

Chapter 2

Receiver Architectures

2.1 Introduction

In Chapter 1 we focused on the blocks enclosed by dotted lines in Figure 1.1. They are the modulator, demodulator, and channel part of the communication system. We now redraw Figure 1.1 as Figure 2.1, with the receiver front end enclosed by dotted lines. In this chapter and the rest of the book, we emphasize this front end part of the communication system. Specifically, in this chapter we discuss the architecture of the front end as well as the filters inside, and in the rest of the book we concentrate on the design of active components inside the front end.

We first review some general philosophy on deciding the front end architecture. We then adopt specific receiver architecture and translate the boundary conditions imposed by the channel and the demodulator on the front end (specified in Chapter 1 in terms of communication concepts [SNR]) to circuit concepts (gain, noise figure, distortion). Next we translate these boundary conditions into boundary conditions of the front end sub blocks. Examples of these sub blocks include low-noise amplifiers (LNA), mixers, and intermediate frequency (IF) amplifiers. In subsequent chapters, these boundary conditions will be translated into the design of various sub component specifications.

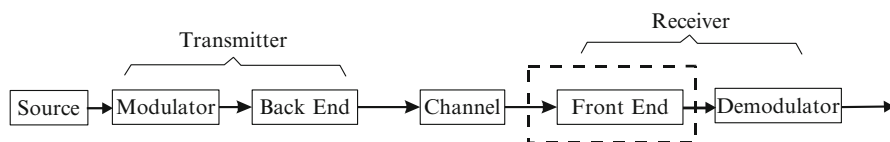


Fig. 2.1 Front end of the receiver enclosed in dotted line

2.2 Receiver Front End: General Discussion

2.2.1 Motivations

The first question we must ask is, why do we need a front end? To answer it, we repeat the boundary conditions imposed by the channel and modulator/demodulator for the front end, derived in Chapter 1 for the DECT example:

1. Due to path loss, the receiver can expect to have a received signal, from the channel, whose minimum carrier's amplitude, P_{\min} , is -77 dBm. Hence the front end's P_{\min} is -77 dBm.
2. Due to envelope's fading and AWGN in the channel, the SNR at the input of the demodulator, denoted as $\text{SNR}_{\text{demod_in}}$, is required to be 25 dB. We assume that no diversity technique is applied and hence the SNR at the output of the front end, denoted as $\text{SNR}_{\text{rec_front_out}}$, is required to be 25 dB.

Even though these boundary conditions have only been determined for DECT, similar situations exist for other standards as well. The difference lies usually in the numbers. For example, in GSM, the P_{\min} number will be much worse due to a larger cell size. To justify the need for a front end, let us go back to the DECT standard. We would want to see, under the worst input condition as described in boundary condition (1), if boundary condition (2) would still be satisfied. In other words, when the antenna has a received power of -77 dBm and there is only AWGN in a fading channel, would the resulting $\text{SNR}_{\text{demod_in}}$ meet the required $\text{SNR}_{\text{demod_in}}$? Let us assume that the AWGN comes just from the thermal noise of the antenna termination resistance of $50\ \Omega$. As shown in Problem 2.1, the $\text{SNR}_{\text{demod_in}}$ is a mere 5 dB. Therefore, we definitely do not meet the required $\text{SNR}_{\text{demod_in}}$ of 25 dB. This conclusion is true in general for other wireless standards, and hence in general a front end is required.

To make the no-front-end option look even worse, let us note that in boundary condition (2) the noise source in the SNR calculation has been assumed to be from AWGN in the channel only. We now state another possible noise source in the channel: interference. Such interference in the channel comes from the following:

1. Adjacent channel interference
2. Co-channel interference
3. Interference from transmit bands of other wireless standards, and from the transmit band of the same channel should frequency division duplexing (FDD) be used

Because these interferences are random in nature, the demodulator cannot differentiate them from AWGN and will process them just like noise during the demodulation operation. With this additional noise source, $\text{SNR}_{\text{demod_in}}$ is reduced even more, further justifying the need for a front end.

2.2.2 General Design Philosophy

The technique to combat a low $\text{SNR}_{\text{demod_in}}$ is by adding a front end block, which processes (conditions) the received signal/AWGN/interference before admitting it to the demodulator. This processing can be done in several ways:

1. Reduce interference/AWGN. This can be done by putting a filter in the front end block, which filters out the interference and AWGN. Filtering for the purpose of reducing interference to an extent that the resulting $\text{SNR}_{\text{demod_in}}$ meets the required $\text{SNR}_{\text{demod_in}}$ is denoted as channel filtering. At what frequency do we want to do channel filtering? Remember, if we want to do channel filtering at RF this involves performing a rather narrowband filtering at RF. Doing narrowband filtering at RF, though conceptually simple, is impractical since high-frequency narrowband filtering is expensive.
2. Amplify the desired signal at RF. This can be done by putting an amplifier inside the front end block. Under this arrangement the antenna will be feeding its output into an amplifier, whose output in turn drives the demodulator. Unfortunately, the desired signal, in the worst case, is so small that a large amplification factor is necessary to boost the $\text{SNR}_{\text{demod_in}}$ to the required level. But remember, when we amplify the desired signal, we are also amplifying the interference and AWGN. In particular, the interference's power, being rather large, when applied to an amplifier with a large gain will have the following undesirable effects:
 - a) It will saturate the amplifier.
 - b) The amplifier, being nonlinear, will generate intermodulation products from these interferencers. Some of these products can have the same frequency as the desired signal and corrupt it.
3. Reapply (2), but this time we want to filter out the interference first. The filtered interference has such a small amplitude that when fed to an amplifier, it will not cause undesirable effects as outlined in (2a) and (2b). This type of filtering is denoted as interference suppression filtering, to be distinguished from channel filtering. Interference suppression filtering requirement is not as stringent as channel filtering because all we need to do is to filter out enough interference so that there are no undesirable effects. Now suppose we do this and repeat (2). Can we amplify the desired signal at RF all the way up so as to achieve the required $\text{SNR}_{\text{demod_in}}$? It turns out that the required amplification for the desired signal at RF is so large that the accompanying interference suppression filtering needs to be very high. Hence it is not practical to do this interference suppression filtering at RF.
4. Having gone through a few alternatives, it seems that the following compromised solution holds the best promise:
 - a) Do partial amplification at RF so that it is practical to implement the required interference suppression filtering at RF.

- b) Frequency translates the partially amplified signal down to a lower frequency, called the intermediate frequency (IF). We can now perform some more amplification since it is more practical to implement the required interference suppression filtering at a lower frequency such as IF. At this lower frequency we can also check and see if the frequency is low enough when channel filtering becomes practical. As a result of the combination of this partial amplification and channel filtering the $\text{SNR}_{\text{demod_in}}$ is improved. This improved $\text{SNR}_{\text{demod_in}}$ is then checked to see if it meets the required $\text{SNR}_{\text{demod_in}}$. If it does not we will repeat the process until the required $\text{SNR}_{\text{demod_in}}$ is met. Then we feed the final amplified and filtered signal to our demodulator. Of course in the process of frequency translation (unless the frequency translation is down to baseband or DC) interference that happens to be present at the image frequency (to be defined later) would also be translated to the desired frequency and has to be filtered out before the mixing. This is called anti-image filtering.

In a sense the process outlined in (4) reverses the attenuation the channel imposes on the desired signal and suppresses the strong interference and noise admitted via the channel, in a couple of steps. There are numerous variations of this approach but they all follow a similar philosophy. To illustrate this we consider one such architecture, called the heterodyne architecture.

2.2.3 *Heterodyne and Other Architectures*

A receiver front end, as we have seen, needs to achieve different objectives: amplification, mixing, filtering, and demodulation. Demodulation may include diversity and equalization. For DECT, given that the speed of the mobiles is typically low, as we assumed in Chapter 1 (Section 1.8), then the resulting fading is slow and does not induce unacceptable BER. Thus diversity (to combat fast fading) is not needed and this simplifies the demodulator.

The resulting heterodyne architecture is shown in Figure 2.2, where we denote all the circuitry enclosed by the dotted line's as the front end. Notice that we have, rather arbitrarily, lumped the antenna as part of the channel. We can observe that the front end includes three bandpass filters (BPF). One design can be such that the bandpass filter right after the antenna (BPF1) is used primarily to select the band of interest of the received signal and is referred to as the band selection filter. This usually provides enough suppression on out-of-band interference that undesirable effects described in (2a) and (2b) of sub-section 2.2.2 are avoided. Note that the band includes the entire spectrum in which the users of a particular standard are allowed to communicate. For the DECT standard, the receive band spans from 1880 to 1900 MHz. Hence BPF1 has a bandwidth of around 20 MHz. As discussed

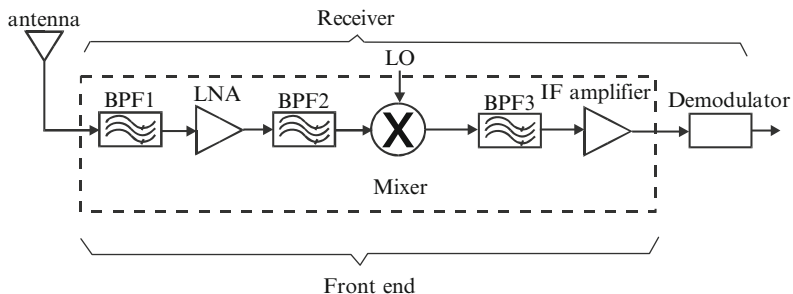


Fig. 2.2 RF receiver using a heterodyne architecture

previously the desired signal, immersed in noise (from both AWGN and interference), can be really small at the Rx antenna (on the order of microvolts) and should be amplified. This amplification should be done with minimal additional noise injected by the amplifier itself, and so a low-noise amplifier (LNA) is used. As mentioned previously, the amplification is followed by mixing, and the mixer should be preceded by an anti-image filter. This is BPF2. This heterodyne architecture adopts one single frequency translation (conversion) and hence is implemented by using a mixer with a variable local oscillator (LO) frequency and a fixed IF. Using this approach, the desired channel can be selected by choosing the proper LO frequency.

Another way to perform channel selection in this architecture consists of using a mixer that mixes down with a fixed LO frequency. The mixer is then followed by a tunable bandpass filter.

In the present approach (variable LO frequency) the mixer is followed by a fixed bandpass filter, BPF3. Since BPF3 operates at IF, we assume the quality factor (Q , not to be confused with Q for complementary error function) required of this filter is now relaxed. Hence we assume that channel filtering becomes possible and therefore BPF3 is also referred to as the channel filter. To perform channel filtering the bandwidth of this filter depends on the channel spacing accorded to each user, which is unique for a particular standard. It is one of the most important constraints on the front end design since this bandwidth is usually very narrow (e.g., 200 kHz in GSM or 1.728 MHz in DECT). The channel filter is followed by an IF amplifier to improve further the $\text{SNR}_{\text{demod_in}}$ of the signal. The demodulator then takes this output and performs demodulation. For a DECT standard, demodulation can be carried out in a manner similar to Figure 1.7 of Chapter 1, which is a form of coherent demodulation. This demodulation can be done in analog or digital domain. In the analog approach, all the correlators, and multipliers will be performed in the analog domain. In the digital approach, analog to digital conversion is performed and all the signal processing (correlations, multiplications) will be performed using a digital signal processor. Also not shown in the diagram is the frequency synthesizer, which generates the LO signal.

Note that there are many other types of possible receiver architectures. For example, in [1] a wideband IF double-conversion scheme has been reported. It uses a similar principle as the heterodyne architecture, but it uses two frequency translation (conversion) blocks, with the first block using a fixed LO frequency and the second block using a variable LO frequency. Further discussion of this approach is carried out in Problem 2.2. Problem 2.9 goes through one design of such architecture. Yet another architecture uses only one conversion stage, where the RF signal is down converted to baseband directly, and is called the homodyne or direct conversion architecture.

Let us assume that we are going to use the heterodyne architecture. We now derive the relevant specifications of the overall receiver front end and then the specifications of the subcomponents. Basically, the interference requirement allows us to determine the bandwidth requirements of the passive subcomponents: filters. The sensitivity requirement will then allow us to determine specifications on the conversion gain, noise, and distortion, first on the front end and then on individual active subcomponents.

In the following section we discuss filter design in some details. The rest of this chapter and subsequent chapters concern the design of the active subcomponents.

2.3 Filter Design

One major technique to combat interference is to filter it out with bandpass filters. For most bandpass filters the relevant design parameters consist of the center frequency, the bandwidth (which together with center frequency defines the quality factor Q), and the out-of-band suppression. We now discuss how to derive the specifications of these parameters. Because these filters are normally not in integrated circuit form, their implementations are mentioned only briefly.

2.3.1 Band Selection Filter (BPF1)

The bandwidth of the band selection filter is typically around the band of interest (for DECT, 17.28 MHz), and the center frequency is the center of the band (for DECT, 1.89 GHz). The Q required is typically high (for DECT, $1.89 \text{ GHz}/17.28 \text{ MHz} \cong 100$) and the center frequency is high as well (as mentioned, for DECT it is 1.89 GHz). On the other hand, the suppression is typically not prohibitive. It only needs to be large enough to ensure that interference is suppressed to a point that it does not cause the undesirable effects as discussed in sub-section 2.2.2. To satisfy these specifications, BPF1 can be implemented using a passive LC filter. This LC filter can be combined with the input matching network of the LNA, which is described in Chapter 3.

