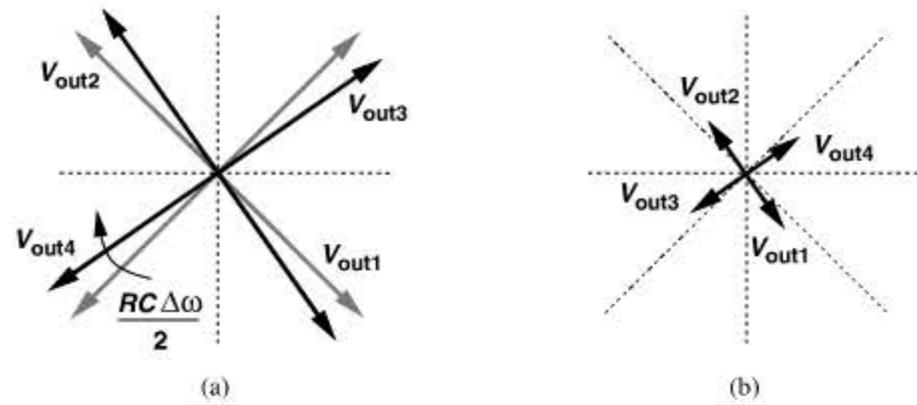


Figure 4.80 Effect of polyphase filter at a frequency offset of $\Delta\omega$ for (a) signal, and (b) image.

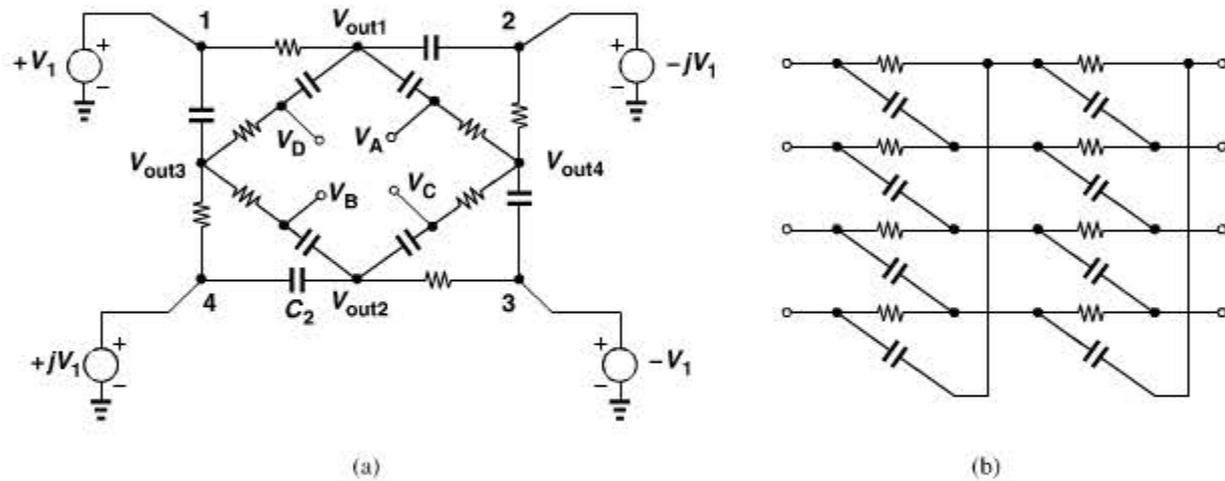


Figure 4.81 (a) Cascaded polyphase sections, (b) alternative drawing.

followed by another, then the image can be further suppressed. Figure 4.81(a) depicts such a cascade and Fig. 4.81(b) shows an alternative drawing that more easily lends itself to cascading.

We must now answer two questions: (1) how should we account for the loading of the second stage on the first? (2) how are the RC values chosen in the two stages? To answer the first question, consider the equivalent circuit shown in Fig. 4.82, where Z_1 - Z_4 represent the RC branches in the second stage. Intuitively, we note that Z_1 attempts to “pull” the phasors V_{out1} and V_{out3} toward each other, Z_2 attempts to pull V_{out1} and V_{out4} toward each other, etc. Thus, if $Z_1 = \dots = Z_4 = Z$, then, V_{out1} - V_{out4} experience no rotation, but the loading may reduce their magnitudes. Since the angles of V_{out1} - V_{out4} remain unchanged, we can express them as $\pm\alpha(1 \pm j)V_1$, where α denotes the attenuation due to the loading of the second stage. The currents drawn from node X by Z_1 and Z_2 are therefore equal to $[\alpha(1 - j)V_1 - \alpha(1 + j)V_1]/Z_1$ and $[\alpha(1 - j)V_1 + \alpha(1 + j)V_1]/Z_2$, respectively. Summing

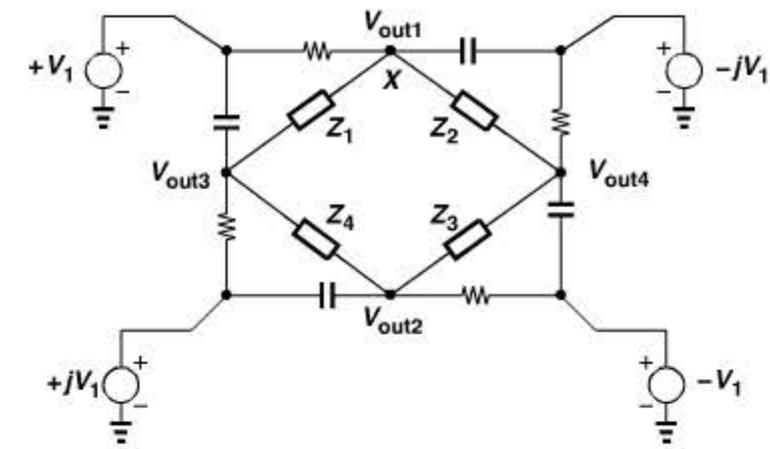


Figure 4.82 Effect of loading of second polyphase section.

all of the currents flowing out of node X and equating the result to zero, we have

$$\frac{\alpha(1 - j)V_1 - V_1}{R} + [\alpha(1 - j)V_1 + jV_1]Cj\omega + \frac{\alpha(1 - j)V_1 - \alpha(1 + j)V_1}{Z} + \frac{\alpha(1 - j)V_1 + \alpha(1 + j)V_1}{Z} = 0. \quad (4.104)$$

This equality must hold for any nonzero value of V_1 . If $RC\omega = 1$, the expression reduces to

$$2\alpha - 2 + \frac{2\alpha(1 - j)R}{Z} = 0. \quad (4.105)$$

That is,

$$\alpha = \frac{Z}{Z + (1 - j)R}. \quad (4.106)$$

For example, if $Z = R + (jC\omega)^{-1}$, then $\alpha = 1/2$, revealing that loading by an identical stage attenuates the outputs of the first stage by a factor of 2.

Example 4.36

If $Z = R + (jC\omega)^{-1}$ and $RC\omega = 1$, determine V_A in Fig. 4.81(a).

Solution:

We have $V_{out1} = (1/2)(1 - j)V_1$ and $V_{out4} = (1/2)(-1 - j)V_1$, observing that V_{out1} and V_{out2} have the same phase relationship as V_1 and V_2 in Fig. 4.76(b). Thus, V_A is simply the vector sum of V_{out1} and V_{out4} :

$$V_A = -jV_1. \quad (4.107)$$

In comparison with Fig. 4.76(b), we note that a two-section polyphase filter produces an output whose magnitude is $\sqrt{2}$ times smaller than that of a single-section counterpart. We say each section attenuates the signal by a factor of $\sqrt{2}$.

The second question relates to the choice of RC values. Suppose both stages employ $RC = R_0C_0$. Then, the cascade of two stages yields an image attenuation equal to the *square* of Eq. (4.95) at a frequency of $(R_0C_0)^{-1} + \Delta\omega$:

$$\left| \frac{V_{im,out}}{V_{im,in}} \right| \approx \frac{(R_0C_0\Delta\omega)^2}{2 + 2R_0C_0\Delta\omega}, \quad (4.108)$$

which reduces to $(R_0C_0\Delta\omega)^2/2$ for $\Delta\omega \ll (R_0C_0)^{-1}$. Figure 4.83 plots this behavior, comparing it with that of a single section.

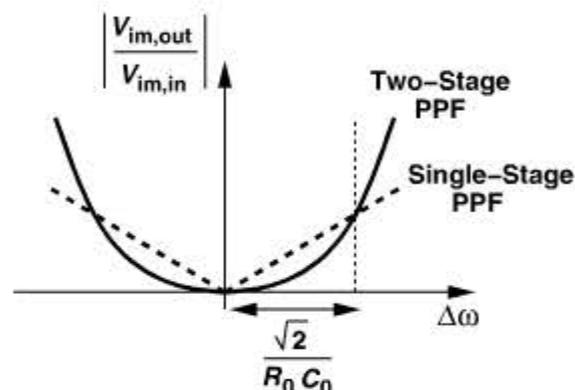


Figure 4.83 Image rejection for single-stage and two-stage polyphase filters.

What happens if the two stages use different time constants? In particular, for a given IF of ω_0 , let us assume that the time constants in the first and second stages are respectively equal to R_1C_1 and R_2C_2 such that $\omega_0 - (R_1C_1)^{-1} = (R_2C_2)^{-1} - \omega_0$, i.e., the center frequencies are shifted up and down. From Eq. (4.96), we plot the image rejections of the two stages as shown in Fig. 4.84(a).²⁴ The product of these two functions is a parabola crossing zero at $\omega_1 = (R_1C_1)^{-1}$ and $\omega_2 = (R_2C_2)^{-1}$ [Fig. 4.84(b)]. The reader can prove that the attenuation at ω_0 is equal to $(\omega_1 - \omega_2)^2 / (8\omega_1\omega_2)$, which must be chosen sufficiently small. The reader can also show that for an attenuation of 60 dB, $\omega_1 - \omega_2$ cannot exceed approximately 18% of ω_0 .

The advantage of splitting the cut-off frequencies of the two stages is the wider achievable bandwidth. Figure 4.85 plots the image rejection for $\omega_1 = \omega_2 = \omega_0$ and $\omega_1 \neq \omega_2$.

The cascading of polyphase filter sections entails both attenuation and additional thermal noise. To alleviate the former, the resistors in the latter stages can be chosen larger than those in the former, but at the cost of higher noise. For this reason, polyphase filters are only occasionally used in RF receivers. In low-IF architectures, the polyphase filters can be realized as “complex filters” so as to perform channel-selection filtering [12].

Double-Quadrature Downconversion In our study of the Hartley architecture, we noted that mismatches arise in both the RF signal path and the LO path. A method of reducing the effect of mismatches incorporates “double-quadrature” downconversion [10]. Illustrated in Fig. 4.86, the circuit decomposes the RF signal into quadrature components, performs

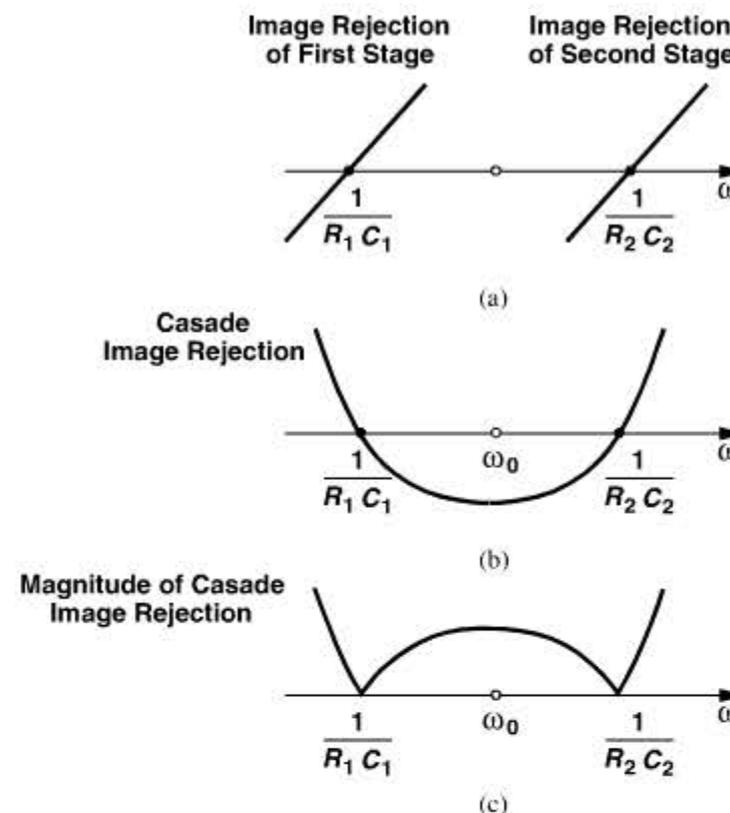


Figure 4.84 (a) Image rejections of two unidentical polyphase stages, (b) cascade image rejection, (c) magnitude of cascade image rejection.

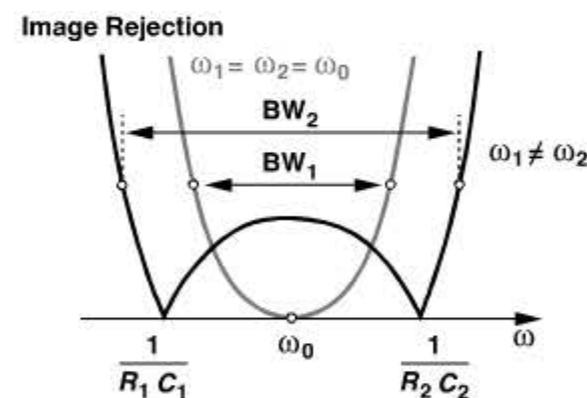


Figure 4.85 Comparison of image rejections of two identical and unidentical polyphase stages.

quadrature downconversion on *each* of the RF components, and subtracts and adds the results to produce net quadrature IF outputs. It can be shown [10] that the overall gain and phase mismatches of this topology are given by

$$\frac{\Delta A}{A} = \frac{\Delta A_{RF}}{A_{RF}} \cdot \frac{\Delta A_{LO}}{A_{LO}} + \frac{\Delta G_{mix}}{G_{mix}} \quad (4.109)$$

$$\tan(\Delta\phi) = \tan(\Delta\phi_{RF}) \cdot \tan(\Delta\phi_{LO}) + \frac{\tan(\Delta\phi_{mix})}{2}, \quad (4.110)$$

24. For clarity, the plots are allowed to be negative even though Eq. (4.96) contains absolute values.

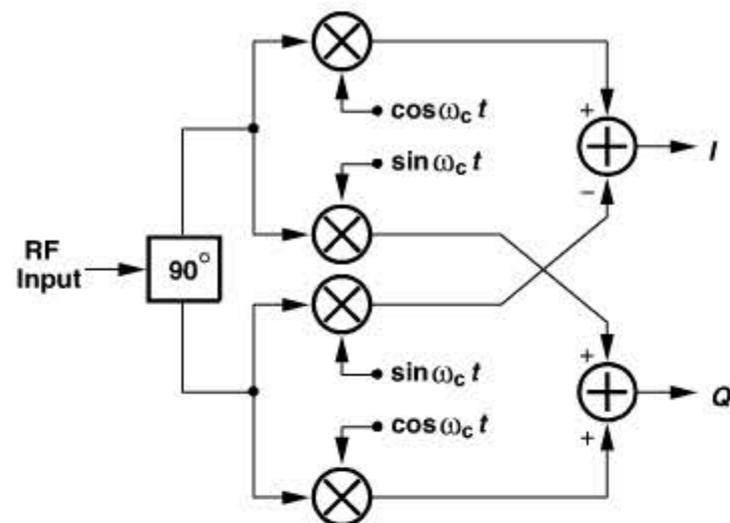


Figure 4.86 Low-IF receiver with double quadrature downconverter.

where $\Delta A_{RF}/A_{RF}$ and $\Delta A_{LO}/A_{LO}$ denote amplitude mismatches in the RF and LO paths, respectively, and $\Delta\phi_{RF}$ and $\Delta\phi_{LO}$ are the corresponding phase mismatches. The quantities $\Delta G_{mix}/G_{mix}$ and $\Delta\phi_{mix}$ denote the conversion gain mismatch and phase mismatch of the mixers, respectively. Thus, the IRR is only limited by mixer mismatches because the first terms on the right-hand sides of (4.109) and (4.110) are very small.

