الجامعة الافتراضية السورية

SYRIAN VIRTUAL UNIVERSITY

نظم الاتصالات الراديوية والمكروية

الدكتور خالد يزبك

ISSN: 2617-989X

Books & Refrences

نظم الاتصالات الراديوية والمكروية

الدكتور خالد يزبك

من منشورات الجامعة الافتراضية السورية

الجمهورية العربية السورية ٢٠٢٠

RF & Microwave telecommunications systems

Khaled Yazbek

Publications of the Syrian Virtual University (SVU)

Syrian Arab Republic, 2020

الفهرس:

Chapter One:

Review of Transmission Line Theory and Scattering Parameters

Keywords

lossless transmission line electrical length characteristic impedance, reflection coefficient, Voltage Standing Wave Ratio, Matched load, quarter-wave transformer, S-parameters, N-port microwave network, Shift in Reference Planes, Power Waves, Generalized Scattering Parameters

Abstract

Transmission line theory and scattering parameters are essential for this course. We will review in this chapter the transmission line theory, where electrical size is important, and voltages and currents can vary in magnitude and phase over its length. The concept of characteristic impedance, input impedance, reflection coefficient, and standing waves will be revised. Microwave networks are characterized by scattering parameters. Then the scattering parameters of a microwave network will be defined and their physical meaning will be interpreted. Power waves and generalized scattering parameters will be revised.

Learning Objectives

1. Review transmission line theory.
2. Review microwave network analysis using S-parameters.

Introduction

Transmission lines are needed in a communication system. Antennas are connected to the transmitter and receiver via transmission lines; Subsystems of a communication system are also connected via transmission lines. On the other hand, we have seen that the design and implementation of different microwave networks is based on transmission lines.

Transmission line theory has been studied in details in chapter 5 of "electromagnetic waves and transmission lines" course. Here we will focus on essential characteristics and parameters of transmission line theory.

Analysis techniques of microwave networks have been studied in details in chapter 2 of "microwave engineering" course. Here we will focus on the scattering parameters and their physical meaning. Scattering parameters are used to characterize microwave networks and subsystems of an RF and microwave system.

Transmission line theory

Electrical size is the key difference between circuit theory and transmission line theory. Electrical size, or electrical length, is the size, or length of a transmission line, a component, or a circuit in terms of wavelength. Circuit theory assumes that the physical dimensions of the network are much smaller than the wavelength, while transmission lines may be a considerable fraction of a wavelength, or many wavelengths, in size. Thus a transmission line is considered as a distributed element, where voltages and currents can vary in magnitude and phase over its length, while ordinary circuit analysis deals with lumped elements, where their physical dimension are much smaller than the wavelength.

Wave Propagation on a Transmission Line

Voltages and currents consist of a superposition of traveling waves on a transmission line

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

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

where the $V_0^+ e^{-\gamma z}$ term represents travelling wave propagation in the $+z$ direction, and the $V_0^- e^{\gamma} z$ term represents travelling wave propagation in the $-z$ direction. γ is the complex propagation constant, which is a function of frequency:

$$\gamma = \alpha + j\beta = \sqrt{(R + j\omega L)(G + j\omega C)}$$

R represents losses is the conductor, and G represents losses is the dielectric. If the transmission line is lossless, then $R = G = 0$ and $\gamma = j\beta = j\omega\sqrt{LC}$, where $\alpha = 0$. Voltages and currents on a lossless transmission line becomes:

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

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

The characteristic impedance of a transmission line is defined as the ratio of the voltage and current travelling waves on the transmission line in one direction. Then we can write:

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

If the transmission line is lossless, then the characteristic impedance is real and equal to:

$$Z_0 = \sqrt{\frac{L}{C}}$$

The current wave along the transmission line can be written as

$$I(z) = \frac{V_0^+}{Z_0} e^{-j\beta z} - \frac{V_0^-}{Z_0} e^{j\beta z}$$

The terminated lossless transmission line

If a lossless transmission line is terminated in an arbitrary load impedance z_l as shown in figure 1, then wave reflection on the transmission line will occur, which is a fundamental property of distributed elements.