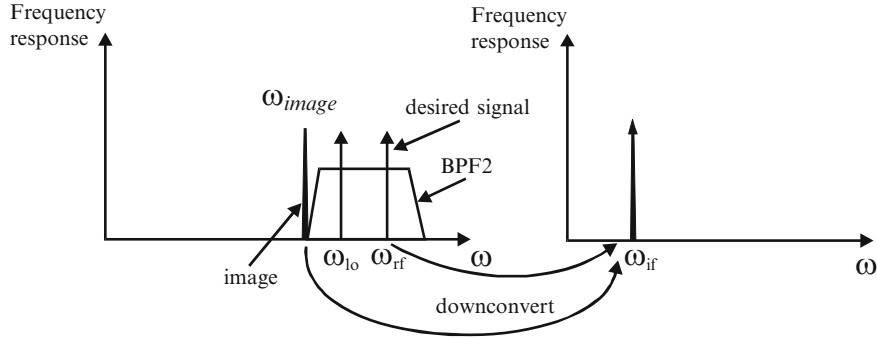


Fig. 2.3 Image rejection problem in the case where FDMA techniques are used to separate different users

2.3.2 Image Rejection Filter (BPF2)

The problem of the image due to mixing was mentioned in sub-section 2.2.2 and is now explained in more detail. As shown in Figure 2.3, during the downconversion process, the desired signal at ω_{rf} , the radio frequency, and its image at ω_{image} , the image frequency, are downconverted to the same frequency of ω_{if} , the intermediate frequency. Hence the desired signal is corrupted. For the downconverted image and downconverted desired signal to overlap, ω_{image} is spaced $2\omega_{if}$ apart from ω_{rf} :

$$\omega_{image} = \omega_{rf} - 2\omega_{if} \quad (2.1)$$

where ω_{if} is from definition, given by

$$\omega_{if} = \omega_{rf} - \omega_{lo} \quad (2.2)$$

Here ω_{lo} is the local oscillator frequency.

To show the image problem mathematically, we model the downconversion (mixing) process to be mathematically equivalent to multiplication (a more complete mathematical treatment of mixing is carried out in Chapter 4). Hence we can write the IF signal as

$$V_{if}(t) = A_{rf} \cos \omega_{rf}t \times A_{lo} \cos \omega_{lo}t \quad (2.3)$$

where $V_{rf} = A_{rf} \cos \omega_{rf}t$ denotes the desired signal whose frequency is at ω_{rf} and $V_{lo} = A_{lo} \cos \omega_{lo}t$ is the local oscillator signal. For simplicity we assume that $A_{lo} = A_{rf} = A$ and hence we have

$$V_{if}(t) = A \cos \omega_{rf}t \times A \cos \omega_{lo}t \quad (2.4)$$

Next we apply trigonometric manipulations and rewrite (2.4) as follows:

$$\begin{aligned} V_{if}(t) &= \frac{1}{2}A^2(\cos(\omega_{rf} + \omega_{lo})t + \cos(\omega_{rf} - \omega_{lo})t) \\ &= \frac{1}{2}A^2(\cos(\omega_{rf} + \omega_{lo})t + \cos \omega_{if}t) \end{aligned} \quad (2.5)$$

The second term is our term of interest. What happens if the received signal consists of the desired signal accompanied by a strong interferer at the image frequency ω_{image} ? Assume that the image is given by $V_{image}=A_{image}\cos(\omega_{image}t)=A\cos(\omega_{image}t)$, where we have again assumed for simplicity that $A_{image}=A$. Then upon downconversion (mixing) the interferer generates the downconverted image:

$$A \cos \omega_{image}t \times A \cos \omega_{lo}t \quad (2.6)$$

Next we will eliminate the term ω_{rf} in (2.1), and (2.2) and express ω_{image} as:

$$\omega_{image} = \omega_{lo} - \omega_{if}. \quad (2.7)$$

Substituting this in (2.6) and applying trigonometric manipulations again, the downconverted image becomes

$$\begin{aligned} &\frac{1}{2}A^2(\cos(\omega_{lo} - \omega_{if} + \omega_{lo})t + \cos(\omega_{lo} - \omega_{if} - \omega_{lo})t) \\ &= \frac{1}{2}A^2(\cos(2\omega_{lo} - \omega_{if})t + \cos \omega_{if}t) = \frac{1}{2}A^2(\cos(2\omega_{lo} - \omega_{if})t) + \frac{1}{2}A^2 \cos \omega_{if}t \end{aligned} \quad (2.8)$$

Comparing (2.8) with (2.5) one can see that both of their second terms are at ω_{if} and so the downconverted image does indeed lie on top of the downconverted desired signal. Thus if there is a strong interference at the image frequency, then at the IF there will be a strong interference sitting on top of the desired signal, resulting in a serious degradation of SNR. Notice that no amount of filtering can help the situation after the downconversion process. Thus, we have to ensure that the image is filtered before downconversion.

When would a strong interference be present at the image frequency? Since the image frequency is only $2\omega_{if}$ from the desired signal, if ω_{if} is small (low IF), then this interference can come from adjacent channels or transmit bands of the same standard (for standards using FDD). If ω_{if} is large, this interference can come from transmit bands of other wireless standards.

To get rid of this image, we should place BPF2 in front of the mixer. The center frequency of BPF2, ω_{center_BPF2} , is typically around the LO frequency. If BPF2 has a bandwidth as shown in Figure 2.3, then the image will be filtered out. How do we determine the bandwidth of this filter?

The answer depends on whether we have a low IF or high IF front end, as the frequencies of the interferers are different. As an example we will treat the low

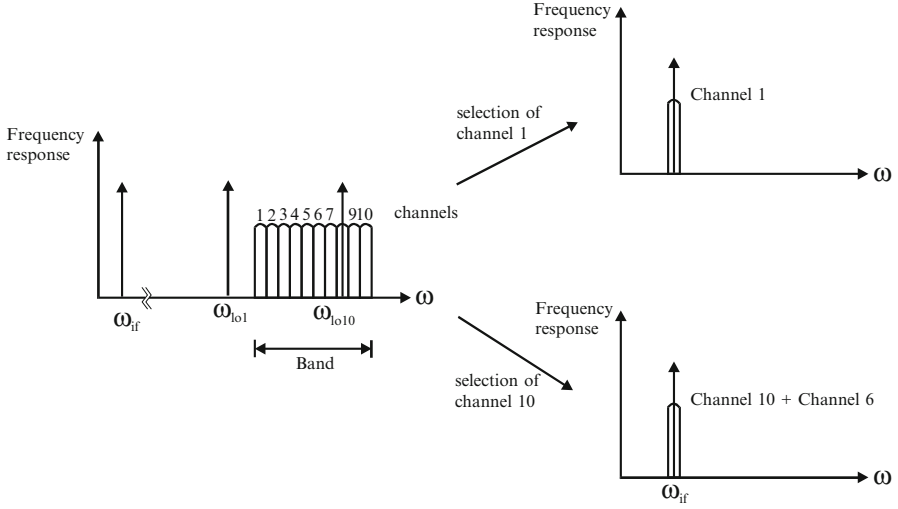


Fig. 2.4 Diagram showing problem of image and function of BPF2

IF case. In this case, the interference comes primarily from an adjacent channel. Let us redraw Figure 2.3 in Figure 2.4, but including adjacent channels. Figure 2.4 is drawn with 10 channels and illustrates the case for the DECT standard. Here different users are separated using FDMA techniques. To maintain the same IF when the desired channel frequency changes, we can change the LO frequency accordingly. We stated that ω_{center_BPF2} is approximately the same as the LO frequency. One way to implement BPF2 is to make ω_{center_BPF2} exactly the same as the LO frequency. However, since the LO frequency changes, this arrangement necessitates that ω_{center_BPF2} be variable, which is impractical because implementing bandpass filters with a variable center frequency at high frequency is very difficult. A better alternative (which is our choice here) is to let ω_{center_BPF2} remain fixed.

Next we determine the bandwidth of a BPF2 that has a fixed center frequency. Referring to Figure 2.4 we show that the 10 channels in DECT are frequency multiplexed over a band that spans from 1880 to 1900 MHz. First let us try to select channel 1. We do this by setting the LO frequency to be ω_{lo1} . This will allow us to select the first channel. Now let us arbitrarily choose

$$\omega_{if} = 2 \times \omega_{channel} \quad (2.9)$$

where $\omega_{channel}$ is the channel frequency.

Substituting (2.9) in (2.7), we have

$$\omega_{image} = \omega_{lo1} - \omega_{if} = \omega_{lo1} - 2 \times \omega_{channel} \quad (2.10)$$

From Figure 2.4 there is no interference at that frequency and we do not have a problem.

Next we try to select some higher number channels. To do this we increase the LO frequency, and the image frequency (at ω_{lo-if}) starts to increase and eventually may move into the band. At that point some other channels of the band will become interferers. This is what we call adjacent channel interference. For example, referring to Figure 2.4, we select the tenth channel by setting the LO frequency to be $\omega_{lo_{10}}$; hence

$$\omega_{rf} = \omega_{10-th-channel} \quad (2.11)$$

Substituting (2.11), and (2.9) into (2.1), the image will now be at a frequency given by

$$\omega_{image} = \omega_{10-th-channel} - 4 \times \omega_{channel} = \omega_{6-th-channel} \quad (2.12)$$

Hence the sixth channel would be downconverted to the same IF frequency as the tenth channel. To get rid of this image we should set BPF2's bandwidth to be less than $4 \times \omega_{channel}$. This means that BPF2 has a center frequency $\omega_{center_BPF2} = \omega_{8-th-channel}$ and span from $\omega_{6-th-channel}$ upward. The lower cutoff frequency is then $\omega_{6-th-channel}$. However, this creates a problem. For example, if the user being assigned this tenth channel is later reassigned to channel 5, then $\omega_{rf} = \omega_{5-th-channel}$. This is below $\omega_{6-th-channel}$ and the user's channel will be cut off. We can avoid this problem by making sure that under the worst case, as the LO frequency varies, the resulting image frequency never falls inside the band.

What is the worst case? Referring to Figure 2.4 the worst case happens when the desired channel is at the highest number channel, or channel 10. Under this worst case we have

$$\omega_{rf} = \omega_{10-th-channel} \quad (2.13)$$

We want to place the desired signal's frequency at the upper cutoff frequency of BPF2 so the upper cutoff frequency of BPF2 is given by:

$$\omega_{upper_cutoff} = \omega_{10-th-channel} \quad (2.14)$$

The lower cutoff frequency ω_{lower_cutoff} of BPF2, ω_{cutoff} , should be larger than the image frequency:

$$\omega_{lower_cutoff} > \omega_{image} \quad (2.15)$$

However, in order not to cut off any channel, this lower cutoff frequency also has to satisfy

$$\omega_{lower_cutoff} < \omega_{1-th-channel} \quad (2.16)$$

As a minimum condition we set

$$\omega_{\text{lower_cutoff}} = \omega_{1\text{-th-channel}} \quad (2.17)$$

Meanwhile, the bandwidth of BPF2, $\omega_{\text{bw_BPF2}}$, is given by definition as

$$\omega_{\text{bw_BPF2}} = \omega_{\text{upper_cutoff}} - \omega_{\text{lower_cutoff}} \quad (2.18)$$

Then, if we substitute (2.14), and (2.17) in (2.18), we will have

$$\omega_{\text{bw_BPF2}} = \omega_{10\text{-th-channel}} - \omega_{1\text{-th-channel}} = \omega_{\text{band}} \quad (2.19)$$

Here ω_{band} is the bandwidth of the band.

For DECT, ω_{band} is around $2\pi \times 20$ Mrad/s and ω_{rf} is around $2\pi \times 1.9$ Grad/s. We assume that ω_{lo} is close to ω_{rf} or close to $2\pi \times 1.9$ Grad/s. Hence $\omega_{\text{center_BPF2}}$ is close to $2\pi \times 1.9$ Grad/s.

Hence we need a filter centered at close to $2\pi \times 1.9$ Grad/s with a $2\pi \times 20$ Mrad/s bandwidth. The amount of suppression must be large enough that in the presence of the strongest adjacent channel interference, the resulting $\text{SNR}_{\text{demod_in}}$ is larger than its required value. With these specifications BPF2 is typically designed using a surface acoustic wave (SAW) filter or ceramic filter. The output matching network of the LNA can also be used to provide some of this filtering, as shown in Chapter 3.

2.3.3 Channel Filter (BPF3)

First we want to find the center frequency of BPF3, which is the same as the IF. To find IF we do the following: from (2.15) $\omega_{\text{image}} < \omega_{\text{lower_cutoff}}$. As a minimum condition we have

$$\omega_{\text{image}} = \omega_{\text{lower_cutoff}} \quad (2.20)$$

(the filter barely suppresses the image). Now we apply the worst case scenario as described in equations (2.13) through (2.19). Substituting (2.17) in (2.20), we have

$$\omega_{\text{image}} = \omega_{1\text{-st-channel}} \quad (2.21)$$

Substituting (2.13), and (2.21) into (2.1), we finally have

$$\omega_{10\text{-th-channel}} - \omega_{1\text{-st-channel}} = 2\omega_{\text{if}} \text{ or } \omega_{\text{if}} = (\omega_{10\text{-th-channel}} - \omega_{1\text{-st-channel}})/2 = \omega_{\text{band}}/2 \quad (2.22)$$

As an example, for DECT, using (2.22) the IF should be larger than $2\pi \times 10$ Mrad/s. However, this would put a stringent requirement on the rolloff of BPF2. Hence the IF is usually increased substantially beyond this bound. In the present case we set IF to be $2\pi \times 100$ Mrad/s:

$$\omega_{\text{if}} = 2\pi \times 100 \text{ Mrad/s} \quad (2.23)$$

Therefore, the center frequency of BPF3, $\omega_{\text{center_BPF3}}$, is

$$\omega_{\text{center_BPF3}} = 2\pi \times 100 \text{ Mrad/s} \quad (2.24)$$

Next we want to find the BW of BPF3. Since at this point we are at a low enough frequency to do channel filtering, without fearing an unrealizable Q , we will attempt to do so. Hence the bandwidth of BPF3 is simply the channel bandwidth, or

$$\omega_{\text{bw_BPF3}} = \omega_{\text{channel}} = 2\pi \times 1.728 \text{ Mrad/s} \quad (2.25)$$

For the Q required [for DECT standard and our choice of intermediate frequency (IF), we have $Q = (2\pi \times 100 \text{ mMrad/s}) / (2\pi \times 1.728 \text{ Mrad/s}) \cong 60$] and the center frequency specified ($2\pi \times 100$ Mrad/s), this filter may be designed with an active filter (continuous time filter) or ceramic filter. The output network of the mixer can also provide some channel filtering.