Equations (4.109) and (4.110) reveal that the IRR of the double-quadrature architecture is likely to fall well below 60 dB. This is because, even with no phase mismatch, $\Delta G_{mix}/G_{mix}$ must remain less than 0.1% for IRR = 60 dB, a very stringent requirement even for matching of simple resistors. Calibration of the mismatches can therefore raise the IRR more robustly [7, 9].

4.3 TRANSMITTER ARCHITECTURES

4.3.1 General Considerations

An RF transmitter performs modulation, upconversion, and power amplification. In most of today's systems, the data to be transmitted is provided in the form of quadrature baseband signals. For example, from Chapter 3 the GMSK waveform in GSM can be expanded as

$$x_{GMSK}(t) = A \cos[\omega_c t + m \int x_{BB}(t) * h(t) dt] \quad (4.111)$$

$$= A \cos \omega_c t \cos \phi - A \sin \omega_c t \sin \phi, \quad (4.112)$$

where

$$\phi = m \int x_{BB} * h(t) dt. \quad (4.113)$$

Thus, $\cos \phi$ and $\sin \phi$ are produced from $x_{BB}(t)$ by the digital baseband processor, converted to analog form by D/A converters, and applied to the transmitter.

We have seen the need for baseband pulse shaping in Chapter 3 and in the above equations for GMSK: each rectangular data pulse must be transformed to a smoother pulse.

Since pulse shaping in the analog domain, especially at low frequencies, requires bulky filters, each incoming pulse is mapped to the desired shape by a combination of digital and analog techniques. Illustrated in Fig. 4.87 is an example [13, 14], where the input pulse generates a sequence of addresses, e.g., it enables a counter, producing a set of levels from two read-only memories (ROMs). (We say the pulse is “oversampled.”) These levels are subsequently converted to analog form, yielding the desired pulse shape at points A and B.

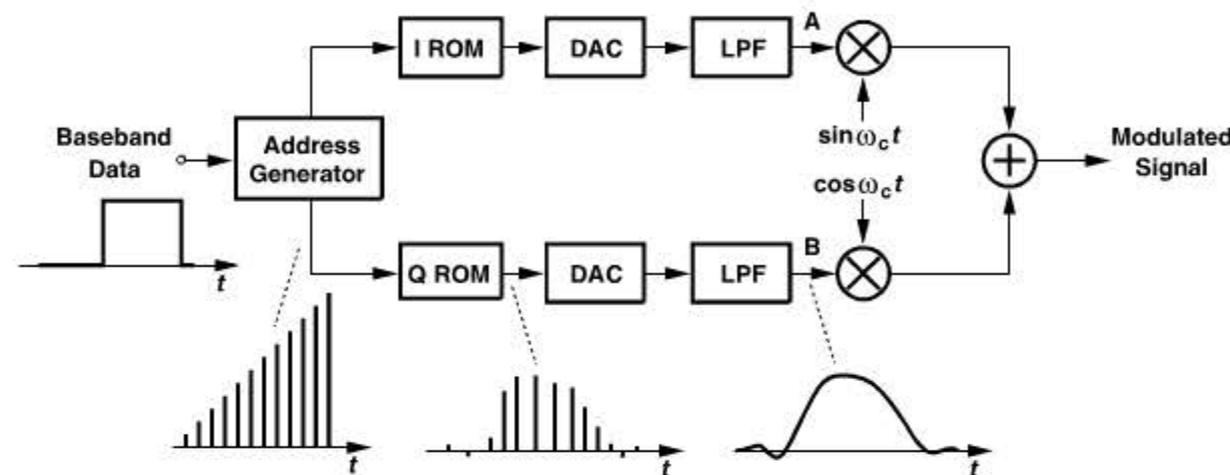


Figure 4.87 Baseband pulse shaping.

4.3.2 Direct-Conversion Transmitters

Most transmitter architectures are similar to the receiver topologies described in Section 4.2, but with the operations performed in “reverse” order. For example, an RX cascade of an LNA and quadrature downconversion mixers suggests a TX cascade of quadrature upconversion mixers and a PA.

The above expression of a GMSK waveform can be generalized to any narrowband modulated signal:

$$x(t) = A(t) \cos[\omega_c t + \phi(t)] \quad (4.114)$$

$$= A(t) \cos \omega_c t \cos[\phi(t)] - A(t) \sin \omega_c t \sin[\phi(t)]. \quad (4.115)$$

We therefore define the quadrature baseband signals as

$$x_{BB,I}(t) = A(t) \cos[\phi(t)] \quad (4.116)$$

$$x_{BB,Q}(t) = A(t) \sin[\phi(t)], \quad (4.117)$$

and construct the transmitter as shown in Fig. 4.88. Called a “direct-conversion” transmitter, this topology directly translates the baseband spectrum to the RF carrier by means of a “quadrature upconverter.”²⁵ The upconverter is followed by a PA and a matching network, whose role is to provide maximum power delivery to the antenna and filter out-of-band

25. Also known as a “quadrature modulator” or a “vector modulator.”

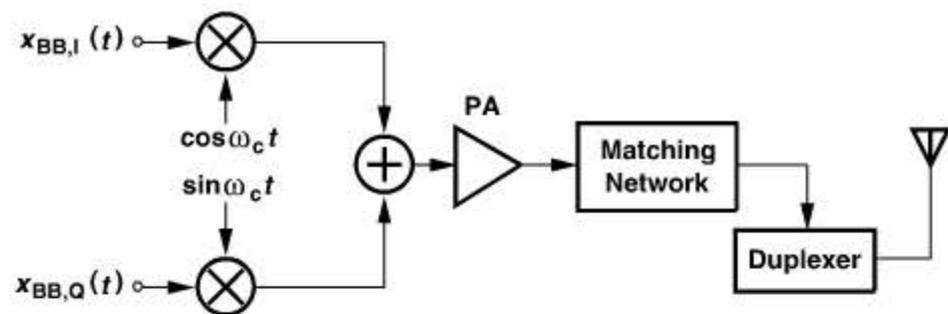


Figure 4.88 Direct-conversion transmitter.

components that result from the PA nonlinearity (Chapter 12). Note that since the baseband signal is produced in the transmitter and hence has a sufficiently large amplitude (several hundred millivolts), the noise of the mixers is much less critical here than in receivers.

The design of the TX begins with the PA. This circuit is crafted so as to deliver the required power to the antenna (and with adequate linearity if applicable). Employing large transistors to carry high currents, the PA exhibits a large input capacitance. Thus, a predriver is typically interposed between the upconverter and the PA to serve as a buffer.

Example 4.37

A student decides to omit the predriver and simply “scale up” the upconverter so that it can drive the PA directly. Explain the drawback of this approach.

Solution:

In order to scale up the upconverter, the width and bias current of each transistor are scaled up, and the resistor and inductor values are proportionally scaled down. For example, if the upconverter is modeled as a transconductance G_m and an output resistance R_{out} (Fig. 4.89),²⁶ then R_{out} can be reduced to yield adequate bandwidth with the input capacitance of the PA, and G_m can be enlarged to maintain a constant $G_m R_{out}$ (i.e., constant voltage swings). In practice, the upconverter employs a resonant LC load, but the same principles still apply.

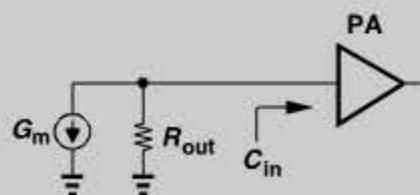


Figure 4.89 Scaling of quadrature upconverter to drive the PA.

The scaling of the transistors raises the capacitances seen at the baseband and LO ports of the mixers in Fig. 4.88. The principal issue here is that the LO now sees a large load capacitance, requiring its own buffers. Also, the two mixers consume a higher power.

Among various TX architectures studied in this chapter, the direct-conversion approach provides the most compact solution and a relatively “clean” output. That is, the output spectrum contains only the desired signal around the carrier frequency (and its harmonics) but no spurious components—an attribute similar to that of direct-conversion receivers. Nonetheless, direct upconversion entails other issues that demand attention.

I/Q Mismatch We noted in Section 4.2 that I/Q mismatch in direct-conversion receivers results in “cross-talk” between the quadrature baseband outputs or, equivalently, distortion in the constellation. We expect a similar effect in the TX counterpart. For example, as derived in Chapter 3, a phase mismatch of $\Delta\theta$ in the upconverter shifts the constellation points of a QPSK signal to $I = \pm \cos \Delta\theta$ and $Q = \pm 1 \pm \sin \Delta\theta$, just as observed for a direct-conversion receiver. The results obtained in Chapter 3 can be extended to include amplitude mismatch by writing

$$x(t) = \alpha_1(A_c + \Delta A_c) \cos(\omega_c t + \Delta\theta) + \alpha_2 A_c \sin \omega_c t \quad (4.118)$$

$$= \alpha_1(A_c + \Delta A_c) \cos \Delta\theta \cos \omega_c t + [\alpha_2 A_c - \alpha_1(A_c + \Delta A_c) \sin \Delta\theta] \sin \omega_c t. \quad (4.119)$$

Since α_1 and α_2 assume ± 1 values, the normalized coefficients of $\cos \omega_c t$ and $\sin \omega_c t$ appear as follows for the four points in the constellation:

$$\beta_1 = + \left(1 + \frac{\Delta A_c}{A_c}\right) \cos \Delta\theta, \quad \beta_2 = 1 - \left(1 + \frac{\Delta A_c}{A_c}\right) \sin \Delta\theta \quad (4.120)$$

$$\beta_1 = + \left(1 + \frac{\Delta A_c}{A_c}\right) \cos \Delta\theta, \quad \beta_2 = -1 - \left(1 + \frac{\Delta A_c}{A_c}\right) \sin \Delta\theta \quad (4.121)$$

$$\beta_1 = - \left(1 + \frac{\Delta A_c}{A_c}\right) \cos \Delta\theta, \quad \beta_2 = 1 + \left(1 + \frac{\Delta A_c}{A_c}\right) \sin \Delta\theta \quad (4.122)$$

$$\beta_1 = - \left(1 + \frac{\Delta A_c}{A_c}\right) \cos \Delta\theta, \quad \beta_2 = -1 + \left(1 + \frac{\Delta A_c}{A_c}\right) \sin \Delta\theta. \quad (4.123)$$

The reader is encouraged to compute the error vector magnitude (Chapter 3) of this constellation.

Another approach to quantifying the I/Q mismatch in a transmitter involves applying two tones $V_0 \cos \omega_{in} t$ and $V_0 \sin \omega_{in} t$ to the I and Q inputs in Fig. 4.88 and examining the output spectrum. In the ideal case, the output is simply given by $V_{out} = V_0 \cos \omega_{in} t \cos \omega_c t - V_0 \sin \omega_{in} t \sin \omega_c t = V_0 \cos(\omega_c + \omega_{in})t$. On the other hand, in the presence of a (relative) gain mismatch of ε and a phase imbalance of $\Delta\theta$, we have

$$V_{out}(t) = V_0(1 + \varepsilon) \cos \omega_{in} t \cos(\omega_c t + \Delta\theta) - V_0 \sin \omega_{in} t \sin \omega_c t \quad (4.124)$$

$$\begin{aligned} &= \frac{V_0}{2} [(1 + \varepsilon) \cos \Delta\theta + 1] \cos(\omega_c t + \omega_{in})t \\ &\quad - \frac{V_0}{2} (1 + \varepsilon) \sin \Delta\theta \sin(\omega_c + \omega_{in})t \\ &\quad + \frac{V_0}{2} [(1 + \varepsilon) \cos \Delta\theta - 1] \cos(\omega_c - \omega_{in})t \\ &\quad - \frac{V_0}{2} (1 + \varepsilon) \sin \Delta\theta \sin(\omega_c - \omega_{in})t. \end{aligned} \quad (4.125)$$

26. In this conceptual model, we omit the frequency translation inherent in the upconverter.

It follows that the power of the unwanted sideband at $\omega_c - \omega_{in}$ divided by that of the wanted sideband at $\omega_c + \omega_{in}$ is given by

$$\frac{P_-}{P_+} = \frac{[(1 + \varepsilon) \cos \Delta\theta - 1]^2 + (1 + \varepsilon)^2 \sin^2 \Delta\theta}{[(1 + \varepsilon) \cos \Delta\theta + 1]^2 + (1 + \varepsilon)^2 \sin^2 \Delta\theta} \quad (4.126)$$

$$= \frac{(1 + \varepsilon)^2 - 2(1 + \varepsilon) \cos \Delta\theta + 1}{(1 + \varepsilon)^2 + 2(1 + \varepsilon) \cos \Delta\theta + 1}, \quad (4.127)$$

which is similar to the image-rejection ratio expression [Eq. (4.77)]. We may even call the unwanted sideband the “image” of the wanted sideband with respect to the carrier frequency. In practice, a P_-/P_+ of roughly -30 dB is sufficient to ensure negligible distortion of the cancellation, but the exact requirement depends on the type of modulation.

Example 4.38

Compute the average power of $V_{out}(t)$ in Eq. (4.125).