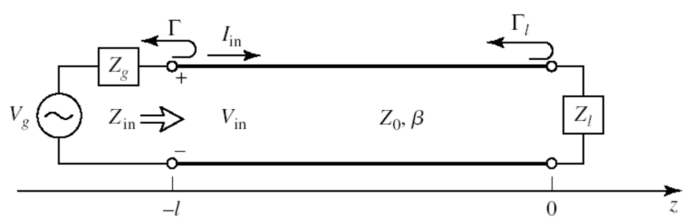


Figure 1: lossless transmission line terminated in an arbitrary load

The voltage reflection coefficient is the ratio of the amplitude of the reflected voltage wave $V_0^- e^{j\beta z}$ normalized to the amplitude of the incident voltage wave $V_0^+ e^{-j\beta z}$. At the load, the reflection coefficient is:

$$\Gamma_L = \Gamma|_{z=0} = \left.\frac{V_0^-}{V_0^+}\right|_{z=0} = \frac{Z_L - Z_0}{Z_L + Z_0}$$

The total voltage and current waves on the line can then be written as

$$V(z) = V_0^+ \left(e^{-j\beta z} + \Gamma_L e^{j\beta z} \right)$$

$$I(z) = \frac{V_0^+}{Z_0} \left(e^{-j\beta z} - \Gamma_L e^{j\beta z} \right)$$

such waves are called standing waves. Only when $\Gamma_L = 0$ is there no reflected wave. To obtain $\Gamma_L = 0$, the load impedance must be equal to the characteristic impedance of the transmission line

$$Z_L = Z_0$$

Such a load is said to be matched to the line since there is no reflection of the incident wave. Practically, this is an important case, where we need to have impedance matching.

Standing wave properties

The voltage magnitude $|V(z)|$ oscillates with position z along the line between maximum and minimum values. The maximum value is $V_{max} = |V_0^+|(1+|\Gamma_L|)$, and the minimum value is $V_{min} = |V_0^+|(1-|\Gamma_L|)$, The distance between two successive voltage maxima (or minima) is $\lambda/2$, while the distance between a maximum and a minimum is $\lambda/4$.

Voltage Standing Wave Ratio is defined as

Then $VSWR$ is a measure of the m$VSWR = \frac{V_{max}}{V_{min}} = \frac{1+|\Gamma_L|}{1-|\Gamma_L|}$ ismatch of a line.

Example 1

Standing wave pattern for a lossless transmission line of terminated in a load with a reflection coefficient $\Gamma_L = 0.3e^{j30°}$. The magnitude of the incident wave $|V_0^+| = 1\text{V}$. The standing–wave ratio is $VSWR = 1.3/0.7 = 1.86$.

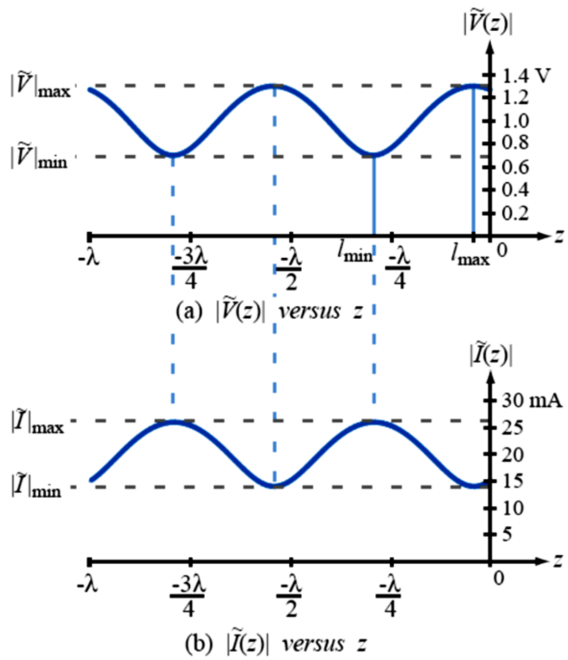


Figure 2: Standing wave pattern for example 1.