2.3.3.1 Trade-Off Between BPF2, and BPF3

We noted that the choice of IF frequency affects both BPF2 and BPF3. We discuss this in a bit more detail here. First we eliminate the ω_{band} term between (2.19) and (2.22) and show that

$$\omega_{\text{bw_BPF2}} = 2\omega_{\text{if}} \quad (2.26)$$

Hence if ω_{if} increases then $\omega_{\text{bw_BPF2}}$ also becomes larger. Consequently, Q_{BPF2} , which is given roughly by $\omega_{\text{lo}} / \omega_{\text{bw_BPF2}}$, becomes smaller and the design of BPF2 becomes more relaxed. Meanwhile, let us look at Q_{BPF3} , which is given roughly by $\omega_{\text{if}} / \omega_{\text{bw_BPF3}}$. Substituting (2.25) in this equation, we have $Q_{\text{BPF3}} \cong \omega_{\text{if}} / \omega_{\text{channel}}$. Hence a larger ω_{if} means that Q_{BPF3} becomes larger, making BPF3 more difficult to design.

2.4 Rest of Receiver Front End: Nonidealities and Design Parameters

Now that we have talked about the design of filters in the receiver front, we turn our attention to the design of the rest of the components. Normally these components consist of circuits such as LNA, mixer, IF amplifier, and analog/digital (A/D)

converter. Unlike filters, their relevant design parameters are different. Hence our first task is to discuss these design parameters.

We start by noting that these design parameters are intimately related to the nonidealities in these circuits. First, these circuits are nonlinear. The primary effect of this nonlinearity is that it can frequency translate interference with frequencies outside the channel frequency onto the channel frequency, thereby corrupting the desired signal. Since this interference behaves like noise, this will degrade the SNR at the output of the front end. Second, these circuits introduce noise of their own. Again, this will degrade the SNR at the front end's output. In the following subsection we translate the effects of these two nonidealities into the corresponding design parameters.

2.4.1 Nonlinearity

In this subsection we limit our analysis to nonlinearities up to the third order, because normally these are the nonlinearities of most interest in a radio environment. For simplicity we further assume that these nonlinearities are memoryless. Mathematically, memoryless third-order nonlinearities are specified by a third-order polynomial, which characterizes the input-output relation of a memoryless nonlinear system:

$$y(t) = \alpha_1 s(t) + \alpha_2 s^2(t) + \alpha_3 s^3(t) \quad (2.27)$$

Here $s(t)$ is the input signal and $y(t)$ is the output signal and the nonlinearity is memoryless. Because of this nonlinearity, distortion is generated.

2.4.1.1 Harmonic Distortion

Using the input-output relation (2.27) with a single tone at the input $s(t) = A \cos \omega_0 t$, the output of the nonlinear systems can be viewed mathematically as

$$\begin{aligned} y(t) &= \alpha_1 A \cos \omega_0 t + \alpha_2 A^2 \cos^2 \omega_0 t + \alpha_3 A^3 \cos^3 \omega_0 t \\ &= \frac{\alpha_2 A^2}{2} + \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) \cos \omega_0 t + \frac{\alpha_2 A^2}{2} \cos 2\omega_0 t + \frac{\alpha_3 A^3}{4} \cos 3\omega_0 t \end{aligned} \quad (2.28)$$

Harmonic distortion is defined as the ratio of the amplitude of a particular harmonic to the amplitude of the fundamental. For example, third-order harmonic distortion (HD_3) is defined as the ratio of amplitude of the tone at $3\omega_0$ to the amplitude of the fundamental at ω_0 . Applying this definition to (2.28) and assuming $\alpha_1 A \gg (3\alpha_3 A^3)/4$, we have

$$HD_3 = \frac{1}{4} \frac{\alpha_3}{\alpha_1} A^2 \quad (2.29)$$

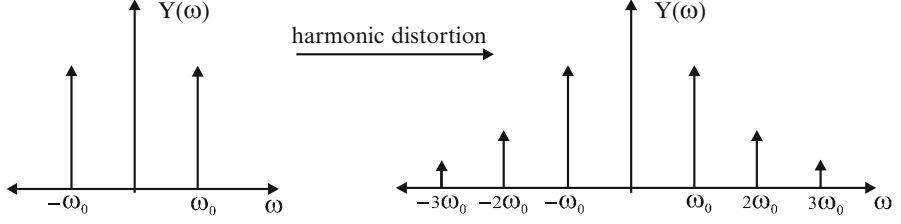


Fig. 2.5 Frequency spectrum of input, output of the nonlinear system

Next we take the Fourier transform of (2.28).

$$\begin{aligned}
 Y(\omega) &= \alpha_2 A^2 \pi \delta(\omega) + \pi \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) [\delta(\omega - \omega_o) + \delta(\omega + \omega_o)] \\
 &+ \pi \frac{\alpha_2 A^2}{2} [\delta(\omega - 2\omega_o) + \delta(\omega + 2\omega_o)] \\
 &+ \pi \frac{\alpha_3 A^3}{4} [\delta(\omega - 3\omega_o) + \delta(\omega + 3\omega_o)]
 \end{aligned} \quad (2.30)$$

Equation (2.30) is plotted in Figure 2.5.

Harmonic distortion is typically not of major concern. As an example, for DECT, $\omega_o = 2\pi \times 1.9$ Grad/s. Suppose that the LNA in Figure 2.2 is nonlinear and generates a second harmonic distortion. However, this is at $2\omega_o$ or $2\pi \times 3.8$ Grad/s and will be filtered by BPF2, hence posing no harm.

2.4.1.2 Intermodulation

Intermodulation arises when more than one tone is present at the input. A common method for analyzing this distortion is the “two-tone” test. We assume that two strong interferers occur at the input of the receiver, specified by $s(t) = A_1 \cos \omega_1 t + A_2 \cos \omega_2 t$. Again, the intermodulation distortion can be expressed mathematically by applying $s(t)$ to (2.27).

$$\begin{aligned}
 y(t) &= \alpha_1 (A_1 \cos \omega_1 t + A_2 \cos \omega_2 t) + \alpha_2 (A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)^2 \\
 &+ \alpha_3 (A_1 \cos \omega_1 t + A_2 \cos \omega_2 t)^3
 \end{aligned} \quad (2.31)$$

Using trigonometric manipulations, we can find expressions for the second and the third-order intermodulation products as follows:

$$\begin{aligned}
 \omega_1 \pm \omega_2 &: \alpha_2 A_1 A_2 \cos(\omega_1 + \omega_2)t + \alpha_2 A_1 A_2 \cos(\omega_1 - \omega_2)t \\
 2\omega_1 \pm \omega_2 &: \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 + \omega_2)t + \frac{3\alpha_3 A_1^2 A_2}{4} \cos(2\omega_1 - \omega_2)t \\
 2\omega_2 \pm \omega_1 &: \frac{3\alpha_3 A_2^2 A_1}{4} \cos(2\omega_2 + \omega_1)t + \frac{3\alpha_3 A_2^2 A_1}{4} \cos(2\omega_2 - \omega_1)t
 \end{aligned} \quad (2.32)$$

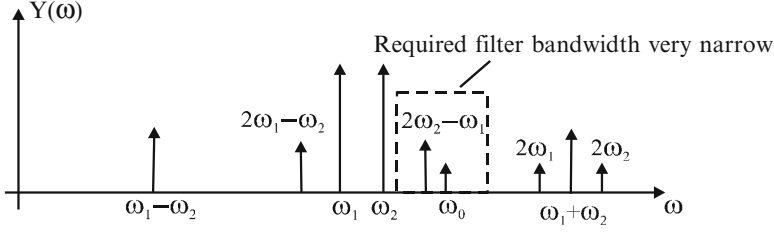


Fig. 2.6 The effect of the intermodulation distortion in the frequency domain, where the frequencies represent the following signals: ω_0 : desired signal; ω_1, ω_2 : strong interferers; $2\omega_1, 2\omega_2$: harmonics of the interferers; $\omega_1 \pm \omega_2$: second order intermodulation products; $2\omega_{1,2} \pm \omega_{2,1}$: third order intermodulation products

The output spectrum in the frequency domain can be determined from (2.32) by evaluating its Fourier transform $Y(\omega)$. This is shown in Figure 2.6.

It can be seen from Figure 2.6 that the intermodulation product with frequency $2\omega_2 - \omega_1$ (denoted as the third order intermodulation product, I_{D3}) lies at ω_0 and corrupts the desired signal at ω_0 . Furthermore ω_1, ω_2 are close to ω_0 and so trying to filter them out requires a filter bandwidth that is very narrow and is impractical. Hence keeping down $2\omega_2 - \omega_1$ by keeping the nonlinearity (which generates them in the first place) is the only solution.

Where do the two tones, at ω_1 and ω_2 , come from? They can be any one of the interferences described in sub-section 2.2.1. Strictly speaking, the interferers are not tones but are more like narrowband noise. For simplicity, for the time being we represent the interfeferer at the band from $\omega_1 - \omega_{\text{channel}}/2$ to $\omega_1 + \omega_{\text{channel}}/2$ as a single tone centered at ω_1 , with the rms value of the narrowband noise set equal to the amplitude of the tone, or A_1 . Similar representation is applied to the desired signal at the band from $\omega_0 - \omega_{\text{channel}}/2$ to $\omega_0 + \omega_{\text{channel}}/2$ and the interferer at the band from $\omega_2 - \omega_{\text{channel}}/2$ to $\omega_2 + \omega_{\text{channel}}/2$. To quantify this distortion we first define the third-order intermodulation distortion, IM_3 , as the ratio of the amplitude of the I_{D3} to the amplitude of the fundamental output component (denoted as I_{D1}) of a linear system given by $y(t) = \alpha_1 A \cos \omega_0 t$, where α_1 is the linear small signal gain. Mathematically, this is written as

$$IM_3 = I_{D3}/I_{D1} \quad (2.33)$$

Note that IM_3 expressed in decibels is simply the difference between the interferer's fundamental output's signal strength in decibels (at ω_0) and the interferer's intermodulated product's strength in decibels (at $2\omega_{2,1} - \omega_{1,2}$).

In order to quantify IM_3 , let us simplify by assuming $A = A_1 = A_2$. Applying (2.31) and (2.32) to (2.33) we get

$$IM_3 = \frac{\frac{3}{4} \alpha_3 A^3}{\alpha_1 A} = \frac{3}{4} \frac{\alpha_3}{\alpha_1} A^2 \quad (2.34)$$

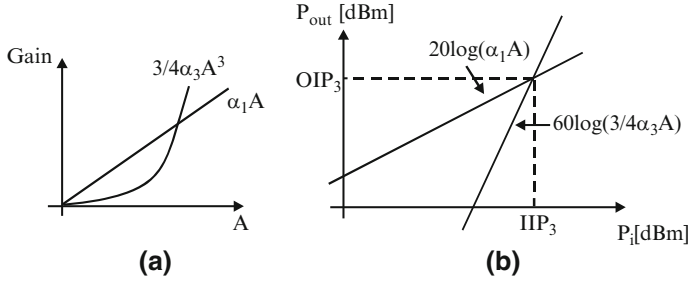


Fig. 2.7 **a)** The linear gain ($\alpha_1 A$) and the nonlinear component ($3/4 \alpha_3 A^3$). **b)** Graphical representation of the input and output third-order intercept point (IIP_3 , OIP_3)

Comparing (2.29) to (2.34), it is seen that

$$IM_3 = 3HD_3. \quad (2.35)$$

Since IM_3 depends on input level and is sometimes not as easy to use, we define another performance metric, called the third order intercept point (IP_3).

2.4.1.3 Third-Order Intercept Point, IP_3

From (2.31), we note that as the input level A increases, the desired signal at the output is proportional to A (by the small signal gain α_1). On the other hand, from (2.32) we can see that the third-order product I_{D3} increases in proportion to A^3 . This is plotted on a linear scale in Figure 2.7a. Figure 2.7a is replotted on a logarithmic scale in Figure 2.7b, where power level is used instead of amplitude level. As shown in Figure 2.7b the power of I_{D3} grows at three times the rate at which the desired signal I_{D1} increases. The third-order intercept point IP_3 is defined to be the intersection of the two lines.

From Figure 2.7b we can see that the amplitude (in voltage) of the input interferer at the third-order intercept point, A_{IP3} , is defined by the relation

$$20 \log(\alpha_1 A_{IP3}) = 20 \log\left(\frac{3}{4} \alpha_3 A_{IP3}^3\right) \quad (2.36)$$

From (2.36) we can solve for A_{IP3} :

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{\alpha_1}{\alpha_3} \right|} \quad (2.37)$$

For a $50 \, \Omega$ load, we define the input third-order intercept point (IIP_3) as $IIP_3 = A_{IP3}^2 / 50 \, \Omega$. (IIP_3 is hence interpreted as the power level of the input interferer for a $50 \, \Omega$ load at the third-order intercept point). Notice that IIP_3 can be interpreted

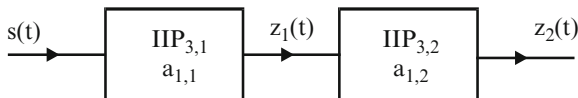


Fig. 2.8 Two cascaded nonlinear stages, where $a_{1,1}$ and $a_{1,2}$ are the linear gains of stage 1 and 2 and $IIP_{3,1}$ and $IIP_{3,2}$ denote the third-order intercept points of stages 1 and 2, $z_1(t)$ denotes the output signal of the first stage, and $z_2(t)$ denotes the output signal of the second stage

in terms of absolute value or decibels. One useful equation that relates IIP_3 to IM_3 , expressed in decibels, is the following [2]:

$$IIP_3|_{dBm} = P_i|_{dBm} - \frac{IM_3|_{dB}}{2} \quad (2.38)$$

Here P_i is the power level of the input interferer and is typically defined for a 50Ω load. Both IIP_3 and P_i have been expressed in dBm, whereas IM_3 is expressed in dB.