Solution:

We add P_- and P_+ and multiply the result by $V_0^2/4$. If $\varepsilon \ll 1$, then

$$\overline{V_{out}^2(t)} = \frac{V_0^2}{2}(1 + \varepsilon). \quad (4.128)$$

Interestingly, the output power is independent of the phase mismatch.

If the raw I/Q matching of circuits in a transmitter is inadequate, some means of calibration can be employed. To this end, the gain and phase mismatch must first be measured and subsequently corrected. Can we use the power of the unwanted sideband as a symptom of the I/Q mismatch? Yes, but it is difficult to *measure* this power in the presence of the large wanted sideband: the two sidebands are too close to each other to allow suppressing the larger one by tens of decibels by means of a filter.

Let us now apply a single sinusoid to both inputs of the upconverter (Fig. 4.90). The output emerges as

$$V_{out3}(t) = V_0(1 + \varepsilon) \cos \omega_{in} t \cos(\omega_c t + \Delta\theta) - V_0 \cos \omega_{in} t \sin \omega_c t \quad (4.129)$$

$$= V_0 \cos \omega_{in} t (1 + \varepsilon) \cos \Delta\theta \cos \omega_c t - V_0 \cos \omega_{in} t [(1 + \varepsilon) \sin \Delta\theta + 1] \sin \omega_c t. \quad (4.130)$$

It can be shown that the output contains two sidebands of equal amplitudes and carries an average power equal to

$$\overline{V_{out3}^2(t)} = V_0^2 [1 + (1 + \varepsilon) \sin \Delta\theta]. \quad (4.131)$$

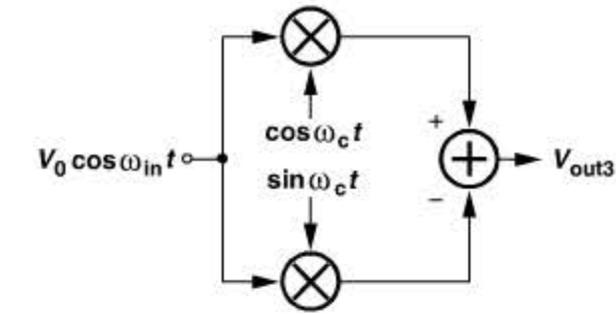


Figure 4.90 Quadrature upconverter sensing a single sinusoid to reveal phase mismatch.

We observe that ε is forced to zero as described above, then

$$\overline{V_{out3}^2} - 2\overline{V_{out1}^2} = \sin \Delta\theta. \quad (4.132)$$

Thus, the calibration of phase mismatch proceeds to drive this quantity to zero.

For gain mismatch calibration, we perform two consecutive tests. Depicted in Fig. 4.91, the tests entail applying a sinusoid to one baseband input while the other is set to zero. For the case in Fig. 4.91(a),

$$V_{out1}(t) = V_0(1 + \varepsilon) \cos \omega_{in} t \cos(\omega_c t + \Delta\theta), \quad (4.133)$$

yielding an average power of

$$\overline{V_{out1}^2(t)} = \frac{V_0^2}{2} + V_0^2 \varepsilon. \quad (4.134)$$

In Fig. 4.91(b), on the other hand,

$$V_{out2}(t) = V_0 \cos \omega_{in} t \sin \omega_c t, \quad (4.135)$$

producing

$$\overline{V_{out2}^2(t)} = \frac{V_0^2}{2}. \quad (4.136)$$

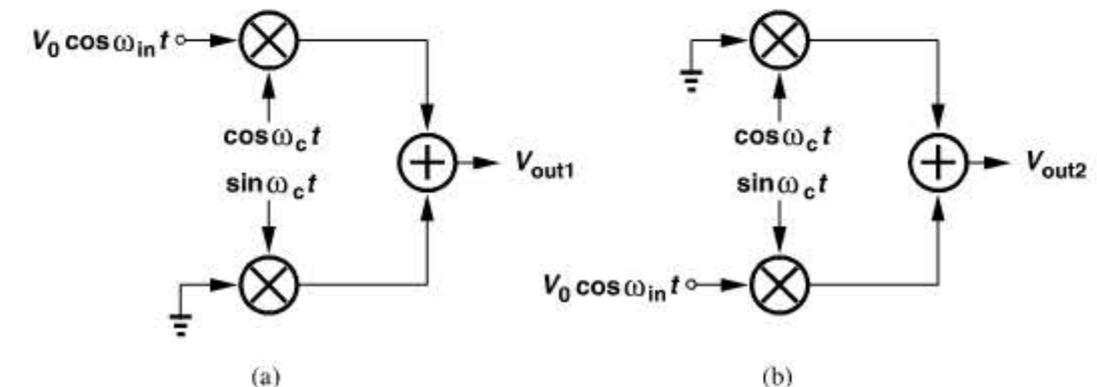


Figure 4.91 Quadrature upconverter sensing a cosine at (a) the I input, (b) at the Q input.

That is,

$$\overline{V_{out1}^2(t)} - \overline{V_{out2}^2(t)} = V_0^2 \varepsilon, \quad (4.137)$$

suggesting that the gain mismatch can be adjusted so as to drive this difference to zero.

The measurement of $\overline{V_{out3}^2(t)}$ and $\overline{V_{out1}^2(t)} - \overline{V_{out2}^2(t)}$ in the above tests requires a relatively high resolution. For example, a residual phase mismatch of $\Delta\theta = 1^\circ$ translates to $\sin \Delta\theta = 1.75\%$, dictating a resolution of about 7–8 bits in the ADC that digitizes $\overline{V_{out3}^2(t)}$ in Eq. (4.131).

We should remark that dc offsets in the baseband may affect the accuracy of I/Q calibration. As explained below, another effect, “carrier leakage,” may also require the removal of dc offsets prior to I/Q calibration.

Carrier Leakage The analog baseband circuitry producing the quadrature signals in the transmitter of Fig. 4.88 exhibits dc offsets, and so does the baseband port of each upconversion mixer. Consequently, the output signal appears as

$$V_{out}(t) = [A(t) \cos \phi + V_{OS1}] \cos \omega_c t - [A(t) \sin \phi + V_{OS2}] \sin \omega_c t, \quad (4.138)$$

where V_{OS1} and V_{OS2} denote the total dc offsets referred to the input port of the mixers. The upconverter output therefore contains a fraction of the *unmodulated* carrier:

$$V_{out}(t) = A(t) \cos(\omega_c t + \phi) + V_{OS1} \cos \omega_c t - V_{OS2} \sin \omega_c t. \quad (4.139)$$

Called “carrier leakage,” and quantified as

$$\text{Relative Carrier Leakage} = \frac{\sqrt{V_{OS1}^2 + V_{OS2}^2}}{\sqrt{A^2(t)}}, \quad (4.140)$$

this phenomenon leads to two adverse effects. First, it distorts the signal constellation, raising the error vector magnitude at the TX output. For example, if $V_{out}(t)$ represents a QPSK signal,

$$V_{out}(t) = \alpha_1(V_0 + V_{OS1}) \cos \omega_c t - \alpha_2(V_0 + V_{OS2}) \sin \omega_c t, \quad (4.141)$$

and is applied to an ideal direct-conversion receiver, then the baseband quadrature outputs suffer from dc offsets, i.e., horizontal and vertical shifts in the constellation (Fig. 4.92).

The second effect manifests itself if the output power of the transmitter must be varied across a wide range by varying the amplitude of the baseband signals. For example, as described in Chapter 3, CDMA mobiles must lower their transmitted power as they come closer to the base station so as to avoid the near-far effect. Figure 4.93(a) conceptually depicts the power control loop. The base station measures the power level received from the mobile and, accordingly, requests the mobile to adjust its transmitted power. With a short distance between the two, the mobile output power must be reduced to very low values, yielding the spectrum shown in Fig. 4.93(b) in the presence of carrier leakage. In this case, the carrier power dominates, making it difficult to measure the actual signal power. This

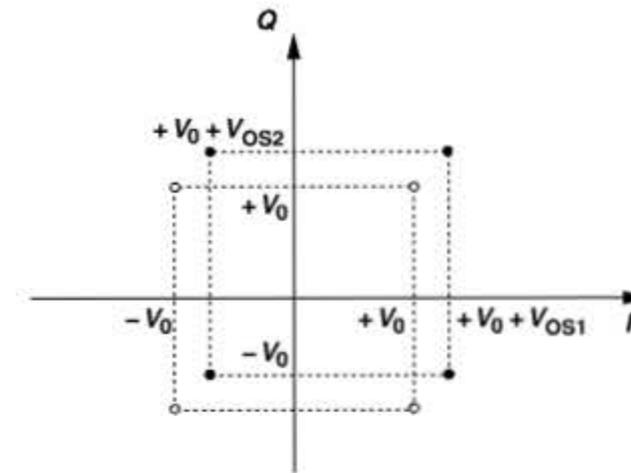


Figure 4.92 Effect of carrier feedthrough on received signal spectrum.

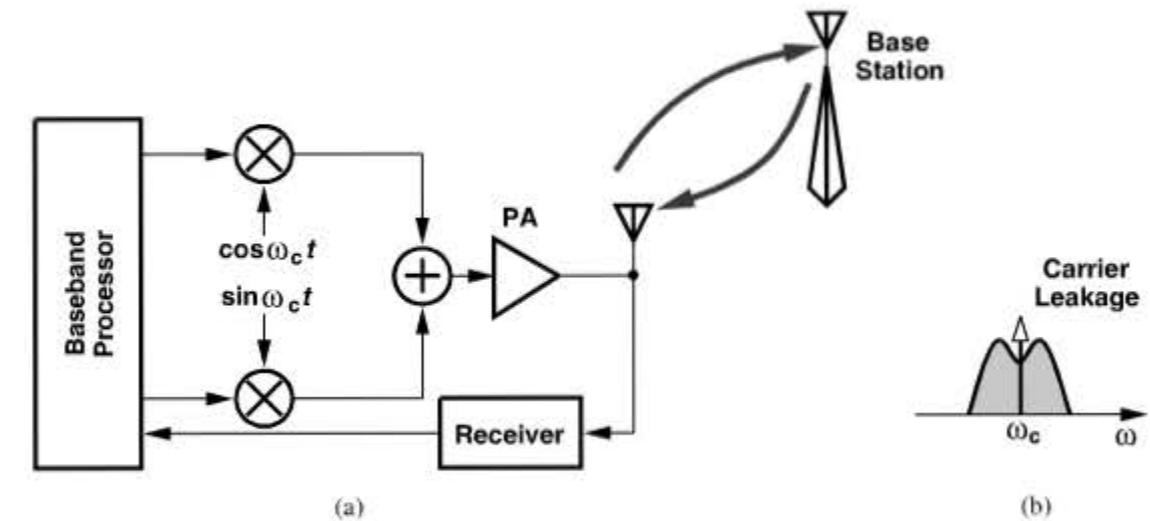


Figure 4.93 (a) Power control feedback loop in CDMA, (b) effect of carrier leakage.

issue arises if the mobile output power is adjusted by varying the baseband swings but not if the PA itself is adjusted.

In order to reduce the carrier leakage, Eq. (4.140) suggests that the baseband signal swing, $A(t)$, must be chosen sufficiently large. However, as $A(t)$ increases, the input port of the upconversion mixers becomes more nonlinear. A compromise is therefore necessary. In stringent applications, the offsets must be trimmed to minimize the carrier leakage. As illustrated in Fig. 4.94, two DACs are tied to the baseband ports of the TX²⁷ and a power detector (e.g., a rectifier or an envelope detector) monitors the output level, and its output is digitized. During carrier leakage cancellation, the baseband processor produces a zero output so that the detector measures only the leakage. Thus, the loop consisting of the TX, the detector, and the DACs drives the leakage toward zero, with the final settings of the DACs stored in the register.

27. The DACs may be embedded within the mixers themselves (Chapter 6).

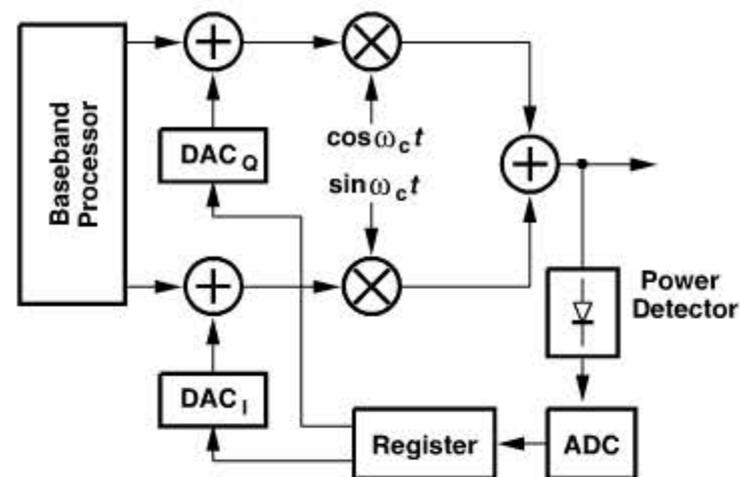


Figure 4.94 Reduction of carrier leakage by baseband offset control.

Example 4.39

Is it possible to cancel the carrier leakage by means of a single DAC?

Solution:

No, it is not. Eq. (4.139) implies that no choice of V_{OS1} or V_{OS2} can force $V_{OS1} \cos \omega_c t - V_{OS2} \sin \omega_c t$ to zero if the other remains finite.

How should the two DACs be adjusted so that the loop in Fig. 4.94 converges? An adaptive loop such as the least mean square (LMS) algorithm can perform this task. Alternatively, an “exhaustive” search can arrive at the optimum settings; e.g., for 8-bit DACs, only 256×256 possible combinations exist, and the system can try all to determine which combination yields the lowest leakage at the output. In this procedure, the system begins with, say, zero settings for both DACs, measures the carrier leakage, *memorizes* this value, increments one DAC by 1 LSB, measures the leakage again, and compares the result with the previous value. The new settings replace the previous ones if the new leakage is lower.

Mixer Linearity Unlike downconversion mixers in a receiver, upconversion mixers in a transmitter sense no interferers. However, excessive nonlinearity in the baseband port of upconversion mixers can corrupt the signal or raise the adjacent channel power (Chapter 3). As an example, consider the GMSK signal expressed by Eq. (4.112) and suppose the baseband I/Q inputs experience a nonlinearity given by $\alpha_1 x + \alpha_3 x^3$. The upconverted signal assumes the form [15]

$$\begin{aligned} V_{out}(t) &= (\alpha_1 A \cos \phi + \alpha_3 A^3 \cos^3 \phi) \cos \omega_c t - (\alpha_1 A \sin \phi + \alpha_3 A^3 \sin^3 \phi) \sin \omega_c t \quad (4.142) \\ &= \left(\alpha_1 A + \frac{3}{4} \alpha_3 A^3 \right) \cos(\omega_c t + \phi) + \frac{\alpha_3 A^3}{4} \cos(\omega_c t - 3\phi). \quad (4.143) \end{aligned}$$

The second term also represents a GMSK signal but with a threefold modulation index, thereby occupying a larger bandwidth. Simulations indicate that this effect becomes negligible if the baseband port of the mixers experiences a nonlinearity less than 1% for the specified baseband swings [15].