2.4.1.4 Cascaded Nonlinear Stages

Next we investigate the impact of the nonlinearity of each stage on the overall nonlinearity performance of the front end. To get a useful design relation, we want to find the overall third-order intercept point at the input in terms of the IIP_3 and the gain of each stage. As a start, let us assume that we have a cascade of two nonlinear stages, as shown in Figure 2.8.

Using the input-output relation (2.27), we can approximate the nonlinear behavior of each subcomponent as follows:

$$z_1(t) = a_{1,1}s(t) + a_{2,1}s^2(t) + a_{3,1}s^3(t) \quad (2.39)$$

$$z_2(t) = a_{1,2}z_1(t) + a_{2,2}z_1^2(t) + a_{3,2}z_1^3(t) \quad (2.40)$$

Here $a_{i,j}$ means the i th order gain of the j th stage. For example, $a_{3,2}$ is the a_3 gain of the second stage.

Now we can write the output of the second stage $z_2(t)$ in terms of the input signal $s(t)$ by applying (2.39) to (2.40):

$$\begin{aligned} z_2(t) = & a_{1,2}(a_{1,1}s(t) + a_{2,1}s^2(t) + a_{3,1}s^3(t)) \\ & + a_{2,2}(a_{1,1}s(t) + a_{2,1}s^2(t) + a_{3,1}s^3(t))^2 \\ & + a_{3,2}(a_{1,1}s(t) + a_{2,1}s^2(t) + a_{3,1}s^3(t))^3 \end{aligned} \quad (2.41)$$

To determine the third-order intercept point, we need to calculate the linear and the third-order terms at the output of the second stage. From (2.41) it can be shown

that the linear term of $z_2(t)$ equals $a_{1,1}a_{1,2}s(t)$ and that the third-order term equals $[a_{3,1}a_{1,2} + 2a_{1,1}a_{2,1}a_{2,2} + a_{1,1}^3a_{3,2}]s^3(t)$.

Since (2.41) describes the overall input-output relation of the two-stage system, we can now treat it as one single stage. We can then reuse the previous formulae derived for a single stage to determine the A_{IP3} of this two-stage system. This consists of setting $\alpha_1 = a_{1,1}a_{1,2}$ and $\alpha_3 = a_{3,1}a_{1,2} + 2a_{1,1}a_{2,1}a_{2,2} + a_{1,1}^3a_{3,2}$ in (2.37). Thus the overall A_{IP3} of the cascaded system can be expressed as

$$A_{IP3} = \sqrt{\frac{4}{3} \left| \frac{a_{1,1}a_{1,2}}{a_{3,1}a_{1,2} + 2a_{1,1}a_{2,1}a_{2,2} + a_{1,1}^3a_{3,2}} \right|} \quad (2.42)$$

Unfortunately, the sign of the coefficient in the denominator in (2.41) is circuit dependent. Considering the worst case, we add the absolute values together. Rearranging, (2.41) becomes

$$\begin{aligned} \frac{1}{A_{IP3}^2} &= \frac{3}{4} \frac{|a_{3,1}a_{1,2}| + |2a_{1,1}a_{2,1}a_{2,2}| + |a_{1,1}^3a_{3,2}|}{|a_{1,1}a_{1,2}|} \\ &= \frac{3}{4} \left| \frac{a_{3,1}}{a_{1,1}} \right| + \frac{3}{2} \left| \frac{a_{2,1}a_{2,2}}{a_{1,2}} \right| + \frac{3}{4} a_{1,1}^2 \left| \frac{a_{3,2}}{a_{1,2}} \right| \end{aligned} \quad (2.43)$$

Remember, our goal is to express the overall A_{IP3} in terms of the A_{IP3} and the linear gain of each stage. Comparing the first term of (2.43) and (2.37), it can be seen that first term gives the reciprocal of the square of $A_{IP3,1}$, the A_{IP3} of the first stage. Following the same reasoning, the third term is seen to consist of the product of the reciprocal of the square of $A_{IP3,2}$, the A_{IP3} of the second stage and the square of $a_{1,1}$, the linear gain of the first stage. Hence (2.43) can be rewritten as

$$\frac{1}{A_{IP3}^2} = \frac{1}{A_{IP3,1}^2} + \frac{|3a_{2,1}a_{2,2}|}{|2a_{1,2}|} + \frac{a_{1,1}^2}{A_{IP3,2}^2} \quad (2.44)$$

We can make further approximations by examining the second term in (2.44). The coefficient $a_{2,1}$ in the second term only contributes to the second-order intermodulation product and harmonic distortion. As shown in Figure 2.6, since the front end operates primarily in a narrow frequency band, the second-order intermodulation products and the harmonic distortion terms lie outside the band of interest and are thus strongly attenuated by the BPFs in the front end. Hence the impact of $a_{2,1}$ in generating the overall front end's distortion component is small. Consequently, in determining the overall front end's A_{IP3} , the second term can be neglected. Accordingly, (2.44) can be further simplified as

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{a_{1,1}^2}{A_{IP3,2}^2} \quad (2.45)$$

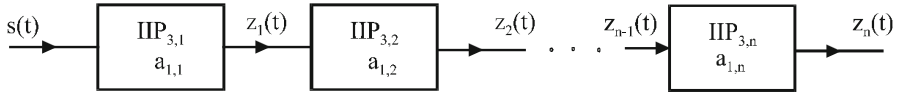


Fig. 2.9 N cascaded nonlinear stages, where $a_{1,1} \dots a_{1,n}$ are the linear gains of stage 1...n and $IIP_{3,1} \dots IIP_{3,n}$ denote the third order intercept points of stage 1...n, $z_1(t)$ denotes the output signal of the first stage, and $z_n(t)$ denotes the output signal of the nth stage

Equation (2.45) leads to a useful observation: since $a_{1,1}$ is large, (2.45) is essentially saying that the overall A_{IP3} is dominated by the A_{IP3} of the second stage. Physically, this is what happens: in the case where the linear gain $a_{1,1}$ of the first stage is large, the input signal to the second stage is large. Thus, the distortion of the second stage becomes more critical since it has to handle a larger input signal.

Next (2.45) can be extended to a cascade of an arbitrary number of nonlinear stages, as shown in Figure 2.9, and the resulting overall input intercept point can be derived to be

$$\frac{1}{A_{IP3}^2} \approx \frac{1}{A_{IP3,1}^2} + \frac{a_{1,1}^2}{A_{IP3,2}^2} + \frac{a_{1,1}^2 a_{1,2}^2}{A_{IP3,3}^2} + \dots + \frac{a_{1,1}^2 a_{1,2}^2 \dots a_{1,n-1}^2}{A_{IP3,n}^2} \quad (2.46)$$

Equation (2.46) can also be restated in a form that involves IIP_3 . For a 50Ω load we stated that IIP_3 is equal $A_{IP3}^2/50$ and hence $IIP_{3,i}$ equals $A_{IP3,i}^2/50$, where i is the stage number. Substituting these relations in (2.46) we have

$$\frac{1}{IIP_3} \approx \frac{1}{IIP_{3,1}} + \frac{a_{1,1}^2}{IIP_{3,2}} + \frac{a_{1,1}^2 a_{1,2}^2}{IIP_{3,3}} + \dots + \frac{a_{1,1}^2 a_{1,2}^2 \dots a_{1,n-1}^2}{IIP_{3,n}} \quad (2.47)$$

Under the assumption that the subcomponents are matched (in order to allow maximum power transfer), G_i , the power gain of each stage, can be expressed in terms of the voltage gain of each stage as

$$G_1 = \text{Power gain of the first stage} = a_{1,1}^2$$

$$G_2 = \text{Power gain of the second stage} = a_{1,2}^2$$

where i is the stage number. (As a side note, the exact relation between the power gain and the voltage gain will be clarified in Problem 2.5). Substituting this in (2.47), we get

$$\frac{1}{IIP_3} \approx \frac{1}{IIP_{3,1}} + \frac{G_1}{IIP_{3,2}} + \frac{G_1 G_2}{IIP_{3,3}} + \dots + \frac{G_1 G_2 \dots G_{n-1}}{IIP_{3,n}} \quad (2.48)$$

Equation (2.48) is a useful approximation of the overall performance of the system due to the nonlinear behavior of each stage. Essentially, it says that the overall input intercept point is dominated by distortion of the last stage.

Finally, the overall output intercept point, OIP_3 , can be expressed in terms of the output intercept point ($OIP_{3,i}$) and power gain (G_i) of each stage ($i = 2, 3, \dots, n$), where n = number of stages. First we start with (2.47) and divide both sides of (2.47) by $a_{1,1}^2 \cdot a_{1,2}^2 \cdot \dots \cdot a_{1,n}^2$:

$$\frac{1}{a_{1,1}^2 \cdot \dots \cdot a_{1,n}^2 IIP_3} \approx \frac{1}{a_{1,1}^2 \cdot \dots \cdot a_{1,n}^2 IIP_{3,1}} + \frac{1}{a_{1,2}^2 \cdot \dots \cdot a_{1,n}^2 IIP_{3,2}} + \dots + \frac{1}{a_{1,n}^2 IIP_{3,n}} \quad (2.49)$$

Next we relate the output intercept point to the input intercept point. For the output intercept point of each stage, $OIP_{3,i}$, we have

$$OIP_{3,i} = a_{1,i}^2 IIP_{3,i}^2 \quad (2.50)$$

where $i = 1, n$. For the output intercept point of the overall front end, OIP_3 , we have

$$OIP_3 = \text{total-gain} \times IIP_3 = a_{1,1}^2 \cdot \dots \cdot a_{1,n}^2 IIP_3 \quad (2.51)$$

Substituting (2.50) and (2.51) into (2.49), we get

$$\frac{1}{OIP_3} = \frac{1}{a_{1,2}^2 \cdot \dots \cdot a_{1,n-1}^2 OIP_{3,1}} + \frac{1}{a_{1,3}^2 \cdot \dots \cdot a_{1,n-1}^2 OIP_{3,2}} + \dots + \frac{1}{OIP_{3,n}} \quad (2.52)$$

Again, under the assumption that the subcomponents are matched (in order to allow maximum power transfer), (2.52) can be expressed in terms of the power gains of each stage:

$$\frac{1}{OIP_3} \approx \frac{1}{G_2 \cdot \dots \cdot G_n OIP_{3,1}} + \frac{1}{G_3 \cdot \dots \cdot G_n OIP_{3,2}} + \dots + \frac{1}{OIP_{3,n}} \quad (2.53)$$

The importance of (2.53) is that it tells us that the output intercept point of the overall front end is dominated by the output intercept point of the last stage.

2.4.1.5 Gain Compression

Another phenomenon caused by the nonlinearity of the receiver is called gain compression. When the input to an amplifier is large, the amplifier saturates, hence clipping the signal. When the strength of the input is further increased, the output signal is no longer amplified. At this point, the output is said to be

compressed. We may now ask what the clipping of a signal has to do with the nonlinear behavior of a system. If we go back to (2.28), we observe that in $y(t)$ there are two terms with frequency ω_0 due to the nonlinear behavior. Let us assume that the other terms in $y(t)$ have frequency outside the band of interest and hence are removed by the BPFs. Thus, $y(t)$ becomes

$$y(t) = \left(\alpha_1 A + \frac{3\alpha_3 A^3}{4} \right) \cos \omega_0 t = \left(\alpha_1 + \frac{3\alpha_3 A^2}{4} \right) A \cos \omega_0 t \quad (2.54)$$

In the case where α_3 is negative, the second term is decreasing the gain. As the input starts to increase, the impact of the second term becomes important in the sense that it saturates the active device. To get a feeling for the input level when considerable gain compression occurs, we can use the concept of the 1-dB compression point, defined as the input level that causes the linear small-signal gain to drop by 1 dB. Thus, the $A_{1\text{-dB}}$ specifies the amplitude (in voltage) of the input signal when the linear voltage gain drops by 1 dB. From (2.54) we see that the 1-dB compression point can be expressed mathematically by

$$\left(\alpha_1 + \frac{3\alpha_3 A_{1\text{-dB}}^2}{4} \right) \Big|_{\text{dB}} = \alpha_1 \Big|_{\text{dB}} - 1 \text{ dB} \quad (2.55)$$

We can rewrite (2.55) in terms of decibels:

$$20 \log \left| \alpha_1 + \frac{3\alpha_3 A_{1\text{-dB}}^2}{4} \right| = 20 \log |\alpha_1| - 20 \log 1.122 \quad (2.56)$$

From (2.56) the $A_{1\text{-dB}}$ input level is given by

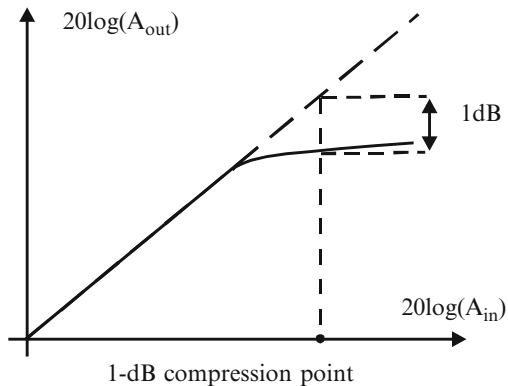
$$A_{1\text{-dB}} = \sqrt{0.145 \left| \frac{\alpha_1}{\alpha_3} \right|} \quad (2.57)$$

The idea of the 1-dB compression point is shown graphically in Figure 2.10.

2.4.1.6 Blocking

A phenomenon closely related to gain compression is blocking. So far, the compressive behavior of a nonlinear subcomponent with a single input signal has been discussed. What happens if a weak desired signal along with a strong interferer occurs at the input of a compressive subcomponent? Assume that we have the input signal $s(t) = A_0 \cos \omega_0 t + A_1 \cos \omega_1 t$ where A_1 is a strong interferer and A_0 is the

Fig. 2.10 Illustration of the 1-dB compression point



desired signal. Applying this $s(t)$ to (2.27) we can express the output terms of interest (the term at the fundamental frequency) as

$$y(t) = \left(\alpha_1 A_0 + \frac{3\alpha_3 A_0^3}{4} + \frac{3\alpha_3 A_0 A_1^2}{2} \right) \cos \omega_o t + \dots \quad (2.58)$$

If the interferer strength is much greater than the desired signal strength (that is, $A_1 \gg A_0$), equation (2.58) can be simplified as

$$y(t) = \left(\alpha_1 + \frac{3\alpha_3 A_1^2}{2} \right) A_0 \cos \omega_o t + \dots \quad (2.59)$$

If α_3 is negative, the small signal gain is attenuated by the interferer. If the attenuation becomes so large that the overall gain drops to zero, we say that the signal is blocked. Many receivers must be able to withstand blocking signals up to 70 dB greater than the desired signal.

To summarize, the effect of gain compression and blocking is that the desired signal amplitude is reduced, and this results in a degraded $\text{SNR}_{\text{demod_in}}$.

2.4.2 Noise

There is circuit noise internal to the subcomponents in the front end. This noise will add on to the AWGN, cause interference, and further degrade the SNR. To explore this effect we have to treat the circuit noise aspect in a more comprehensive manner. First we talk about noise sources. Circuit noise is associated with the electrical components that build the subcomponents, such as resistors and transistors. Circuit noise is further subdivided into thermal, shot, and flicker noise.

2.4.2.1 Noise sources

In this subsection we define the effects of the circuit noise. The noise phenomena considered here are caused by the small current and voltage fluctuations that are generated within the devices themselves.

Thermal noise

Thermal noise basically arises due to the random thermally generated motion of electrons. It occurs in resistive devices and is proportional to the temperature. Fundamentally, the thermal energy of electrons causes them to move randomly, thus causing local concentrations of electrons. This net concentration of negative charge in a local spot (balanced by net concentration of positive charge in another spot, as the total charge must remain zero) will result in a local nonzero voltage. As the concentration changes randomly, the resulting voltage also changes randomly, resulting in a noiselike behavior. Note that the noise exists even though the resistor is not connected and no current is flowing through it (as opposed to shot noise, to be discussed in the next subsection). It is white and has a flat power spectral density (PSD) whose value can be given as follows (depending on whether it is modeled by its Thévenin or Norton equivalent):

$$\frac{\overline{v^2}}{\Delta f} = 4kTR \quad \text{or} \quad \frac{\overline{i^2}}{\Delta f} = 4kT \frac{1}{R} \quad (2.60)$$

Here $\overline{v^2}$ and $\overline{i^2}$ are the mean square noise voltage and current, respectively; k is the Boltzmann constant; T is the temperature in Kelvin; and R is the resistance value. The unit is V^2/Hz or A^2/Hz

Shot noise

Shot noise occurs in all energy barrier junctions, namely in diodes and bipolar transistors. Actually, it happens whenever a flux of carriers (possessing potential energy) passes over an energy boundary. Since the potential energy of the carrier is random, the number of carriers that possess enough energy to cross the barrier is random in nature, resulting in a flux (current if the carrier are electrons) whose density is also random in nature. This random nature gives rise to shot noise. It is thus obvious that shot noise exists only when there is a current, as opposed to thermal noise. For example, a bipolar junction transistor (BJT) is a device whose current I is composed of holes and electrons that have sufficient energy to overcome the potential barrier at the junction. Thus, the current consists of discrete charges and not of continuum of charges. The fluctuation in the current I is termed shot noise.

This current I is composed of random pulses with average value I_{DC} . Now it can be shown that shot noise is white with a flat PSD whose value is given as follows:

$$\frac{\overline{i^2}}{\Delta f} = 2qI_{DC} \quad (2.61)$$

Here q is the charge of an electron in coulomb and I_{DC} is the value of the dc current in amperes.

Flicker noise

Flicker noise or $1/f$ noise arises from random trapping of charge at the oxide-silicon interface of MOS transistors and in some resistive devices. Obviously, the more current is flowing in a nonideal silicon interface, the higher is the rate of electrons that are trapped. On the other hand, at high frequency the electrons are varying too fast to be trapped and no noise-like current variation is caused. Consequently, the noise density is given by

$$\frac{\overline{i^2}}{\Delta f} = K \frac{I_{DC}^\alpha}{f} \quad (2.62)$$

where K and α (ranging from 0.5 to 2) are constants that depend on the nature of the device. I_{DC} is the dc current in amperes and f is frequency in hertz. As we can see from (2.62), flicker noise is most significant at low frequencies. However, it can still be troublesome for frequencies up to a few megahertz.

Additive Noise Versus Phase Noise

Now we discuss the manifestation of circuit noise in the front end. The noise of a subcomponent manifests itself as either additive noise or phase noise. Additive noise is described by noise adding to the amplitude of the desired signal. Phase noise is noise adding to the phase of the desired signal. In a front end the noise inherent in a LNA, and a mixer is best described by additive noise, whereas the noise inherent in a frequency synthesizer is best described by phase noise. Both will degrade the $SNR_{\text{demod_in}}$ and hence sensitivity of the overall receiver chain. In this chapter we will focus on the additive noise of the LNA and mixer blocks. In Chapters 7 and 8 we return to the discussion of phase noise.

2.4.2.2 Noise Figure

A parameter called noise figure (NF) is a commonly used method of specifying the additive noise inherent in a circuit or system. Use of this parameter is limited to situations where the source impedance is resistive. However, this is often the case in

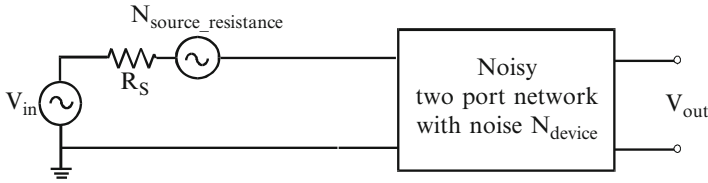


Fig. 2.11 Model used in noise figure calculations

a front end, and so this method of specifying noise is adopted here. The noise figure describes how much the internal noise of an electronic element degrades its SNR. It is often specified for a 1 Hz bandwidth at a given frequency. In this case, the noise figure is also called the spot noise figure to emphasize the very small bandwidth as opposed to the average noise figure, where the band of interest is taken into account. The interpretation should be clear from the context. Mathematically the noise figure is defined as

$$NF = \frac{SNR_{in}}{SNR_{out}} = \frac{S_{in}N_{out}}{S_{out}N_{in}} \quad (2.63)$$

where N_{in} is the input noise power and is always taken as the noise in the source resistance; and N_{out} is the output noise power including the circuit contribution and noise transmitted from the source resistance. By inserting $S_{out} = GS_{in}$ into (2.63), where G is the power gain of the corresponding stage, we get

$$NF = \frac{N_{out}}{GN_{in}} \quad (2.64)$$

Let us refer to Figure 2.11, the model for NF calculation. In Figure 2.11,

$$N_{in} = N_{source_resistance} \quad (2.65)$$

N_{out} , the noise occurring at the output, is given by the input noise multiplied by its power gain G plus the additional device noise:

$$N_{out} = N_{Device} + G \cdot N_{Source_resistance} \quad (2.66)$$

Substituting (2.65), and (2.66) into (2.64), we have

$$NF = \frac{N_{Device} + G \cdot N_{Source_resistance}}{G \cdot N_{Source_resistance}} \quad (2.67)$$

As a final remark, note that the noise figure is specified by a power ratio and given in decibels. We usually refer to the corresponding numerical ratio as the noise factor. Thus, the relation between the two is given by noise figure = $10 \log_{10}(\text{noise factor})$. Which one we are referring to should be clear from the context.

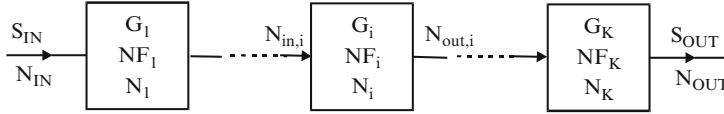


Fig. 2.12 Cascade of k gain stages, each with noise

2.4.2.3 Cascaded Noisy Stages

The noise figure of a cascade of noisy stages can be derived in terms of the noise figures of the individual blocks. Consider Figure 2.12 where a cascade of noisy stages each with available power gain G_i and noise figure NF_i is shown. N_i , $N_{in,i}$, $N_{out,i}$ specify the device noise, input noise, output noise, respectively, of the i th stage. Since we have a number of stages, we further make the assumptions that we have the same source impedance for each stage. This assures that we have the same input noise $N_{in,i}$ occurring at the input of each stage. Referring to Figure 2.12 the derivation consists of the following four steps:

1. We find the output power, which is given by the input signal power multiplied by product of the power gain of each stage.

$$S_{out} = S_{in}(G_1 G_2 \cdot \dots \cdot G_k) \quad (2.68)$$

2. We want to relate the noise N_i occurring in each stage to the corresponding noise figure of that stage. We start by applying (2.64) to stage 1 and we have

$$NF_i = \frac{S_{in,i} N_{out,i}}{S_{out,i} N_{in,i}} = \frac{(G_1 \cdot \dots \cdot G_{i-1}) S_{IN} N_{out,i}}{(G_1 \cdot \dots \cdot G_i) S_{IN} N_{in,i}} = \frac{N_{out,i}}{G_i N_{in,i}} \quad (2.69)$$

Then we apply (2.66) to the i th stage and we have

$$N_{out,i} = N_i + G_i N_{in,i} \quad (2.70)$$

But what is $N_{in,i}$? This, being the input noise of the i th stage, comes from the input resistance of the i th stage. Since we assume that every stage has the same source resistance, then every stage has the same input noise. For ease of reference, we label all of them by the input noise of stage 1, which from Figure 2.12 is simply given by N_{IN} . Therefore, $N_{in,i} = N_{IN}$. Substitute this in (2.70), we have

$$N_{out,i} = N_i + G_i N_{IN} \quad (2.71)$$

Now we can substitute (2.71) in (2.69) and we have

$$NF_i = \frac{N_i + G_i N_{IN}}{G_i N_{IN}} = \frac{N_i}{G_i N_{IN}} + 1 \quad (2.72)$$

Finally, since we are interested in relating N_i to NF_i , we can rearrange (2.72) as

$$N_i = (NF_i - 1)G_i N_{IN} \quad (2.73)$$

We want to find the output noise power of this cascade of stages. We will use (2.66) and iteratively apply this formula, starting from the first stage and then to a combination of first and second stage, until we include all k stages. The resulting output noise is the output noise from the k th stage, $N_{out,k}$. From Figure 2.12 we know that $N_{out,k}$ is also denoted as N_{out} . The final expression therefore is given as

$$N_{OUT} = (G_1 G_2 \cdots G_K) N_{IN} + N_1 (G_2 \cdots G_K) + N_2 (G_3 \cdots G_K) + \cdots + N_{K-1} G_K + N_K \quad (2.74)$$

We then substitute (2.73) from step 2 in (2.74) and we get

$$N_{out} = (G_1 G_2 \cdots G_K) N_{in} + (NF_1 - 1)(G_1 G_2 \cdots G_K) N_{in} + \cdots + (NF_K - 1) G_K N_{in} \quad (2.75)$$

4. Finally we want to derive the total noise figure of the cascaded chain. To do this we substitute (2.68) from step 1, and (2.75) from step 3 into (2.63). Canceling S_{in} , N_{in} and doing the proper simplification, we get

$$NF = NF_1 + \frac{NF_2 - 1}{G_1} + \frac{NF_3 - 1}{G_1 G_2} + \cdots + \frac{NF_k - 1}{G_1 G_2 \cdots G_{k-1}} \quad (2.76)$$

So we have finally derived the equation that relates the total noise figure to the individual noise figure. This equation is called the Friis formula. Note that the NF used in the Friis formula is specified as a ratio and not in decibels. Equation (2.76) predicts that NF is dominated by the first stage NF, NF_1 . When compared with (2.48) we see that both noise and distortion are dominated by one stage. The only difference is that for noise this is the first stage whereas for distortion this is the last stage. Of course, a front end with a small NF and a large IIP₃ is desirable.

2.5 Derivation of NF, IIP₃ of Receiver Front End

We have now finished discussing the NF and IIP₃ of the receiver front end. We have also developed formulas that relate the IIP₃ and NF of this receiver front end to those of the individual subcomponents. Our next step is to translate the boundary conditions on this front end, imposed by wireless standards and derived in Chapter 1, to its key design parameters. We assume that the key design parameters are IIP₃, NF, G . Boundary conditions such as SNR_{demod_in} , and P_{min} (communication

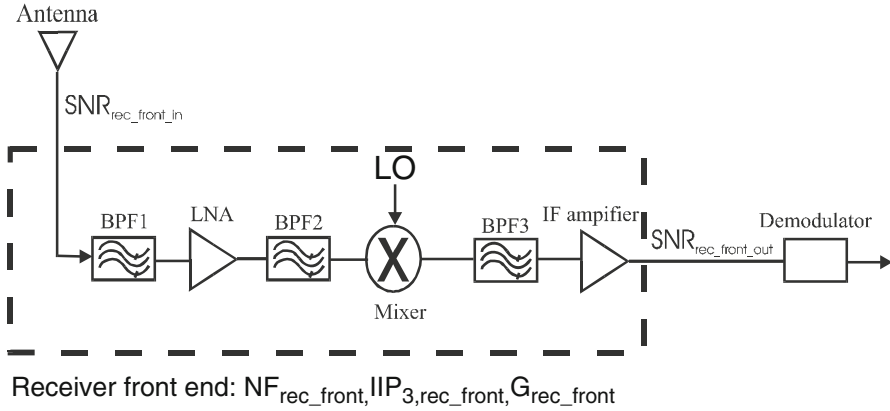


Fig. 2.13 Receiver block diagram showing the G , NF , IIP_3 and SNR of the front end

terminology) that we derived in Chapter 1 are now translated into the required IIP_3 and NF (circuit terminology) of the front end.