For variable-envelope signals, $A^3(t)$ appears in both terms of Eq. (4.143), exacerbating the effect. The required mixer linearity is typically determined by simulations. However, in most cases (i.e., in a good design), as the baseband signal swings increase, the PA output begins to compress *before* the mixer nonlinearity manifests itself. This is explained below.

TX Linearity The linearity of transmitters must be chosen according to the spectral regrowth (adjacent channel power) requirements and/or the tolerable distortion of the signal to be transmitted. As mentioned in Chapter 3, both of these effects become critical for variable-envelope modulation schemes. We defer the former to Chapter 12 and deal with the latter here.

The distortion of a variable-envelope signal is typically characterized by the compression that it experiences. As shown in Fig. 4.95, the signal is subjected to a nonlinear characteristic in simulations and it is determined how close its level can come to the 1-dB compression point while the degradation in the constellation or the bit error rate is negligible. For example, the average power of the 64-QAM OFDM signal in 802.11a must remain about 8 dB below P_{1dB} of a given circuit. We say the circuit must operate at “8-dB back-off.” In other words, if a peak swing of V_0 places the circuit at the 1-dB compression point, then the average signal swing must not exceed $V_0/2.51$.

In a TX chain, the signal may experience compression in any of the stages. Consider the example depicted in Fig. 4.96, where the signal levels (in dB) along the chain are also shown. Since the largest voltage swing occurs at the output of the PA, this stage dominates the compression of the TX; i.e., in a good design, the preceding stages must remain well below compression as the PA output approaches P_{1dB} . To so ensure, we must *maximize* the gain of the PA and minimize the output swing of the predriver and the stages preceding it. This requirement places additional burden on the PA design (Chapters 12 and 13).

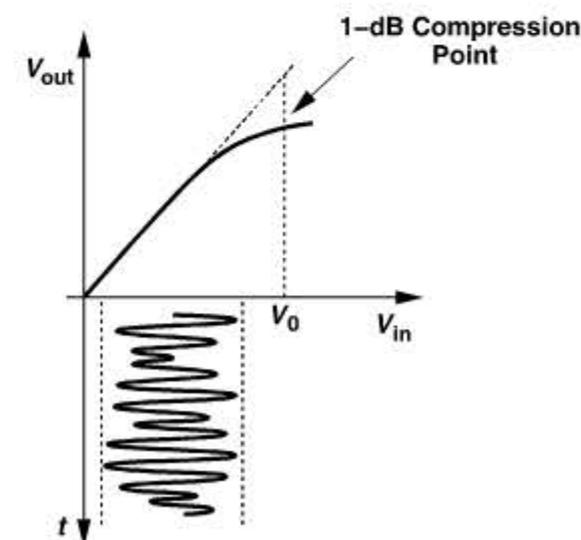


Figure 4.95 Variable-envelope signal applied to a compressive system.

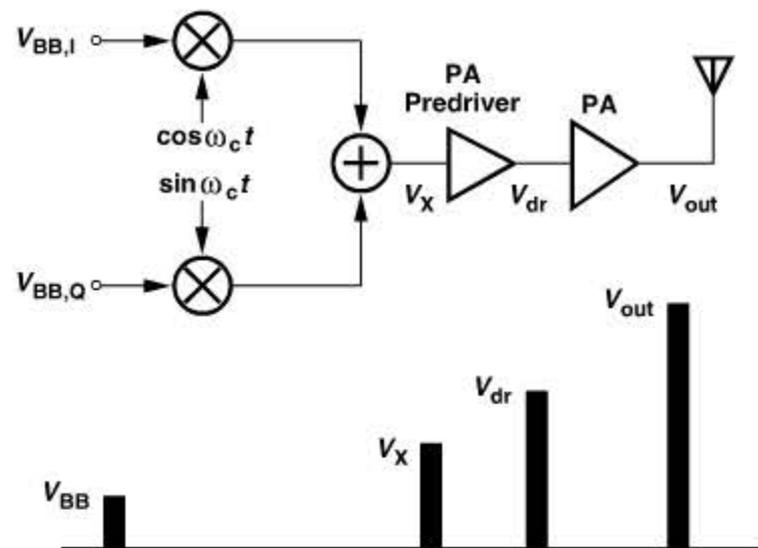


Figure 4.96 Level chart for the signals along a TX chain.

Example 4.40

If the predriver and the PA exhibit third-order characteristics, compute the 1-dB compression point of the cascade of the two.

Solution:

Assuming a nonlinearity of $\alpha_1 x + \alpha_3 x^3$ for the predriver and $\beta_1 x + \beta_3 x^3$ for the PA, we write the PA output as

$$y(t) = \beta_1(\alpha_1 x + \alpha_3 x^3) + \beta_3(\alpha_1 x + \alpha_3 x^3)^3 \quad (4.144)$$

$$= \beta_1 \alpha_1 x + (\beta_1 \alpha_3 + \beta_3 \alpha_1^3) x^3 + \dots \quad (4.145)$$

If the first two terms are dominant, then the input 1-dB compression point is given by the coefficients of x and x^3 as follows:

$$A_{1dB,in} = \sqrt{0.145 \left| \frac{\beta_1 \alpha_1}{\beta_1 \alpha_3 + \beta_3 \alpha_1^3} \right|}. \quad (4.146)$$

The reader is encouraged to consider the special cases $\beta_3 = 0$ or $\alpha_3 = 0$ and justify the results intuitively. It is interesting to note that if $\beta_1 \alpha_3 = -\beta_3 \alpha_1^3$, the coefficient of x^3 falls to zero and $A_{1dB,in} \rightarrow \infty$. This is because the compression in one stage is cancelled by the expansion in the other.

In transmitters, the *output* power is of interest, suggesting that the compression behavior must also be quantified at the output. Since at $A_{1dB,in}$, the output level is 1 dB below its ideal value, we simply multiply $A_{1dB,in}$ by the total gain and reduce the result by 1 dB so as to determine the output compression point:

$$A_{1dB,out} = A_{1dB,in} \times |\alpha_1 \beta_1| \times \frac{1}{1.12}, \quad (4.147)$$

Example 4.40 (Continued)

where the factor $1/1.12$ accounts for the 1-dB gain reduction. It follows that

$$A_{1dB,out} = \frac{0.34 |\alpha_1 \beta_1| \sqrt{|\beta_1 \alpha_1|}}{\sqrt{\beta_1 \alpha_3 + \beta_3 \alpha_1^3}}. \quad (4.148)$$

Oscillator Pulling While the issues described above apply to most transmitter architectures, another one becomes particularly critical in direct-conversion topologies. As illustrated in Fig. 4.97(a), the PA output exhibits very large swings ($20 V_{pp}$ for 1 W delivered to a $50\text{-}\Omega$ load), which couple to various parts of the system through the silicon substrate, package parasitics, and traces on the printed-circuit board. Thus, it is likely that an appreciable fraction of the PA output couples to the local oscillator. Even if the PA is off-chip, the PA driver may still pull the LO. Note that the center frequency of the PA output spectrum is equal to ω_{LO} in direct-conversion transmitters.

Let us consider a “sliver” of the output spectrum centered at $\omega_{LO} + \Delta\omega$ and model it by an impulse of equal energy [Fig. 4.97(b)]. We therefore inquire what happens if a sinusoid at a frequency of $\omega_1 = \omega_{LO} + \Delta\omega$ is “injected” into an oscillator operating at a frequency of ω_{LO} , where $\Delta\omega \ll \omega_{LO}$. Called “injection pulling,” this phenomenon has been studied extensively [16, 17] and is analyzed in Chapter 8. In such a condition, the output phase of the oscillator, ϕ_{out} , is modulated *periodically*. In fact, as depicted in Fig. 4.98(a), ϕ_{out} remains around 90° (with respect to the input phase) for part of the period, subsequently experiencing a rapid 360° rotation. The input and output waveforms therefore appear as in Fig. 4.98(b). It can be proved that the output spectrum is heavily asymmetric [Fig. 4.98(c)], with most of the impulses located *away* from the input frequency, $\omega_{inj} (= \omega_{LO} + \Delta\omega)$. Note that the spacing between the impulses is equal to the frequency of the phase variation in Fig. 4.98(a) and *not* equal to $\Delta\omega$.²⁸

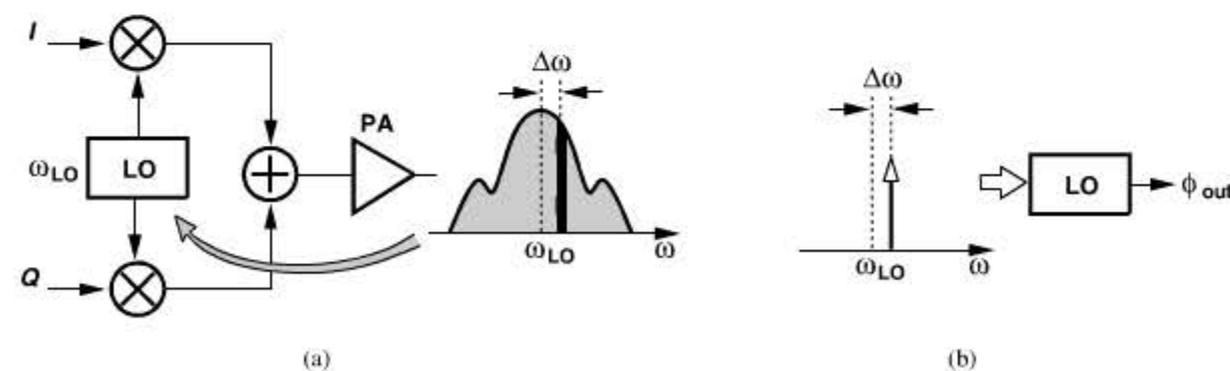


Figure 4.97 (a) Injection pulling of LO by PA in a direct-conversion TX, (b) conceptual illustration of injection into an oscillator.

28. Pulling can also occur if the injected signal frequency is close to a *harmonic* of the oscillator frequency, e.g., in the vicinity of $2\omega_{LO}$. We call this effect “superharmonic pulling.”

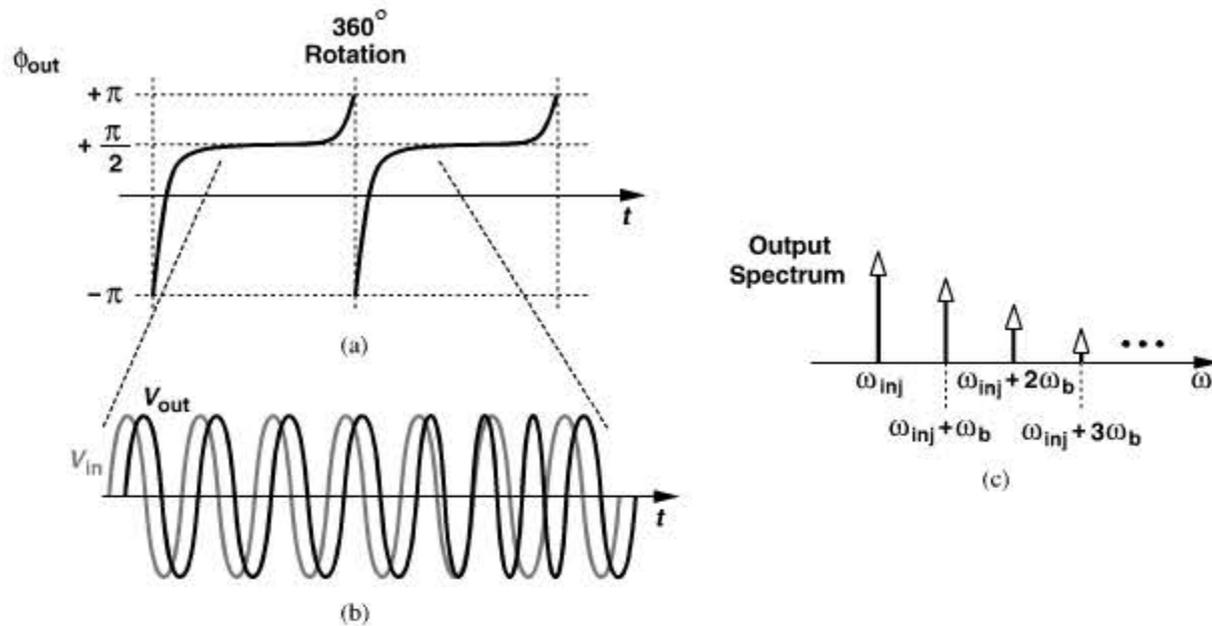


Figure 4.98 (a) Behavior of LO output phase in the presence of injection pulling, (b) cycle slips in time domain, (c) resulting spectrum.

At what output power level does injection pulling become significant? The answer depends on several factors: (1) the internal voltage and current swings of the LO (the larger they are, the less is the effect of pulling); (2) the Q of the tank used in the oscillator; (3) whether the PA output is differential, in which case its coupling to the LO is 30–40 dB lower than if it is single-ended;²⁹ (4) how much the feedback loop controlling the LO (the synthesizer) counteracts the pulling [17]; (5) the symmetry of the layout and the type of packaging. Nonetheless, for typical designs, as the PA output exceeds 0 dBm, injection pulling may prove serious.

In order to avoid injection pulling, the PA output frequency and the oscillator frequency must be made sufficiently different (e.g., by more than 20%), an impossible task in the architecture of Fig. 4.97. This principle has led to a number of transmitter architectures and frequency plans that are described below.

Noise in RX Band As explained in Chapter 3, some standards (e.g., GSM) specify the maximum noise that a TX can transmit in the RX band. In a direct-conversion transmitter, the baseband circuits, the upconverter, and the PA may create significant noise in the RX band. To resolve this issue, “offset-PLL” transmitters can be used (Chapter 9).

4.3.3 Modern Direct-Conversion Transmitters

Most of today’s direct-conversion transmitters avoid an oscillator frequency equal to the PA output frequency. To avoid confusion, we call the former the LO frequency, ω_{LO} , and the

29. This is true only if the differential PA incorporates “single-ended” inductors rather than one symmetric inductor (Chapter 7).

latter, the carrier frequency, ω_c . The task is accomplished by choosing ω_{LO} sufficiently far from ω_c and deriving ω_c from ω_{LO} by operations such as frequency division and mixing.