To illustrate this process we use DECT as an example standard and we use the heterodyne architecture as the example architecture. We will have a chance to redo this using zero IF (homodyne) architecture in Problem 2.9. The heterodyne architecture is redrawn in Figure 2.13, where as in Figure 2.2 we separate it into two parts: receiver front end followed by the demodulator. The receiver front end consists of the part between the antenna and the demodulator and contains the LNA, mixer, IF amplifier, and so forth. It is characterized by having its own gain G_{rec_front} , noise figure NF_{rec_front} , and third order intercept point IIP_{3rec_front} . This front end takes an input signal from the antenna with a signal to noise ratio denoted as $SNR_{rec_front_in}$, processes it, and generates a signal at its output with a signal to noise ratio denoted as $SNR_{rec_front_out}$. As stated previously, our goal in this section is to find the required G_{rec_front} , NF_{rec_front} , and IIP_{3rec_front} of this front end. To simplify notation, unless otherwise specified all parameters used in this section are to be interpreted in terms of decibels.

2.5.1 Required NF_{rec_front}

In this subsection, we want to find the required NF_{rec_front} . Let us apply (2.63), interpreted in decibels, to Figure 2.13. We have

$$NF_{rec_front} = SNR_{rec_front_in} - SNR_{rec_front_out} \quad (2.77)$$

Step 1: Calculate $SNR_{rec_front_out}$.

From Section 2.2, boundary condition (2) we know that

$$SNR_{rec_front_out} = 25dB. \quad (2.78)$$

Step 2: Calculate $SNR_{rec_front_in}$

$SNR_{rec_front_in}$ should be calculated under the worst-case situation. Obviously, the worst case happens when the received signal from the antenna is at its minimum and the noise is at its maximum. We have minimum signal power when the receiver is farthest away from the base station. From Section 2.2, boundary condition (1) we know that $P_{min} = -77$ dBm. Hence

$$S_{rec_front_in} = -77dBm \quad (2.79)$$

At this point we want to calculate the AWGN in the channel. This noise is hard to calculate. However, we know for sure that since the antenna has a 50Ω resistive load, there will be thermal noise coming from this load. We further assume that this is the only AWGN in the channel, and we use (2.60) to calculate this noise. Applying (2.60), interpreted in decibels, and integrating throughout the noise bandwidth B , and we have

$$N_{rec_front_in} = 10 \log \overline{v^2} = 10 \log 4kTR_s B \quad (2.80)$$

Here $R_s = 50 \Omega$. How do we calculate B ? Referring to Figure 2.13, we note that the receiver front end has 3 different filters. We take the one with the narrowest bandwidth to define B . From subsection 2.3.3 we note that BPF3 has the narrowest bandwidth, whose bandwidth is given by (2.25) as $2\pi \times 1.728$ Mrad/s or 1.728 MHz.

Substituting this value into (2.80) and working in decibels we have

$$\begin{aligned} N_{rec_front_in} &= 10 \log_{10}(4kTR_s) - 10 \log_{10}(B/1Hz) = -174dBm + 62dB \\ &= -112 dBm \end{aligned} \quad (2.81)$$

Now from the SNR definition in decibels,

$$SNR_{rec_front_in} = S_{rec_front_in} - N_{rec_front_in} \quad (2.82)$$

Substituting (2.79), and (2.81) in (2.82), we have

$$SNR_{rec_front_in} = -77dBm - (-112 dBm) = 35dB \quad (2.83)$$

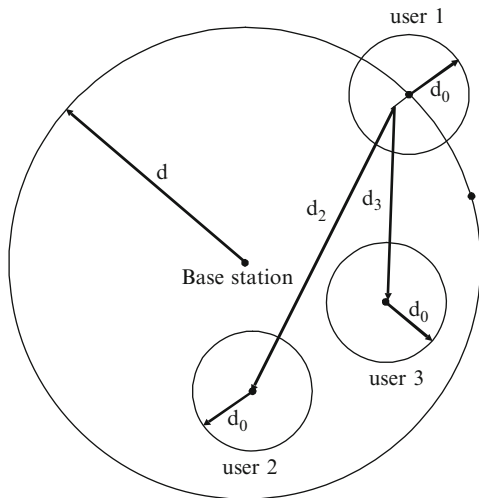
Step 3: Calculate NF_{rec_front}

Let us substitute (2.78), and (2.83) into (2.77). We get

$$NF_{rec_front} = 35dB - 25dB = 10dB \quad (2.84)$$

This is the required NF of the front end to satisfy the DECT standard.

Fig. 2.14 Four users are using their cellular phones at the same time, where d is the maximum range from user to the base station (dictated by standards; in our example this equals 400 m); d_0 is the free space distance; d_2 is the minimum allowable distance between user 1 and user 2, where user 2 uses the next channel; and d_3 is the minimum allowable distance between user 1 and user 3



2.5.2 Required IIP_{3,rec_front}

In this subsection we want to derive the required IIP_{3,rec_front_end} . The larger this IIP_3 , the smaller is the third-order intermodulation product generated by the interferers. We denote this intermodulation product, calculated at the output of the receiver front end, as I_{D3}'' . These interferers, as mentioned in subsection 2.4.1, have the same characteristics as that of narrowband noise. The maximum I_{D3}'' must be made small enough (achieved with a large enough IIP_{3,rec_front_end}) that its power is below the minimum $S_{rec_front_out}$ (minimum signal power at the output of the receiver front end), by a sufficiently large margin.

Step 1: State the minimum desired signal power and the maximum interferer power at the input of the front end.

To calculate the required IIP_3 , we must look at the condition when we have the minimum $S_{rec_front_out}$ and the maximum I_{D3}'' . This corresponds to the condition where at the input of the front end, the desired signal power is at a minimum, while the interferers are at their maximum.

From Section 2.1, boundary condition (1), the minimum desired signal power at the input of the front end is given as -77 dBm. To find the power of the interferer, we assume that the interferences are from adjacent channel interference. To understand adjacent channel interference, let us refer to Figure 2.14. Figure 2.14 shows three users, where user 1 uses the desired channel. Let us assume he is assigned the channel at 1.89 GHz. Next we assign the next two channels to users 2 and 3. Now let us consider user 3. User 3 is transmitting signal and interfering with user 1. From Figure 2.14 user 3 is closer to user 1 than user 2 and hence the interfering signal from him is larger. Hence he is placed farther away in frequency. Specifically, user 3 is assigned a channel at frequency $= 1.89 \text{ GHz} + 2 \times 1.728 \text{ MHz} = 1.8934 \text{ GHz}$. User 2 is then assigned a channel at frequency 1.8917 GHz. The maximum power

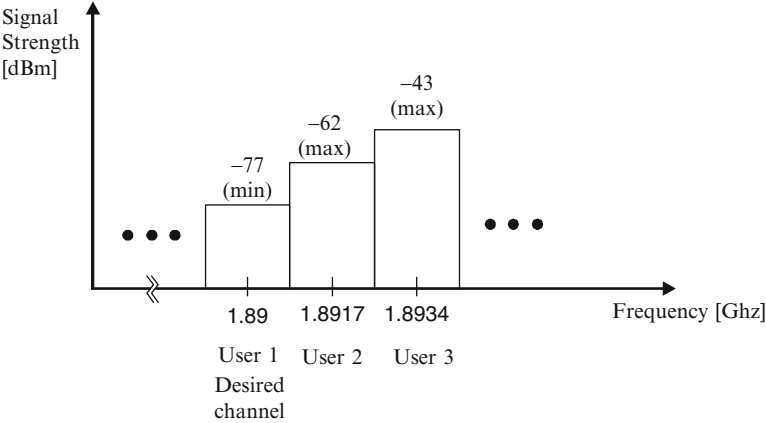


Fig. 2.15 The signal strength requirements of adjacent channels in the DECT standard

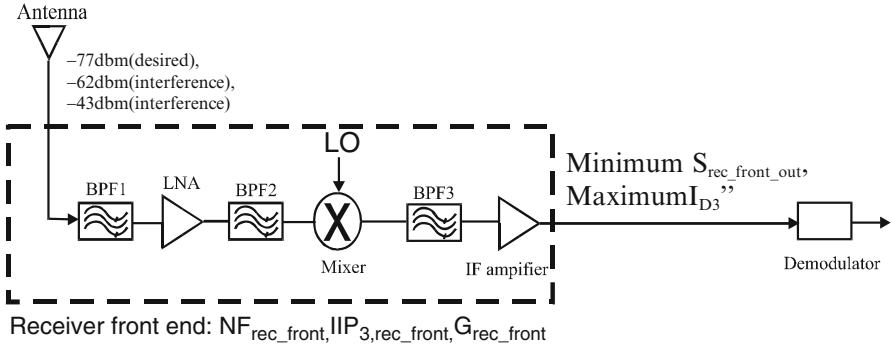


Fig. 2.16 Block diagram showing signal and interference level at input and output of receiver front end

from user 2 and user 3 is given in the DECT blocking specifications. This is shown in Figure 2.15. For example, for a channel away it is around -62 dBm [3]. Since we are interested in IIP_3 , we need to show only the two adjacent channels. Hence Figure 2.15 summarizes the minimum desired signal power and the maximum interferer power at the input of the front end that is required to calculate $\text{IIP}_{3,\text{rec_front}}$.

Step 2: Calculate the minimum $S_{\text{rec_front_out}}$

We now redraw Figure 2.13, but we specify the relevant power levels at the input and output of the front end. The resulting diagram is shown in Figure 2.16.

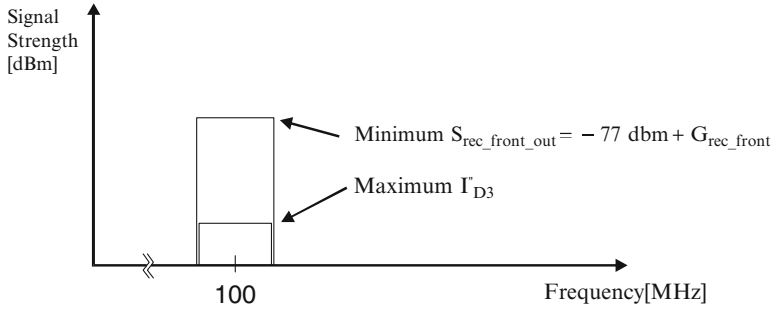


Fig. 2.17 Frequency spectrum at the output of the receiver front end

In Figure 2.16, at the input of the receiver front end the received signal is assumed to consist of the desired signal and two interferers. Their power are specified as

Minimum power of desire dsignal at 1.89 GHz = -77 dBm

Maximum power of interference from user 2 at 1.8917 GHz = -62 dBm

Maximum power of interference from user 3 at 1.8917 GHz = -43 dBm.

Next we determine minimum $S_{\text{rec_front_out}}$. Ignoring the nonlinearity of the receiver front end for a moment, we will see that for the desired signal at 1.89 GHz, interferers at 1.8917 GHz, and 1.8934 GHz all got mixed down to the IF. In (2.23) we chose the IF to be 100 MHz. Hence the interferers will be mixed down to 100 MHz, 101.7 MHz, and 103.4 MHz, respectively. Furthermore, from (2.24), and (2.25) we know that BPF3 has a center frequency of 100 MHz and a bandwidth of 1.7 MHz. Hence the interferers at 101.7 MHz and 103.4 MHz will be filtered out. The only signal at the front end output will be the one at 100 MHz. Since the receiver front end has a gain of $G_{\text{rec_front}}$, the signal at 100 MHz will have a power level given by -77 dBm + $G_{\text{rec_front}}$. Therefore we have

$$\text{Minimum } S_{\text{rec_front_out}} = -77\text{dBm} + G_{\text{rec_front}}$$

This is shown in Figure 2.16. The frequency spectrum is shown in Figure 2.17.

Step 3: Derive the maximum I_{D3} , in terms of the maximum power of interferers at the receiver front end input, P_i , and $IIP_{3,\text{rec_front}}$.

This step is subdivided into three substeps.

Step 3a: Derive equivalent block diagram for receiver front end and find the frequency spectrum for I_{D3} .

In step 2 we have ignored the nonlinearity of the front end. Let us bring back this nonlinearity and see what happens. First let us redraw Figure 2.16 in Figure 2.18(a).