Figure 4.99 depicts a common example where $\omega_{LO} = 2\omega_c$. A divide-by-2 circuit follows the LO, thereby generating $\omega_{LO}/2$ with quadrature phases. This architecture is popular for two reasons: (1) injection pulling is greatly reduced, and (2) the divider readily provides quadrature phases of the carrier, an otherwise difficult task (Chapter 8).

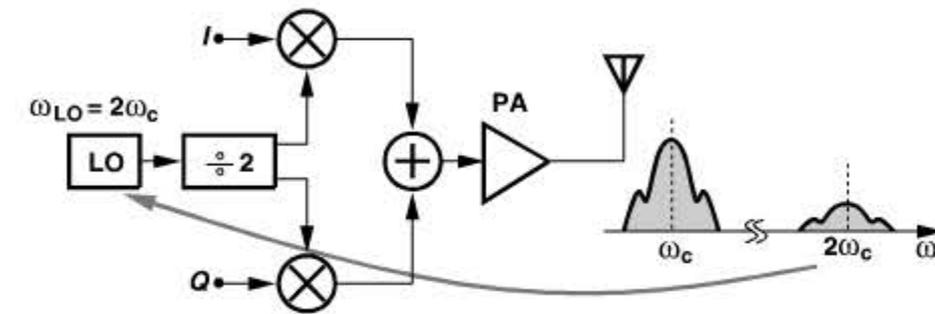


Figure 4.99 Use of an LO running at twice the carrier frequency to minimize LO pulling.

The architecture of Fig. 4.99 does not entirely eliminate injection pulling. Since the PA nonlinearity produces a finite amount of power at the *second* harmonic of the carrier, the LO may still be pulled. Nonetheless, proper layout and isolation techniques can suppress this effect.

Example 4.41

Is it possible to choose $\omega_{LO} = \omega_c/2$ and use a frequency doubler to generate ω_c ?

Solution:

It is possible, but the doubler typically does not provide quadrature phases, necessitating additional quadrature generation stages. Figure 4.100 shows an example where the doubler output is applied to a polyphase filter (Section 4.2.5). The advantage of this architecture is that no harmonic of the PA output can pull the LO. The serious disadvantage is that the doubler and the polyphase filter suffer from a high loss, requiring the use of power-hungry buffers.

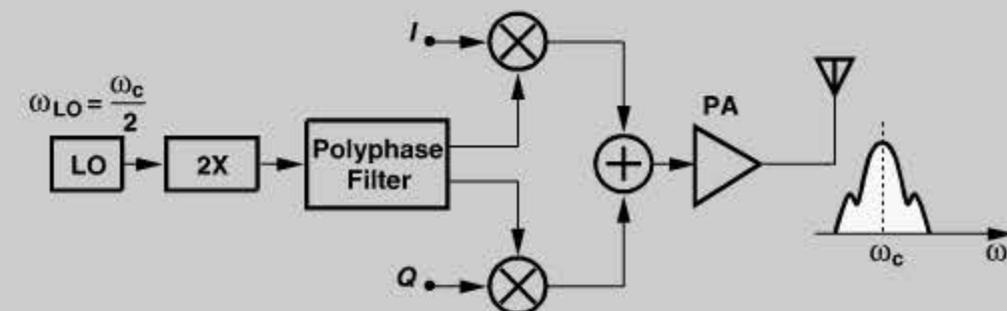


Figure 4.100 Use of an LO running at half the carrier frequency to minimize LO pulling.

The principal drawback of the architecture of Fig. 4.99 stems from the speed required of the divider. Operating at twice the carrier frequency, the divider may become the speed bottleneck of the entire transceiver. (We deal with divider design in Chapter 10.) Nevertheless, as seen in the following discussion, other transmitter architectures suffer from more serious drawbacks. Thus, even a substantial effort on divider design to enable this architecture is well justified.

Another approach to deriving frequencies is through the use of mixing. For example, mixing the outputs of two oscillators operating at ω_1 and ω_2 can yield $\omega_1 + \omega_2$ or $\omega_1 - \omega_2$. Nonetheless, as with the receivers studied in Section 4.2, it is desirable to employ a *single* oscillator and utilize division to obtain subharmonics. To this end, let us consider the arrangement shown in Fig. 4.101(a), where the oscillator frequency is divided by 2 and the two outputs are mixed. The result contains components at $\omega_1 \pm \omega_1/2$ with equal magnitudes. We may call one the “image” of the other with respect to ω_1 .

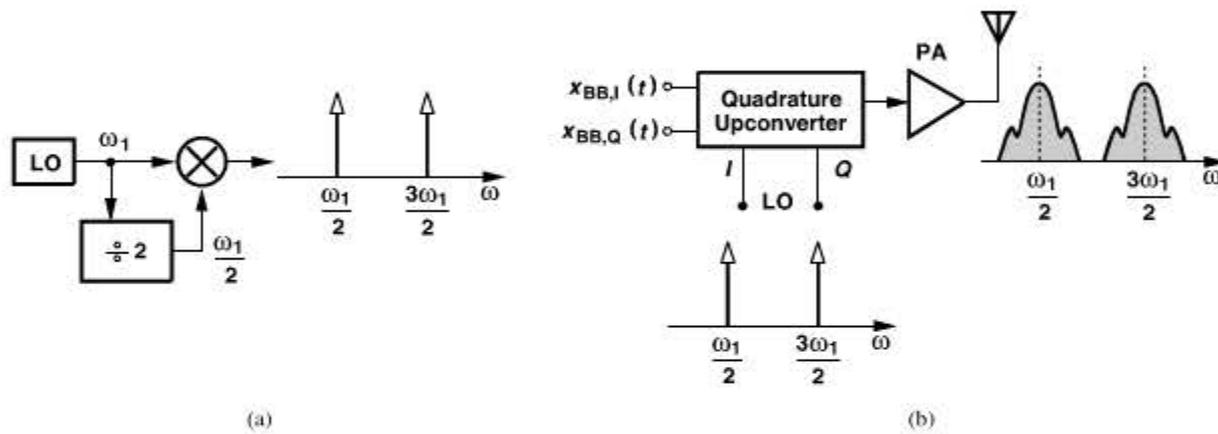


Figure 4.101 (a) Mixing of an LO output with half of its frequency, (b) effect of two sidebands on transmitter output.

Can both components be retained? In a transmitter using such an LO waveform, the upconverter output would contain, with equal power, the signal spectrum at both carrier frequencies [Fig. 4.101(b)]. Thus, half of the power delivered to the antenna is wasted. Furthermore, the power transmitted at the unwanted carrier frequency corrupts communication in other channels or bands. One component (e.g., that at $\omega_1/2$) must therefore be suppressed.

It is difficult to remove the unwanted carrier by means of filtering because the two frequencies differ by only a factor of 3. For example, in Problem 4.24 we show that a second-order *LC* filter resonating at $3\omega_1/2$ attenuates the component at $\omega_1/2$ by a factor of $8Q/3$. For a Q in the range of 5 to 10, this attenuation amounts to 25 to 30 dB, sufficient for minimizing the waste of power in the unwanted sideband but inadequate for avoiding corruption of other channels. The other issue is that the output in Fig. 4.101(a) is not available in quadrature form.

An alternative method of suppressing the unwanted sideband incorporates “single-sideband” (SSB) mixing. Based on the trigonometric identity $\cos \omega_1 t \cos \omega_2 t - \sin \omega_1 t \sin \omega_2 t = \cos(\omega_1 + \omega_2)t$ and illustrated in Fig. 4.102(a), SSB mixing involves multiplying the quadrature phases of ω_1 and ω_2 and subtracting the results—just as realized by the quadrature upconverter of Fig. 4.88. We denote an SSB mixer by the abbreviated symbol

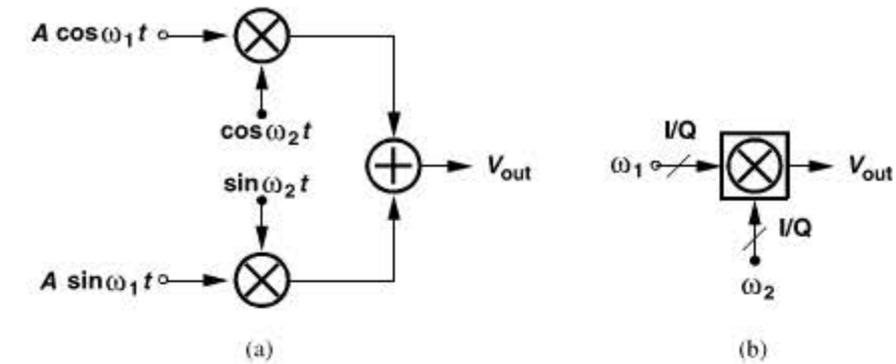


Figure 4.102 Single-sideband mixer (a) implementation, (b) simplified symbol.

shown in Fig. 4.102(b). Of course, gain and phase mismatches lead to an unwanted sideband, as expressed by Eq. (4.127). With typical mismatches, P_-/P_+ falls in the vicinity of 30 to 40 dB, and filtering by a second-order *LC* stage attenuates the sideband by another 25 to 30 dB.

In addition to the image sideband, the *harmonics* of the input frequencies also corrupt the output of an SSB mixer. For example, suppose each mixer in Fig. 4.102(a) exhibits third-order nonlinearity in the port sensing $A \sin \omega_1 t$ or $A \cos \omega_1 t$. If the nonlinearity is of the form $\alpha_1 x + \alpha_3 x^3$, the output can be expressed as

$$V_{out}(t) = (\alpha_1 A \cos \omega_1 t + \alpha_3 A^3 \cos^3 \omega_1 t) \cos \omega_2 t - (\alpha_1 A \sin \omega_1 t + \alpha_3 A^3 \sin^3 \omega_1 t) \sin \omega_2 t \quad (4.149)$$

$$= \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) \cos \omega_1 t \cos \omega_2 t - \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) \sin \omega_1 t \sin \omega_2 t + \frac{\alpha_3 A^3}{4} \cos 3\omega_1 t \cos \omega_2 t + \frac{\alpha_3 A^3}{4} \sin 3\omega_1 t \sin \omega_2 t \quad (4.150)$$

$$= \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) \cos(\omega_1 + \omega_2)t + \frac{\alpha_3 A^3}{4} \cos(3\omega_1 - \omega_2)t. \quad (4.151)$$

The output spectrum contains a spur at $3\omega_1 - \omega_2$. Similarly, with third-order nonlinearity in the mixer ports sensing $\sin \omega_2 t$ and $\cos \omega_2 t$, a component at $3\omega_2 - \omega_1$ arises at the output. The overall output spectrum (in the presence of mismatches) is depicted in Fig. 4.103.

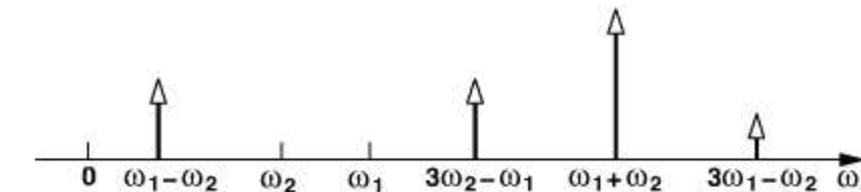


Figure 4.103 Output spectrum of SSB mixer in the presence of nonlinearity and mismatches.

Figure 4.104 shows a mixer example where the port sensing V_{in1} is linear while that driven by V_{in2} is nonlinear. As explained in Chapter 2, the circuit multiplies V_{in1} by a *square wave* toggling between 0 and 1. That is, the third harmonic of V_{in2} is only one-third of its fundamental, thus producing the strong spurs in Fig. 4.103.

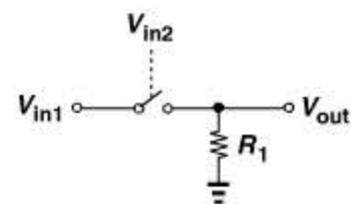


Figure 4.104 Simple mixer.

How serious are the above spurs? In a typical mixer design (Chapter 6), it is possible to linearize only one port, thus maintaining a small third harmonic in that port. The other is highly nonlinear so as to retain a reasonable gain (or loss). We therefore conclude that, between the two spurs at $3\omega_1 - \omega_2$ and $3\omega_2 - \omega_1$, only one can be reduced to acceptably low levels while the other remains only 10 dB (a factor of one-third) below the desired component. As an example, if $\omega_2 = \omega_1/2$, then $3\omega_1 - \omega_2 = 5\omega_1/2$ and $3\omega_2 - \omega_1 = \omega_1/2$; we can linearize the port sensing ω_2 to suppress the latter, but the former still requires substantial filtering.

For use in a direct-conversion TX, the SSB mixer must provide the quadrature phases of the carrier. This is accomplished by noting that $\sin \omega_1 t \cos \omega_2 t + \cos \omega_1 t \sin \omega_2 t = \sin(\omega_1 + \omega_2)t$ and duplicating the SSB mixer as shown in Fig. 4.105.

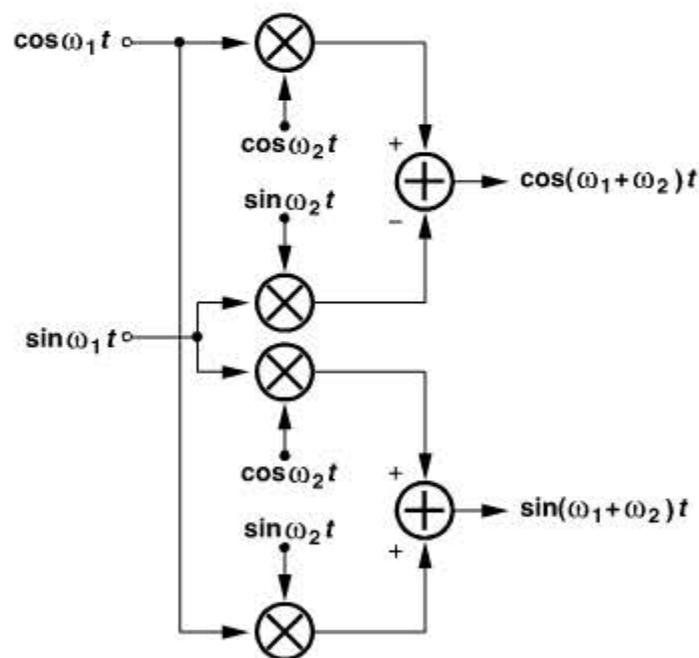


Figure 4.105 SSB mixer providing quadrature outputs.

Figure 4.106 shows a direct-conversion TX with SSB mixing for carrier generation. Since the carrier and LO frequencies are sufficiently different, this architecture remains free from injection pulling.³⁰ While suppressing the carrier sideband at $\omega_1/2$, this architecture

30. This is not strictly correct because the second harmonic of the PA output is also the third harmonic of the LO, potentially causing “superharmonic” pulling.

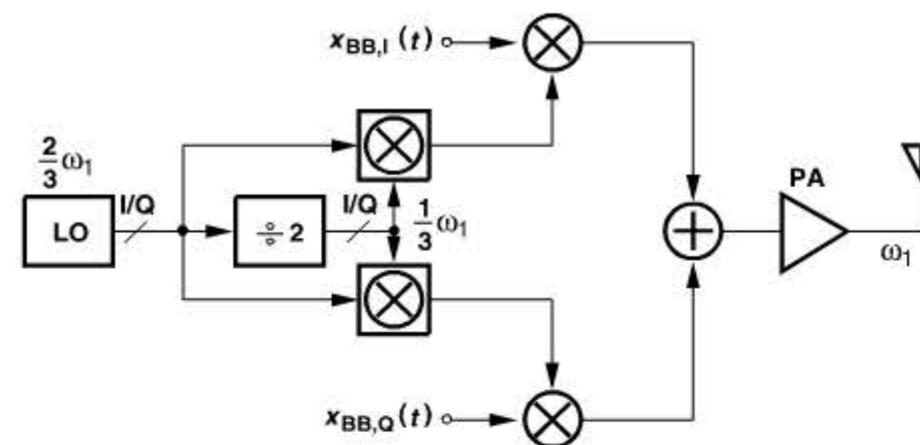


Figure 4.106 Direct-conversion TX using SSB mixing in LO path.

presents two drawbacks: (1) the spurs at $5\omega_1/2$ and other harmonic-related frequencies prove troublesome, and (2) the LO must provide quadrature phases, a difficult issue (Chapter 8).

Example 4.42

A student replaces the $\div 2$ circuit in Fig. 4.106 with a $\div 4$ topology. Analyze the unwanted components in the carrier.

Solution:

Upon mixing ω_1 and $\omega_1/4$, the SSB mixer generates $5\omega_1/4$ and, due to mismatches, $3\omega_1/4$. In the previous case, these values were given by $3\omega_1/2$ and $\omega_1/2$, respectively. Thus, filtering the unwanted sideband is more difficult in this case because it is closer to the wanted sideband.

As for the effect of harmonics, the output contains spurs at $3\omega_1 - \omega_2$ and $3\omega_2 - \omega_1$, which are respectively equal to $11\omega_1/4$ and $\omega_1/4$ if $\omega_2 = \omega_1/4$. The spur at $11\omega_1/4$ remains slightly higher than its counterpart in the previous case ($5\omega_1/2$), while that at $\omega_1/4$ is substantially lower and can be filtered more easily. Figure 4.107 summarizes the output components.

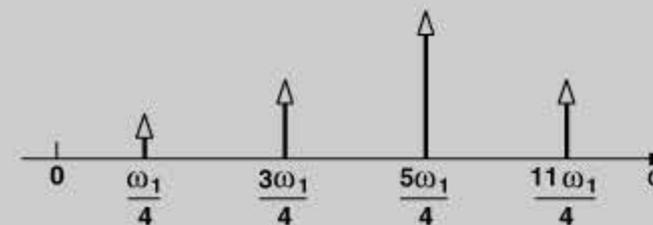


Figure 4.107 Output spurs with a divide-by-4 circuit used in LO mixing.

4.3.4 Heterodyne Transmitters

Another approach to avoiding injection pulling involves performing the signal upconversion in *two* steps so that the LO frequency remains far from the PA output spectrum. Shown in Fig. 4.108, such a TX topology is the “dual” of the heterodyne receiver studied in Section 4.2.1. Here, the baseband I and Q signals are upconverted to an IF of ω_1 , and the result is mixed with ω_2 and hence translated to a carrier frequency of $\omega_1 + \omega_2$. Since the second mixer also produces an output at $\omega_1 - \omega_2$, a band-pass filter follows this stage. As with the receiver counterpart, one advantage of this architecture is that the I/Q upconversion occurs at a significantly lower frequency than the carrier, exhibiting smaller gain and phase mismatches. The equations quantifying the effect of mismatches are the same as those derived in Section 4.3.2 for the direct-conversion TX.

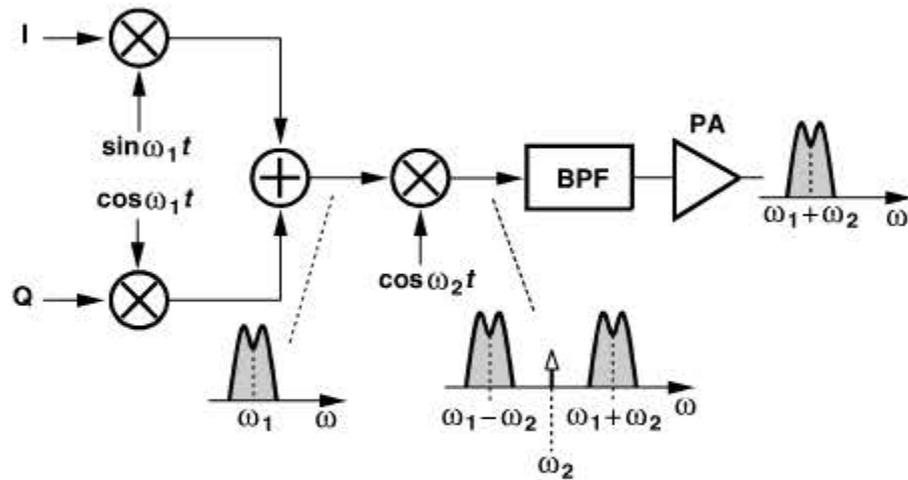


Figure 4.108 Two-step TX.

In analogy with the sliding-IF receiver architecture of Fig. 4.26(a), we eliminate the first oscillator in the above TX and derive the required phases from the second oscillator (Fig. 4.109). The carrier frequency is thus equal to $3\omega_1/2$. Let us study the effect of nonidealities in this architecture. We call the LO waveforms at $\omega_1/2$ and ω_1 the first and second LOs, respectively.

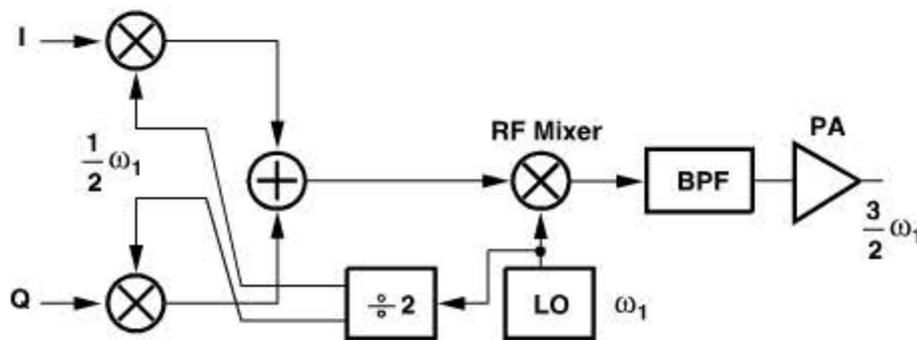


Figure 4.109 Sliding-IF TX.

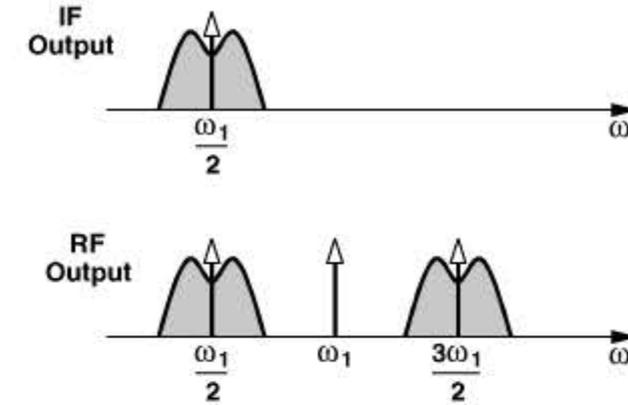


Figure 4.110 Carrier leakage in heterodyne TX.

Carrier Leakage The dc offsets in the baseband yield a component at $\omega_1/2$ at the output of the quadrature upconverter, and the dc offset at the input of the RF mixer produces another component at ω_1 (Fig. 4.110). The former can be minimized as described in Section 4.3.2. The latter, and the lower sideband at $\omega_1/2$, must be removed by filtering. The leakage at ω_1 is closer to the upper sideband than the lower sideband is, but it is also much smaller than the lower sideband. Thus, the filter following the RF mixer must be designed to attenuate both to acceptably low levels.

Mixing Spurs The heterodyne TX of Fig. 4.109 displays various mixing spurs that must be managed properly. The spurs arise from two mechanisms: the harmonics of the first LO and the harmonics of the second LO.

The quadrature upconverter mixes the baseband signals with the third and fifth harmonics of the first LO,³¹ thus creating replicas of the signal spectrum at $\omega_1/2$, $3\omega_1/2$, and $5\omega_1/2$. The result is shown in Fig. 4.111(a) for an asymmetrically-modulated signal. Note that the harmonic magnitudes follow a sinc envelope if the mixers operate as the switching network depicted in Fig. 4.104. In other words, the magnitudes of the replicas at $3\omega_1/2$ and $5\omega_1/2$ are one-third and one-fifth of the desired signal magnitude, respectively. Upon mixing with the second LO (ω_1), the components in Fig. 4.111(a) are translated up and down by an amount equal to ω_1 , yielding the spectrum illustrated in Fig. 4.111(b). Interestingly, the desired sideband at $+3\omega_1/2$ is *enhanced* by a smaller replica that results from the mixing of $5\omega_1/2$ and ω_1 . The unwanted sidebands at $\omega_1/2$, $5\omega_1/2$, and $7\omega_1/2$ must be suppressed by an RF band-pass filter.

The second mechanism relates to the harmonics of the second LO. That is, the spectrum shown in Fig. 4.111(a) is mixed with not only ω_1 but $3\omega_1$, $5\omega_1$, etc. Illustrated in Fig. 4.112 is the resulting output, revealing that, upon mixing with $+3\omega_1$, the IF sideband at $-3\omega_1/2$ is translated to $+3\omega_1/2$, thereby *corrupting* the wanted sideband (if the modulation is asymmetric). Similarly, the IF sideband at $-5\omega_1/2$ is mixed with $+5\omega_1$ and falls atop the desired signal.

31. The higher harmonics are neglected here.

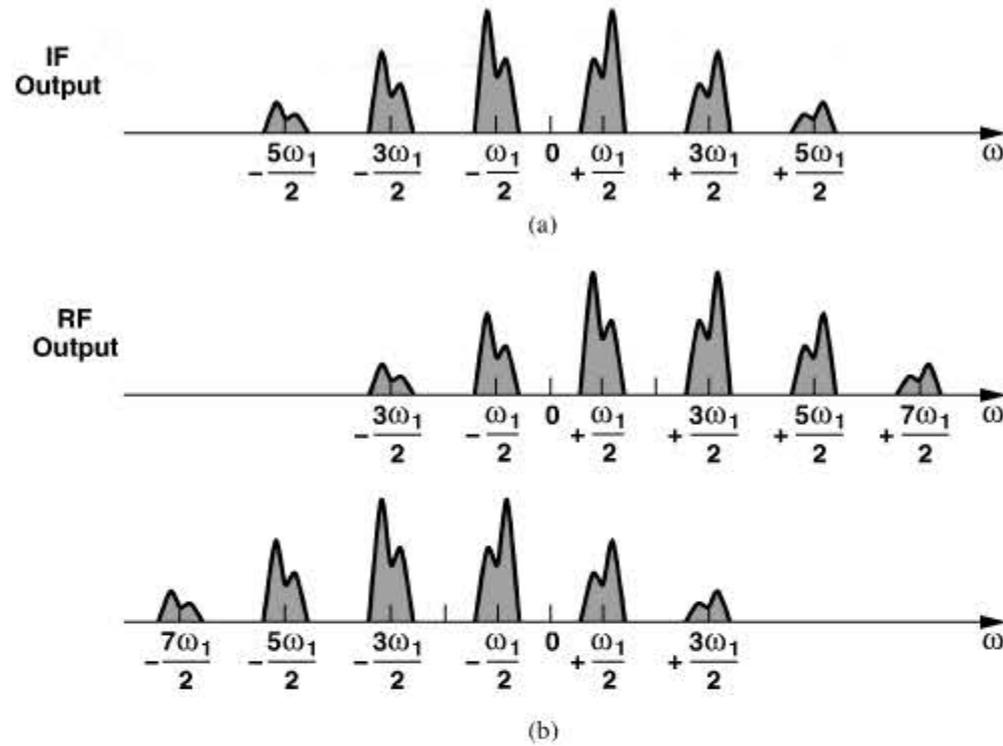


Figure 4.111 Spurs at (a) IF and (b) RF outputs of a heterodyne TX.

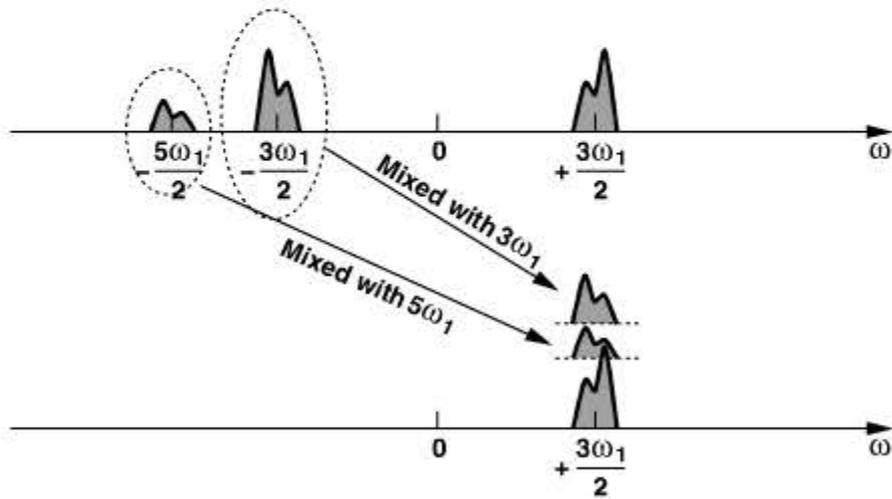


Figure 4.112 Effect of harmonics of second LO on TX output.

How serious is this corruption? Since the IF sideband at $-3\omega_1/2$ is 10 dB below the desired signal, and since mixing with $3\omega_1$ entails another 10 dB attenuation (why?), the level of corruption is at -20 dB. This value is acceptable only for modulation schemes that require a moderate SNR (10–12 dB) (e.g., QPSK) or systems with a moderate bit error rate (e.g., 10^{-2}). Even in these cases, some IF filtering is necessary to suppress the unwanted sidebands before they are upconverted to RF and fall into other users' channels.