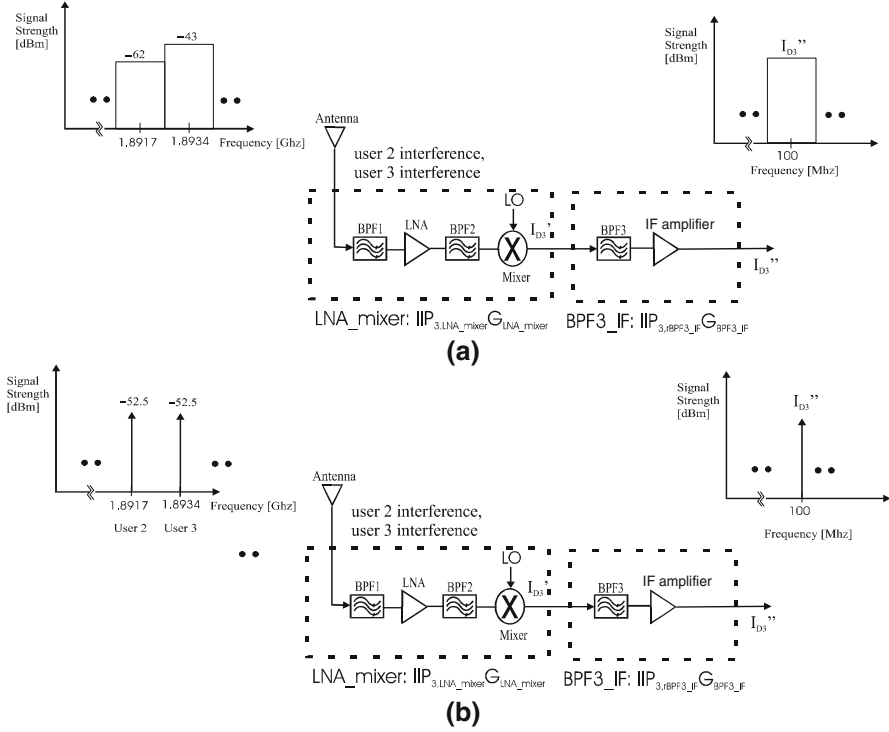


Fig. 2.18 Two equivalent representations of the front end: (a) Interferers represented as narrow-band noise. (b) Interferers represented as tones

Here we subdivide the receiver front end into two blocks: LNA_mixer block and BPF3_IF block. Each block has its own gain and IIP₃. Their gain relationship with $G_{\text{rec_front}}$ is given as

$$G_{\text{rec_front}} = G_{\text{LNA_mixer}} + G_{\text{BPF3_IF}} \quad (2.85)$$

We now concentrate on the LNA_mixer block, which consists of BPF1, LNA, BPF2, and the mixer. The nonlinearity in the LNA_mixer block, as characterized by IIP_{3,LNA_mixer}, will take the two interferers from user 2 and user 3 (whose frequencies are at 1.8917GHz and 1.8934 GHz, respectively, and at maximum power) and generate a third-order intermodulation product, denoted as I_{D3} (whose frequency is at 1.89 GHz and at maximum power). I_{D3} will then be mixed down by the mixer to an IF of 100 MHz. The mixed down I_{D3} is denoted as I_{D3}' . Now I_{D3}' , being at 100MHz, will not be filtered by BPF3. It will pass through the filter and got amplified by the IF amplifier and generate the I_{D3}'' (at maximum power). To save notations, we will also use I_{D3} , I_{D3}' and I_{D3}'' to denote the power level of the respective intermodulation products. Which meaning they refer to should be clear from the context. Let us further assume that BPF3 and IF amplifier are very linear. Accordingly, IIP_{3,BPF3_IF} is practically given as

$$\text{IIP}_{3,\text{BPF3_IF}} \cong \infty \quad (2.86a)$$

which means that IIP_{3,rec_front} is given by

$$IIP_{3,rec_front} \cong IIP_{3,LNA_mixer} \quad (2.86b)$$

Hence when we apply I_{D3}' to the BPF3_IF block no distortion occurs and no new frequency components are generated. Consequently, I_{D3}'' will also have only one frequency component at 100 MHz. This is shown in the frequency plot in Figure 2.17. Likewise this maximum I_{D3}'' is also shown in Figure 2.16. [We will show in Problem 2.4(c) that even if the approximation given in (2.86a) is not observed, the maximum I_{D3}'' derived in step 3 remains practically the same.]

Step 3b: Find P_i .

We know from step 3a that maximum interferences from user 2 and user 3 at the front end input got intermodulated and mixed to generate the maximum I_{D3}'' at the front end output (at 100 MHz). In subsection 2.4.1.2, we represented each interferer, which is like narrowband noise, by a tone (with the same power) in order to simplify the IM_3 calculation and subsequently the IIP_3 calculation. We will use the same representation here. Hence the interferer from user 2 at the front end input (at 1.8917 GHz and with a power of -62 dBm) in Figure 2.18 (a) is now represented by a tone at the same frequency (1.8917 GHz) and with the same power (-62 dBm) in Figure 2.18 (b). Similarly, the interferer from user 3 at the front end input (at 1.8934 GHz and with a power of -43 dBm) in Figure 2.18 (a) is now represented by a tone at the same frequency (1.8934 GHz) and with the same power (-43 dBm) in Figure 2.18 (b). To further simplify the derivation, instead of having two different power levels for two tones, we assign one power level, taken to be the average of the two, to both tones. We arbitrarily decide to take the geometric average of the two power levels, which equals -52.5 dBm and assign it to both tones. Hence we have

$$P_i = -52.5 \text{ dBm}. \quad (2.87)$$

Step 3c: Derive maximum I_{D3}'' in terms of IIP_{3,rec_front}

First we apply (2.38) to the LNA_mixer block of Figure 2.18(b) and we have

$$IIP_{3,LNA_mixer} = P_i - \frac{IM_3}{2} \quad (2.88)$$

Remember from (2.87) $P_i = -52.5$ dBm. Rearranging (2.88) and substituting this value for P_i , we have

$$IM_{3,LNA_mixer} = 2(P_i - IIP_{3,LNA_mixer}) = 2(-52.5 \text{ dBm} - IIP_{3,LNA_mixer}) \quad (2.89)$$

However, from the definition as given in (2.33) (interpreted in decibels)

$$IM_{3,LNA_mixer} = I_{D3}' - I_{D1}' \quad (2.90a)$$

To reiterate, I_{D3}' and I_{D1}' are the third-order intermodulation product and fundamental component generated by the two interferers which got mixed down and appear at the output of the mixer. In the present situation the two interferers are at the maximum power and so the intermodulation product and the fundamental component are also at their maximum power level. Rewriting (2.90a) under this situation, we have

$$IM_{3,LNA_mixer} = \text{maximum } I_{D3}' - \text{maximum } I_{D1}' \quad (2.90b)$$

Rearranging (2.90b), we have

$$\text{maximum } I_{D3}' = \text{maximum } I_{D1}' + IM_{3,LNA_mixer} \quad (2.91)$$

Since maximum I_{D1}' is the fundamental component generated by the maximum interferer with a power level of -52.5 dBm, it is given by

$$\text{maximum } I_{D1}' = -52.5 \text{ dBm} + G_{LNA_mixer} \quad (2.92)$$

Now we can substitute (2.89), and (2.92) into (2.91), and we get

$$\begin{aligned} \text{maximum } I_{D3}' &= -52.5 \text{ dBm} + G_{LNA_mixer} \\ &+ 2(-52.5 \text{ dBm} - IIP_{3,LNA_mixer}) \end{aligned} \quad (2.93)$$

In discussing (2.86a) we stated that BPF3 and IF AMPLIFIER are practically linear. When we apply maximum I_{D3}' to the input of the BPF3_IF block, the maximum I_{D3}'' generated will contain only one frequency component at 100 MHz, whose level is given by

$$\text{maximum } I_{D3}'' = \text{maximum } I_{D3}' + G_{BPF3_IF} \quad (2.94)$$

Substituting (2.93) into (2.94), we get

$$\text{maximum } I_{D3}'' = -52.5 \text{ dBm} + G_{LNA_mixer} + 2(-52.5 \text{ dBm} - IIP_{3,LNA_mixer}) + G_{BPF3_IF} \quad (2.95)$$

Applying (2.85) and (2.86b) to (2.95), we have

$$\begin{aligned} \text{maximum } I_{D3}'' &= -52.5 \text{ dBm} + G_{rec_front} + 2(-52.5 \text{ dBm} - IIP_{3,rec_front}) \\ &= -3 \times 52.5 \text{ dBm} + G_{rec_front} - 2 \times IIP_{3,rec_front} \end{aligned} \quad (2.96)$$

This gives the maximum I_{D3}'' in terms of IIP_{3,rec_front} .

Step 4: Relate $SNR_{rec_front_out}$ to maximum I_{D3}'' and hence relate $SNR_{rec_front_out}$ to IIP_{3,rec_front} . From the required $SNR_{rec_front_out}$, calculate the required IIP_{3,rec_front} .

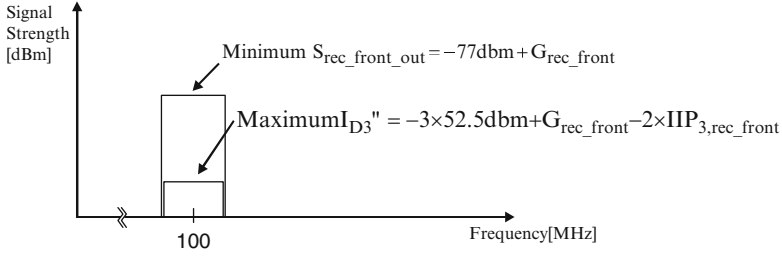


Fig. 2.19 Signal strength at the output of the receiver front end

In this step we refer again to Figure 2.18(b). As shown at the output of the receiver front end, we have I_{D3}'' , a tone at 100 MHz. At this point we want to express I_{D3}'' using a narrowband noise representation again. We denote this noise as $n_{I_{D3}''}(t)$. We first assume that $n_{I_{D3}''}(t)$, like the AWGN channel noise, is also additive and has a Gaussian distribution. The power of $n_{I_{D3}''}(t)$ is, of course, the same as I_{D3}'' and is still given by (2.96). We redraw Figure 2.17 as Figure 2.19, where the value of maximum I_{D3}'' [as expressed in (2.96)] is explicitly shown.

If we assume that there is no other noise (e.g., no AWGN from the channel, antenna, or circuit noise), then at the receiver front end output we have a signal immersed in noise and the SNR is given by

$$\text{SNR}_{\text{rec_front_out}} = \text{minimum } S_{\text{rec_front_out}} - \text{maximum } I_{D3} \quad (2.97)$$

Substituting the appropriate values for minimum $S_{\text{rec_front_out}}$ and maximum I_{D3}'' from Figure 2.19 into (2.97) and simplifying, we have

$$\text{SNR}_{\text{rec_front_out}} = 80.5\text{dBm} + 2 \times \text{IIP}_{3,\text{rec_front}} \quad (2.98)$$

This signal and noise is applied to the demodulator and hence we have

$$\text{SNR}_{\text{demod_in}} = \text{SNR}_{\text{rec_front_out}} = 80.5\text{dBm} + 2 \times \text{IIP}_{3,\text{rec_front}} \quad (2.99)$$

Now what is the required $\text{SNR}_{\text{demod_in}}$? To derive the required $\text{SNR}_{\text{demod_in}}$ with this noise, $n_{I_{D3}''}(t)$, we can replace $n(t)$ in (6.14), Chapter 1 by $n_{I_{D3}''}(t)$, so that (6.14) now becomes

$$x_R(t) = \alpha(t)s_1(t)\exp(-j\theta(t)) + n_{I_{D3}''}(t) \quad (2.100)$$

This equation is not to be interpreted in decibels. We can then start from (2.100) and go through the rest of the derivation in subsection 6.2.2, Chapter 1, and derive the required $\text{SNR}_{\text{demod_in}}$ for DECT. Since the noise characteristics are assumed to

be the same as the AWGN in the channel, the required $SNR_{\text{demod_in}}$ for DECT should stay the same, which is given in (1.51), Chapter 1, as 25 dB. Hence if we substitute 25 dB for $SNR_{\text{rec_front_out}}$ in (2.99) we get

$$25 \text{ dB} = 80.5 \text{ dBm} + 2 \times IIP_{3,\text{rec_front}} \quad (2.101)$$

Solving, we get

$$IIP_{3,\text{rec_front}} = -27.75 \text{ dBm} \quad (2.102)$$

This is the required IIP_3 of the front end to satisfy the DECT standard.

2.6 Partitioning of required $NF_{\text{rec_front}}$ and $IIP_{3,\text{rec_front}}$ into individual NF, IIP_3

Our strategy here is to start with a set of power gains, NF, IIP_3 , of the individual stages, based on some existing receiver front end. This will provide us with an initial design. We then calculate the $NF_{\text{rec_front}}$ and $IIP_{3,\text{front_end}}$ of this existing front end and see if it satisfies the required $NF_{\text{rec_front}}$ and $IIP_{3,\text{front_end}}$. Iterations can be carried out if necessary. In subsequent chapters, we go through the actual design of these stages and figure out if the power gains, NF, IIP_3 of the individual stages are achievable. Further iterations can then be carried out if necessary.

Before we carry out this strategy, there is one more point to be noted. Up to now we have assumed that the power gain G_i is given as the square of the voltage gain. However, this is only true if termination resistances are the same. In general, G is given by

$$G_i = A_{v,i}^2 \cdot \frac{R_{\text{in},i}}{R_{\text{out},i}} \quad (2.103)$$

where G_i , $A_{v,i}$, $R_{\text{in},i}$, $R_{\text{out},i}$ are the power gain, voltage gain, and input and output termination resistance of the i th stage, respectively [4]. Therefore, instead of specifying G , we would specify the termination resistances and the voltage gain of the individual stages and then calculate the corresponding G .