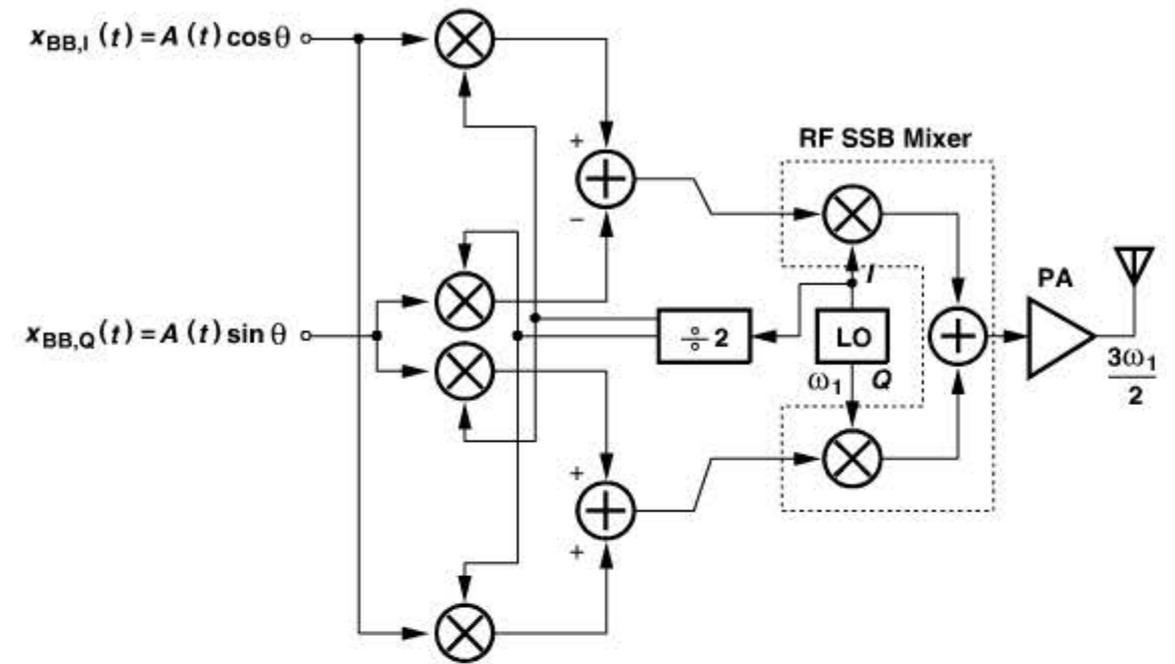


Figure 4.113 Use of baseband quadrature SSB mixing and IF SSB mixing to reduce the unwanted component.

Example 4.43

Compare the spurious behavior of the TX architectures shown in Figs. 4.106 and 4.109.

Solution:

In the direct-conversion TX of Fig. 4.106, the primary spur appears at $5\omega_1/2$, and no self-corruption similar to that illustrated in Fig. 4.112 exists. The heterodyne topology, on the other hand, suffers from more spurs.

The unwanted sideband at $\omega_1 - \omega_1/2$ produced by the RF mixer in Fig. 4.109 can be greatly suppressed through the use of SSB mixing. To this end, the IF signal must be generated in *quadrature* form. Figure 4.113 shows such a topology [15, 18], where two quadrature upconverters provide the quadrature components of the IF signal:

$$x_{IF,I}(t) = A(t) \cos \theta \cos \frac{\omega_1 t}{2} - A(t) \sin \theta \sin \frac{\omega_1 t}{2} \tag{4.152}$$

$$= A(t) \cos \left(\frac{\omega_1 t}{2} + \theta \right) \tag{4.153}$$

$$x_{IF,Q}(t) = A(t) \cos \theta \sin \frac{\omega_1 t}{2} + A(t) \sin \theta \cos \frac{\omega_1 t}{2} \quad (4.154)$$

$$= A(t) \sin \left(\frac{\omega_1 t}{2} + \theta \right). \quad (4.155)$$

The RF SSB mixer then translates the result to $\omega_1 + \omega_1/2$. The reader is encouraged to study the mixing spurs in this architecture.

While attenuating the sideband at $\omega_1 - \omega_1/2$, the architecture of Fig. 4.113 suffers from three drawbacks: (1) the oscillator must provide quadrature outputs, a difficult issue (Chapter 8), (2) the circuit employs twice as many mixers as those in the original architecture (Fig. 4.109), and (3) the loading seen by the $\div 2$ circuit is doubled. The first issue can be alleviated by operating the oscillator at $2\omega_1$ and following it with a $\div 2$ stage, but such a design is only slightly simpler than the direct-conversion architecture of Fig. 4.106.

Our study of the heterodyne sliding-IF TX has thus far assumed that the first LO frequency is half of the second LO frequency. It is possible to replace the $\div 2$ circuit with a $\div 4$ stage so as to produce the IF signal at $\omega_1/4$ and the RF output at $\omega_1 + \omega_1/4 = 5\omega_1/4$ [19]. We study the spurious effects in such an architecture in Problem 4.25.

4.3.5 Other TX Architectures

In addition to the TX architectures described above, several others find usage in some applications. These include “offset PLL” topologies, “in-loop modulation” systems, and “polar modulation” transmitters. We study the first two in Chapter 10 and the last in Chapter 12.

4.4 OOK TRANSCEIVERS

“On-off keying” (OOK) modulation is a special case of ASK where the carrier amplitude is switched between zero and maximum. Transceivers employing OOK lend themselves to a compact, low-power implementation and merit some study here. Figure 4.114 illustrates two TX topologies. In Fig. 4.114(a), the LO is directly turned on and off by the binary baseband data. If the LO swings are large enough, the PA also experiences relatively complete switching and delivers an OOK waveform to the antenna. In contrast to the transmitter architectures studied in the previous sections, OOK does not require quadrature baseband or LO waveforms or a quadrature upconverter. Of course, it is also less bandwidth-efficient as unshaped binary pulses modulated on one phase of the carrier occupy a wide spectrum.

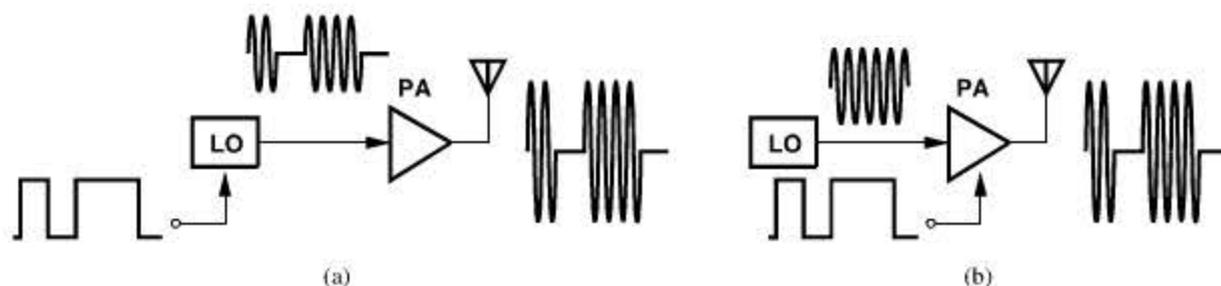


Figure 4.114 OOK TX with (a) direct LO switching (b) PA switching.



Figure 4.115 OOK receiver.

Nonetheless, the simplicity of the architecture makes it attractive for low-cost, low-power applications.

The principal issue in the TX of Fig. 4.114(a) is that the LO cannot be easily controlled by a phase-locked loop (Chapter 9). The TX of Fig. 4.114(b), on the other hand, keeps the LO on and directly switches the PA. We study the injection-pulling properties of the two architectures in Problem 4.29.

OOK receivers are also simple and compact. As shown in Fig. 4.115, an LNA followed by an envelope detector can recover the binary data, with no need for an LO. Of course, such a receiver has little tolerance of interferers.

REFERENCES

- [1] B. Razavi et al., “Multiband UWB Transceivers,” *Proc. CICC*, pp. 141–148, Sept 2005.
- [2] B. Razavi, “Design Considerations for Direct-Conversion Receivers,” *IEEE Trans. Circuits and Systems*, vol. 44, pp. 428–435, June 1997.
- [3] A. A. Abidi, “Direct-conversion Radio Transceivers for Digital Communications,” *IEEE Journal of Solid-State Circuits*, vol. 30, pp. 1399–1410, Dec. 1995.
- [4] R. Hartley, “Modulation System,” US Patent 1,666,206, April 1928.
- [5] D. K. Weaver, “A Third Method of Generation and Detection of Single-Sideband Signals,” *Proc. IRE*, vol. 44, pp. 1703–1705, Dec. 1956.
- [6] J. Rudell et al., “A 1.9-GHz Wideband IF Double Conversion CMOS Receiver for Cordless Telephone Applications,” *IEEE Journal of Solid-State Circuits*, vol. 32, pp. 2071–2088, Dec. 1997.
- [7] L. Der and B. Razavi, “A 2-GHz CMOS Image-Reject Receiver with LMS Calibration,” *IEEE Journal of Solid-State Circuits*, vol. 38, pp. 167–175, Feb. 2003.
- [8] M. Gingell, “Single-Sideband Modulation Using Sequence Asymmetric Polyphase Networks,” *Elec. Comm.*, vol. 48, pp. 21–25, 1973.
- [9] S. Lerstaveesin and B. S. Song, “A Complex Image Rejection Circuit with Sign Detection Only,” *IEEE J. Solid-State Circuits*, vol. 41, pp. 2693–2702, Dec. 2006.
- [10] J. Crols and M. S. J. Steyaert, “A Single-Chip 900-MHz CMOS Receiver Front End with a High-Performance Low-IF Topology,” *IEEE J. Solid-State Circuits*, vol. 30, pp. 1483–1492, Dec. 1995.
- [11] F. Behbahani et al., “CMOS Mixers and Polyphase Filters for Large Image Rejection,” *IEEE J. Solid-State Circuits*, vol. 36, pp. 873–887, June 2001.
- [12] J. Crols and M. S. J. Steyaert, “Low-IF Topologies for High-Performance Analog Front Ends of Fully Integrated Receivers,” *IEEE Tran. Circuits and Sys., II*, vol. 45, pp. 269–282, March 1998.
- [13] K. Feher, *Wireless Digital Communications*, New Jersey: Prentice-Hall, 1995.
- [14] R. Steele, Ed., *Mobile Radio Communications*, New Jersey: IEEE Press, 1992.
- [15] B. Razavi, “A 900-MHz/1.8-GHz CMOS Transmitter for Dual-Band Applications,” *IEEE Journal of Solid-State Circuits*, vol. 34, pp. 573–579, May 1999.

- [16] R. Adler, "A Study of Locking Phenomena in Oscillators," *Proc. of the IEEE*, vol. 61, No. 10, pp. 1380–1385, Oct. 1973.
- [17] B. Razavi, "A Study of Injection Locking and Pulling in Oscillators," *IEEE J. of Solid-State Circuits*, vol. 39, pp. 1415–1424, Sep. 2004.
- [18] M. Zargari et al., "A 5-GHz CMOS Transceiver for IEEE 802.11a Wireless LAN Systems," *IEEE J. of Solid-State Circuits*, vol. 37, pp. 1688–1694, Dec. 2002.
- [19] S. A. Sanielevici et al., "A 900-MHz Transceiver Chipset for Two-Way Paging Applications," *IEEE J. of Solid-State Circuits*, vol. 33, pp. 2160–2168, Dec. 1998.
- [20] M. Conta, private communication, Feb. 2011.
- [21] B. Razavi, "A 5.2-GHz CMOS Receiver with 62-dB Image Rejection," *IEEE Journal of Solid-State Circuits*, vol. 36, pp. 810–815, May 2001.
- [22] A. Parsa and B. Razavi, "A New Transceiver Architecture for the 60-GHz Band," *IEEE Journal of Solid-State Circuits*, vol. 44, pp. 751–762, Mar. 2009.

PROBLEMS

- 4.1. For the sliding-IF architecture of Fig. 4.26(a), assume the $\div 2$ circuit is replaced with a $\div 4$ circuit.
- Determine the required LO frequency range and steps.
 - Determine the image frequency range.
- 4.2. Since the image band of the sliding-IF receiver of Fig. 4.26(a) is narrower than the signal band, is it possible to design an 11g receiver whose image is confined to the GPS band? Explain your reasoning.
- 4.3. A sliding-IF receiver with $f_{LO} = (2/3)f_{in}$ is designed for the 11g band. Determine some of the mixing spurs that result from the harmonics of the first LO and the second LO. Assume the second IF is zero.
- 4.4. Consider the 11g sliding-IF receiver shown in Fig. 4.116.
- Determine the required LO frequency range.
 - Determine the image frequency range.
 - Is this architecture preferable to that in Fig. 4.26(a)? Why?

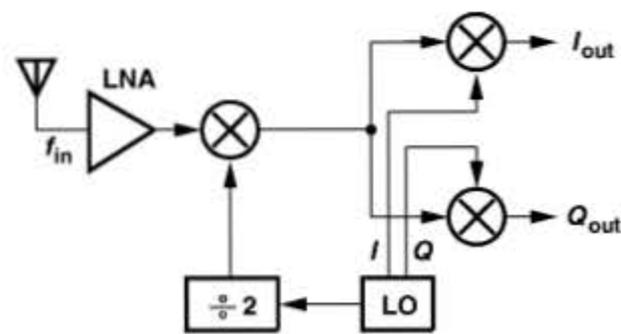


Figure 4.116 Sliding-IF RX for 11g.

- 4.5. Determine some of the mixing spurs in the architecture of Fig. 4.116.
- 4.6. The sliding-IF architecture shown in Fig. 4.117 is designed for the 11a band.
- Determine the image band.
 - Determine the interferer frequencies that can appear in the output baseband as a result of mixing with the third harmonic of the first LO or the third harmonic of the second LO.

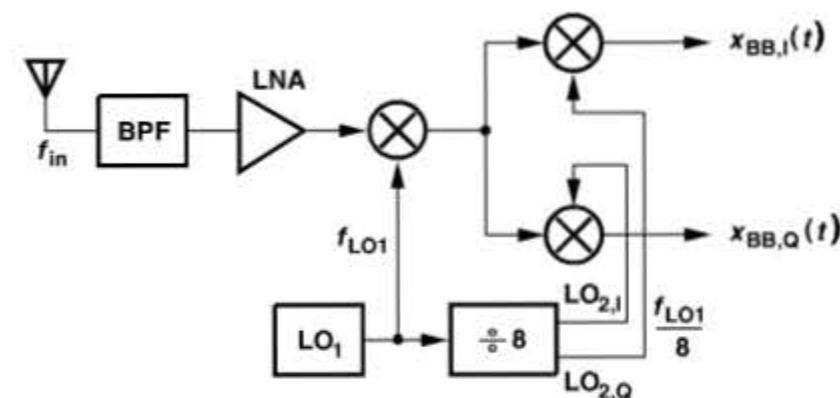


Figure 4.117 Sliding-IF RX for 11a.