Step 1: Specify the voltage gain and termination resistance of the subcomponents. The heterodyne architecture is redrawn in Figure 2.20 with the corresponding termination resistances. BPF1 and BPF2 both need 50 Ω termination resistance; otherwise the filters would lose their frequency responses. For this example we choose a BPF3 that has $R_{\text{in}} = 1 \text{ k}\Omega$ and $R_{\text{out}} = 1 \text{ k}\Omega$. The LNA is specified to have input and output resistances of 50 Ω . The output resistance of the mixer is much lower than its input resistance. In this case it is a nice feature since it maximizes the voltage gain of the mixer. Thus the mixer is specified with an input resistance of 50 Ω

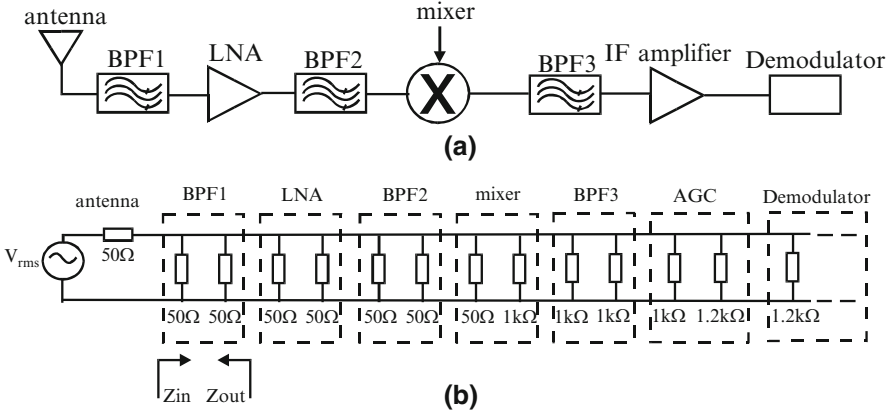


Fig. 2.20 a) The heterodyne architecture b) The heterodyne architecture with the corresponding input and output impedance of the different subcomponents

and an output resistance of 1 kΩ. Finally, the input resistance of the demodulation block is specified to be 1.2 kΩ.

The voltage gains of the various subcomponents are given in the second row of Table 2.1. The power conversion gains can now be calculated using equation (2.103) and are given in the third row of Table 2.1. Overall conversion gain G can be calculated by summing all the terms in row 3 and we have

$$G = 31.6 \text{ dB} \quad (2.104)$$

Step 2: Specify a possible set of NF of the subcomponents that meets the required $NF_{\text{front_end}}$.

First, we need to specify the noise figure of the individual subcomponents. An initial set of values is given in row 4 of Table 2.1. The general philosophy is for the early stages of the front end (basically the LNA) to dominate the NF (i.e., the early stages should have good NF [small values] and the latter stages can afford to have poorer NF [larger values]). The exact values (i.e. 3 dB for LNA and 12 dB for mixer) depend on the circuit and technology details and will be covered in later chapters.

Substituting the G and NF values from rows 3 and 4 of Table 2.1 into the Friis' formula (2.76) and carrying out calculation in ratio (not in decibels), we have the noise contributions calculated for individual components, which are shown in row 5 of Table 2.1.

Finally, we get the total noise figure by adding the corresponding components of the Friis formula.

$$\therefore NF_{\text{rec_front}} = 11.54 \text{ or } 10.6 \text{ dB} \quad (2.105)$$

Table 2.1 Characteristics of the subcomponents used in the receiver of Figure 2.20

	BPF1	LNA	BPF2	Mixer	BPF3	IF AMPLIFIER
A_v (isolated components) [dB]	-2	12	-2	10	-2.6	30
Power Conversion Gain G (isolated components)[dB]	-2	12	-2	-3	-2.6	29.2
NF (isolated components) [dB]	2	3	2	12	6	8
Components of Friis formula	1.58	1.587	0.058	2.368	0.95	3.073
IIP_3 (isolated components) [dBm]	90	-10	90	-10	90	20
Components of overall IIP_3 formula (2.48)	ignored	$(0.15 \cdot 10^{-3})^{-1}$	Ignored	$(0.01 \cdot 10^{-3})^{-1}$	ignored	ignored

Comparing this value and the required NF_{rec_front} [calculated in (2.84) to be 10 dB], we see that we have selected a set of G , NF values for the subcomponents that allows the front end to meet the required NF.

Step 3: Specify a possible set of IIP_3 of the subcomponents that meets the required $IIP_{3,front_end}$.

As in the NF case, we need to specify the IIP_3 of the individual subcomponents. The values are given in row 6 of Table 2.1. The general philosophy is for the latter stages of the receiver (basically the mixer) to dominate the distortion. The exact values (i.e. -10 dBm for LNA and -10 dBm for mixer) again depend on the circuit and technology details and will be covered in later chapters.

Substituting G from row 3 and IIP_3 from row 6 into the overall IIP_3 formula, (2.48), and carrying out calculation in ratio (not in decibels), we have the IIP_3 contributions calculated for individual components, which are shown in row 7 of Table 2.1. Finally, we get the IIP_3 by adding the corresponding components in row 7 and then taking the reciprocal.

$$\begin{aligned}
 \therefore IIP_{3,rec_front} &= \left(\left\{ 10 \log \left[1 / \left((0.15 \cdot 10^{-3})^{-1} + (0.01 \cdot 10^{-3})^{-1} \right) \right] \right\} + 30 \right) \text{dBm} \\
 &= -20 \text{ dBm}
 \end{aligned} \tag{2.106}$$

Comparing this value and the required IIP_{3,rec_front} [calculated in (2.102) to be -27.75 dBm], we see that we have selected a set of G , IIP_3 values for the subcomponents that allows the front end to meet the required $IIP_{3,front_end}$.

2.7 Problems

- 2.1 For DECT's standard, for worst-case reception, calculate the SNR at the input of the demodulator (Figure 2.1). We assume no front end is used and the only AWGN in the channel comes from a $50\ \Omega$ input resistance.
- 2.2 In the chapter we discussed that we want to go for a fixed f_{if} (and hence a variable f_{io}) scheme because this allows the use of a fixed frequency BPF2, which is easier to implement. We showed how, with this scheme, image can become a problem. A system designer suggests that we should instead use a variable f_{if} , fix f_{io} scheme, together with a fixed frequency BPF2 because he believes this will fix the problem. Is he correct? To answer this you can follow parts (a), (b), and (c). Assume that the receive band spans from 824 to 894 MHz.
- Assume that we use a fixed f_{if} of 10 MHz and a variable f_{io} . Draw the frequency spectrum, including all relevant frequencies, when f_{if} is 894 and 824 MHz. Repeat the case for f_{if} of 100 MHz.
 - Now assume that we use a variable f_{if} and a fixed f_{io} of 760 MHz. Again, draw the spectrum when f_{if} is 894 and 824 MHz. Repeat the case for f_{io} of 850 MHz.
 - Now determine the f_{image} for all the cases in (a) and (b) and comment on whether scheme (b) is better than scheme (a) as far as making it easier for BPF2 to filter out the image.
- 2.3 This problem concerns the qualitative understanding of the nonlinear behaviour of the receiver front end.
- We mentioned that harmonic distortion is not an issue for the heterodyne architecture described in Figure 2.2 since the bandpass filters (BPF1, 2, 3) will filter them out. What happens if they do not filter them out completely?
 - We have shown mathematically what blocking is. Explain, from first principle and in words (no equations), the mechanism of blocking. Offer an explanation in words (no equations) that distinguishes how blocking and intermodulation of interferers affect $SNR_{rec_front_out}$ differently.
 - Is it possible to have a receiver front end such that it generates poor IM_3 but does not block?
- 2.4 This question clarifies subtleties encountered in Section 2.5.
- At the beginning of Section 2.5, we assumed that IIP_3 is the only key design parameter as far as characterizing the receiver front end's distortion performance. Comment on the validity of this assumption.
 - In (2.81) we used 1.728 MHz (=bandwidth of BPF3) as the noise bandwidth B . Comment on the validity of this assertion.
 - In step 3 of subsection 2.5.2 we said that BPF3's and IF amplifier's nonlinearity does not matter [refer to (2.86a)]. Justify this.

- 2.5 In Section 2.6 we stated that power gain G is the same as the voltage gain squared a_v^2 only if $R_L = R_S$, otherwise it is given by (2.103). Derive (2.103) and show that it is only true if the input of the i th stage is matched to the output of the previous stage [the $(i-1)$ th stage].
- 2.6 Calculate the individual components of the Friis formula in Table 2.1.
- 2.7 For the heterodyne architecture whose subcomponents are described in Table 2.1, plot the change of the overall NF as a function of the isolated NF of the LNA with a fixed voltage gain of 12 dB and as a function of voltage gain of the LNA with a fixed NF of 3 dB.
- 2.8 Let us reexamine the architecture described in Figure 2.20 and Table 2.1 as follows: we change the IF amplifier's IIP_3 to be -20 dBm. Now the mixer, and the IF amplifier have $IIP_3 = -10$ dBm, -20 dBm respectively. Hence the IF amplifier is poorer than the mixer in terms of IIP_3 . On the other hand, the OIP_3 of the IF amplifier, even with this new IIP_3 value, is 9.2 dBm. This remains better than the OIP_3 of the mixer (which is -13 dBm). Which component is better? Explain.
- 2.9 In this problem we deal with a different standard and a different receiver front end architecture. The standard is given as follows: the carrier frequency is 800 MHz and the channel bandwidth is 200 kHz. The input signal plus interference spans from -104 dBm to -10 dBm. Assume that a BPSK modulation/demodulation scheme is used and the required BER is 10^{-3} . Furthermore, TDMA and FDMA are used to separate the different users. To simplify matters, we will neglect the impact of fading in our calculation.

Figure P.2(a) and Figure P.2(b) show the block diagram describing a different receiver front end architecture (called zero IF or homodyne receiver architecture), together with a diagram that specifies the input and output impedance of each subcomponent (matched).

The receiver front end takes the signal, filters it with a BPF for anti-aliasing, and amplifies it with a LNA. (The LNA is assumed to have a output matching network that performs further noise rejecting filtering. This output matching network [noise rejection filter] is included in the LNA block and not explicitly shown. The noise it is rejecting comes mainly from the LNA itself.) The signal is then fed into an automatic gain circuit (AGC) which is then converted by the A/D converter. The sample and hold circuit in the A/D converter (to be discussed in Chapter 6) acts as a sampling mixer (to be discussed in Chapter 5). This sampling mixer mixes the signal to DC before getting it quantized. That is why this is called a zero IF architecture. It should be noted that the mixing to baseband (zero IF) is inherently performed as the sampling operation of the A/D converter. In this problem, subsampling (or bandpass sampling) is performed. This means that the sampling frequency in the A/D converter is much lower than the carrier frequency (800 MHz) and at least larger than the Nyquist frequency of the baseband signal (2×200 kHz). As a side note, as a result of using subsampling, noise from the channel and LNA at around multiples of the subsampled frequency will be aliased and hence there is the

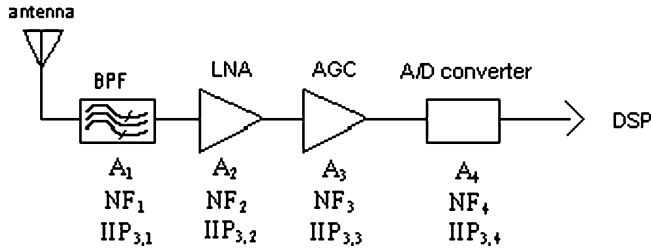


Fig. P.2(a) Receiver architecture assumed in this example, where A_i denotes the voltage gain, NF_i denotes the isolated noise figure and $IIP_{3,i}$ denotes the isolated third-order intercept point (power) at the input ($i = 1, \dots, 4$)

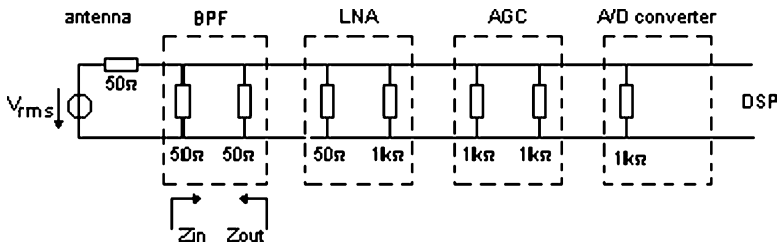


Fig. P.2(b) Input and output impedance are characterized for each subcomponent

need for BPF and noise rejection filter. Further demodulation and other signal processing is done in a digital signal processor (DSP) at DC.

- Find the NF required of the receiver front end.
- Find the NF of the A/D converter. An A/D converter is characterized by another type of additive noise: quantization noise. As shown in Chapter 6,

the noise is white and the spectral density of noise is given by $\frac{\overline{v_{n,out}^2}}{\Delta f} = \frac{\Delta^2}{12 \times f_b}$ where Δ = step size of the least significant bit (LSB) and is given by $V_{FS}/2^n$. Here V_{FS} is the full-scale voltage and n is the number of bits in the A/D converter. f_b is the bandwidth of the A/D converter. Suppose $V_{FS}=3.13V$, $n=12$. What should f_b be? What is the gain of the A/D converter? What is $\frac{\overline{v_{n,out}^2}}{\Delta f}$ and what is the NF of the A/D converter?

- Assume that the following information for the individual stage is given:

$$NF_1 = 2 \text{ dB} \quad A_1 = -2 \text{ dB} \\ A_2 = 20 \text{ dB}$$

$$NF_3 = 10 \text{ dB} \quad A_3 = 31 \text{ dB}$$

Calculate the noise figure for the LNA.

References

1. Rudell, J.C.; Ou, J.-J.; Cho, T.B.; Chien, G.; Brianti, F.; Weldon, J.A.; Gray, P.R, "A 1.9-GHz Wide-Band IF Double Conversion CMOS Receiver for Cordless Telephone Applications," *IEEE Journal of Solid State Circuits*, Vol. 32, No. 12, P. 2071–2088, December 1997.
2. D. Johns and K. Martin, *Analog Integrated Circuit Design*, Wiley, 1997.
3. European Telecommunication Standard, ETS 300 175–1, October 1992.
4. G. Gonzalez, *Microwave Transistor Amplifiers*, 2nd ed., Prentice Hall, 1997.
5. P. Gray, P. Hurst, S. Lewis, R. Meyer, "Analysis and design of analog integrated circuits", 5th edition, Wiley, 2009.



<http://www.springer.com/978-1-4614-0985-4>

VLSI for Wireless Communication

Leung, B.

2011, XXVI, 546 p., Hardcover

ISBN: 978-1-4614-0985-4