- 4.7. Figure 4.118 shows a "half-RF" architecture, where $f_{LO} = f_{in}/2$ [21, 22].
- Assume the RF input is an asymmetrically-modulated signal. Sketch the spectra at the first and second IFs if the mixers are ideal multipliers.
 - Repeat part (a) but assuming that the RF mixer also multiplies the RF signal by the third harmonic of the LO.
 - The flicker noise of the LNA may be critical here. Explain why.

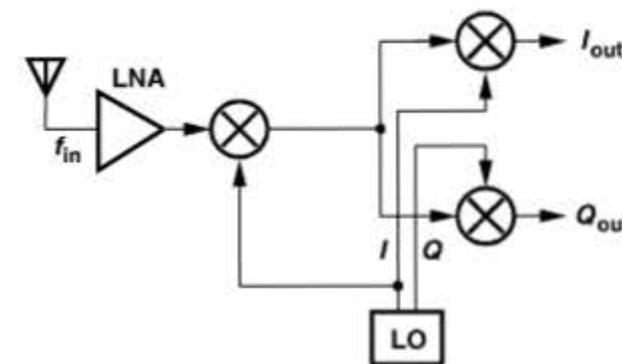


Figure 4.118 Half-RF RX.

- 4.8. Suppose an AM signal, $A(t) \cos \omega_c t$, is applied to a single mixer driven by an LO.
- If the LO waveform is given by $\cos \omega_c t$, determine the baseband signal.
 - If the LO waveform is given by $\sin \omega_c t$, what happens? Why does this indicate the need for quadrature downconversion?

- 4.9. In this problem we wish to study an interesting effect that arises from LO leakage in direct-conversion receivers [20]. Consider the LO leakage in a direct-conversion receiver, $V_0 \cos \omega_{LO} t$. Suppose this leakage is added to an amplitude-modulated interferer, $V_{int}(t) \cos \omega_{int} t$, and the result experiences third-order nonlinearity in the LNA (or downconversion mixer).
- Determine the components near the carrier at the LNA output.
 - Determine the resulting baseband components and whether they corrupt the desired signal.
- 4.10. In Example 4.24, how much gain must precede the given noise spectrum so that the penalty remains below 1 dB?
- 4.11. In Example 4.24, what flicker noise corner frequency is necessary if the penalty must remain below 1 dB?
- 4.12. An ASK waveform is applied to a direct-conversion receiver. Plot the baseband I and Q waveforms.
- 4.13. Does the quadrature mixing of Fig. 4.59(a) perform a Hilbert transform if the *upconverted* outputs at $\omega_c + \omega_{LO}$ are considered?
- 4.14. Repeat the analysis in Fig. 4.59 if $\omega_{IF} > \omega_c$.
- 4.15. Does the Hartley architecture cancel the image if the IF low-pass filters are replaced with high-pass filters and the upconverted components are considered?
- 4.16. In the architecture of Fig. 4.64, assume the two resistors have a mismatch of ΔR . Compute the IRR.
- 4.17. Prove that the IRR of the Hartley architecture is given by $(\omega_{IF}/\Delta\omega)^2$ at an intermediate frequency of $\omega_{IF} + \Delta\omega$ if $\omega_{IF} = (R_1 C_1)^{-1}$.
- 4.18. Considering only the thermal noise of the resistors in Fig. 4.64 and assuming a voltage gain of A_1 for each mixer, determine the noise figure of the receiver with respect to a source impedance of R_D .
- 4.19. In the Weaver architecture of Fig. 4.67, both quadrature downconversions were performed with low-side injection. Study the other three combinations of high-side and low-side injection with the aid of signal spectra at nodes A-F.
- 4.20. Figure 4.119 shows three variants of the Hartley architecture. Explain which one(s) can reject the image.
- 4.21. If $\sin \omega_{LO} t$ and $\cos \omega_{LO} t$ in the Hartley architecture are swapped, does the RX still reject the image?
- 4.22. Repeat the above problem for the first or second LO in a Weaver architecture.
- 4.23. Using Eq. (4.96), compute the IRR of the receiver shown in Fig. 4.77(b) at an IF of $\omega + \Delta\omega$.
- 4.24. Assume a second-order parallel RLC tank is excited by a current source containing a component at ω_0 and another at $3\omega_0$. Prove that, if the tank resonates at $3\omega_0$, then the first harmonic is attenuated by approximately a factor of $8Q/3$ with respect to the third harmonic.

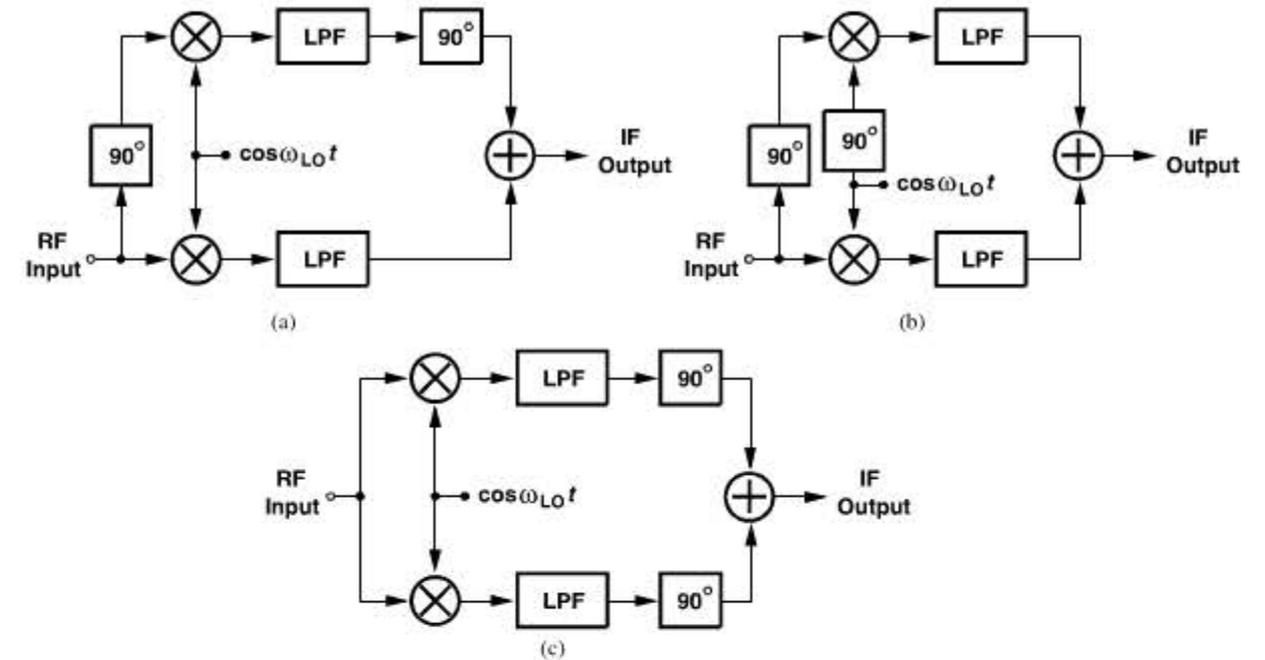


Figure 4.119 Possible variants of Hartley RX.

- 4.25. If the $\div 2$ circuit in Fig. 4.109 is replaced with a $\div 4$ circuit, study the spurious components arising from the third and fifth harmonics of the first and second LO frequencies.
- 4.26. The simplified Hartley architecture shown in Fig. 4.120 incorporates mixers having a voltage conversion gain of A_{mix} and an infinite input impedance. Taking into account only the noise of the two resistors, compute the noise figure of the receiver with respect to a source resistance of R_S at an IF of $1/(R_1 C_1)$.

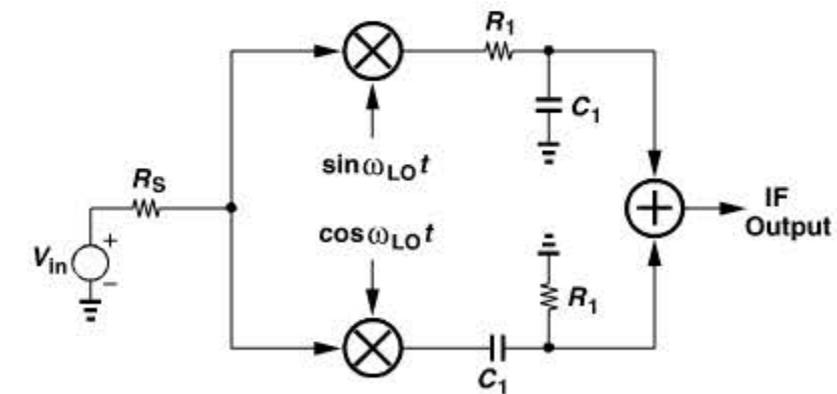


Figure 4.120 Simplified Hartley RX.

- 4.27. A dual-band receiver employing a Weaver architecture is shown in Fig. 4.121. The first LO frequency is chosen so as to create high-side injection for the 2.4-GHz band and low-side injection for the 5.2-GHz band. (The receiver operates only in one band at a given time.) Neglect the noise and nonlinearity of the receiver itself and assume

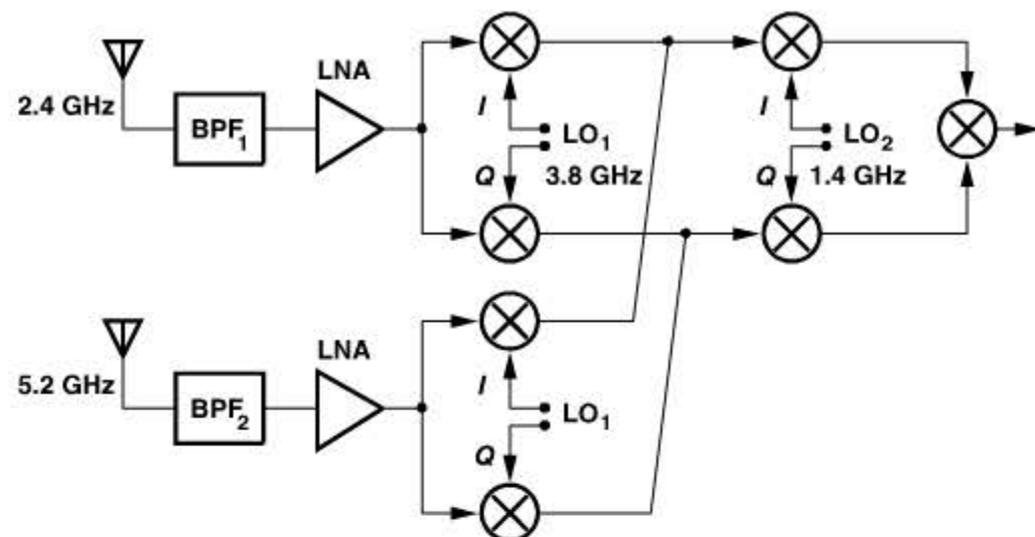


Figure 4.121 Dual-band RX.

an SNR of 20 dB is required for the signal to be detected properly. The Weaver architecture provides an image rejection ratio of 45 dB.

- (a) Suppose the receiver must detect a -85 -dBm signal in the 2.4-GHz mode while receiving at the same antenna a -10 -dBm 5.2-GHz component as well. Determine the amount of rejection required of BPF₁ at 5.2 GHz.
- (b) Suppose the receiver operates in the 5.2-GHz band but it also picks up a strong component at 7.2 GHz. It is possible for this component to be mixed with the third harmonics of LO₁ and LO₂ and appear in the baseband. Does the Weaver architecture prohibit this phenomenon? Explain in detail.
- 4.28. Consider the single-sideband mixer shown in Fig. 4.122. In the ideal case, the output has only one component at $\omega_1 + \omega_2$. Now suppose the ports sensing ω_2 suffer from third- and fifth-order nonlinearity. Plot the output spectrum if (a) $\omega_1 > 3\omega_2$ or (b) $\omega_1 < 3\omega_2$. Identify the frequency of each component.

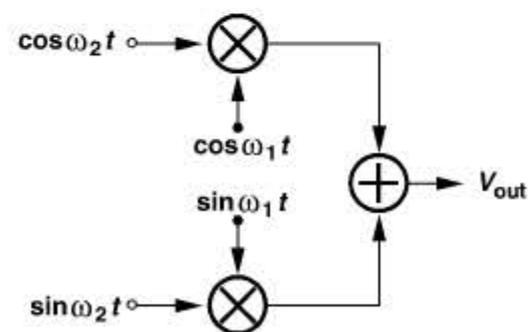


Figure 4.122 SSB mixer.

- 4.29. Explain why injection pulling is more serious in Fig. 4.114(b) than in Fig. 4.114(a).

CHAPTER

5

LOW-NOISE AMPLIFIERS

Following our system- and architecture-level studies in previous chapters, we move farther down to the circuit level in this and subsequent chapters. Beginning with the receive path, we describe the design of low-noise amplifiers. While our focus is on CMOS implementations, most of the concepts can be applied to other technologies as well. The outline of the chapter is shown below.

Basic LNA Topologies	Alternative LNA Topologies	Nonlinearity of LNAs
<ul style="list-style-type: none"> ▪ CS Stage with Inductive Load ▪ CS Stage with Resistive Feedback ▪ CG Stage ▪ CS Stage with Inductive Degeneration 	<ul style="list-style-type: none"> ▪ Variants of CS LNA ▪ Noise-Cancelling LNAs ▪ Differential LNAs 	<ul style="list-style-type: none"> ▪ Nonlinearity Calculations ▪ Differential and Quasi-Differential LNAs

5.1 GENERAL CONSIDERATIONS

As the first active stage of receivers, LNAs play a critical role in the overall performance and their design is governed by the following parameters.

Noise Figure The noise figure of the LNA directly adds to that of the receiver. For a typical RX noise figure of 6 to 8 dB, it is expected that the antenna switch or duplexer contributes about 0.5 to 1.5 dB, the LNA about 2 to 3 dB, and the remainder of the chain about 2.5 to 3.5 dB. While these values provide a good starting point in the receiver design, the exact partitioning of the noise is flexible and depends on the performance of each stage in the chain. In modern RF electronics, we rarely design an LNA in isolation. Rather, we view and design the RF chain as one entity, performing many iterations among the stages.

To gain a better feel for a noise figure of 2 dB, consider the simple example in Fig. 5.1(a), where the noise of the LNA is represented by only a voltage source. Rearranging