Input impedance concept of a transmission line

At a distance $\ell = -z$ rom the load, the input impedance seen looking toward the load is

$$Z_{in} = Z_0 \frac{1+\Gamma_{in}}{1-\Gamma_{in}} = Z_0 \frac{1+\Gamma_L e^{-2j\beta\ell}}{1-\Gamma_L e^{-2j\beta\ell}} = Z_0 \frac{Z_L + jZ_0 \tan\beta\ell}{Z_0 + jZ_L \tan\beta\ell}$$

This is an important result giving the input impedance of a length of transmission line with an arbitrary load impedance. The input impedance is function of Z_0, $\beta\ell$, and Z_L it is the equivalent impedance of the circuit composed of a transmission line defined by Z_0 $\beta\ell$ and a load Z_L , as seen by the generator. The **electrical length** of the transmission line $\beta\ell$ is equivalent to the phase delay of a travelling wave along a distance ℓ from the input to the load.

Special cases

- Matched load $Z_L = Z_0$

When we have matching between the line and the load, there is no reflection (all the available power from the source is delivered to the load), nor standing wave on the line. Then $\Gamma_L = 0$; $VSWR = 1$;. $Z_{in} = Z_0$

- Short-circuit $Z_L = 0$

When we have a short-circuit as a load, we will have total reflection $|\Gamma_L| = 1$ (no power is delivered to the load). Then $\Gamma_L = -1; VSWR = +\infty$ $Z_{in} = jZ_0 \tan\beta\ell$, which is seen to be purely imaginary for any length ℓ , and to take on all values between $+j\infty$ and $-j\infty$.

- Open-circuit $Z_L = \infty$

When we have an open-circuit as a load, we will have total reflection $|\Gamma_L| = 1$ (no power is delivered to the load). Then $\Gamma_L = 1; VSWR = +\infty$ $Z_{in} = -jZ_0 \tan\beta\ell$ which is seen to be purely imaginary for any length ℓ , and to take on all values between $+j\infty$ and $-j\infty$ Special length $\ell = \lambda / 2$

For this special length, the electrical length of the line is

$$\beta\ell = \frac{2\pi}{\lambda}\ell = \pi$$

We obtain $Z_{in} = Z_L$, meaning that a half–wavelength line (or any multiple of $\lambda / 2$) does not alter or transform the load impedance, regardless of its characteristic impedance. $\ell = \lambda / 2$ is the periodicity of the impedance on a transmission line.

- Special length $\ell = \lambda / 4$

For this special length, the electrical length of the line is

$$\beta\ell = \frac{2\pi}{\lambda}\ell = \frac{\pi}{2}$$

If the line is a quarter–wavelength long, we obtain $Z_{in} = \dfrac{Z_0^2}{Z_L}$

Such a line is known as a quarter–wave transformer because it has the effect of transforming the load impedance in an inverse manner, depending on the characteristic impedance of the line. For example, if $Z_L = +\infty$, we obtain $Z_{in} = 0$; and if $Z_L = 0$, we obtain $Z_{in} = \infty$. The quarter–wave transformer has many important practical applications in circuit design and real impedance matching.

Return loss and power

For the lossless transmission line of figure 1, the average power flow is constant at any point on the line and the total power delivered to the load is equal to the difference between the incident power and the reflected power. The average incident wave power, which represent the available power from the generator, is

$$P_{inc} = \frac{1}{2}\frac{\left|V_0^+\right|^2}{Z_0}$$

And the average reflected wave power is

$$P_{ref} = \frac{1}{2}\frac{\left|V_0^-\right|^2}{Z_0} = \frac{1}{2}\frac{\left|V_0^+\right|^2}{Z_0}\left|\Gamma_L\right|^2 = \left|\Gamma_L\right|^2 P_{inc}$$

The average power flow on the line is

$$P_{av} = \frac{1}{2}\mathcal{R}e\left\{V(z)I^*(z)\right\} = \frac{1}{2}\frac{\left|V_0^+\right|^2}{Z_0}\left(1-\left|\Gamma_L\right|^2\right)$$

Then the total average power delivered to the load is

$$P_L = P_{av} = P_{inc} - P_{ref} = \frac{1}{2}\frac{\left|V_0^+\right|^2}{Z_0}\left(1-\left|\Gamma_L\right|^2\right)$$

Note that

$$\frac{P_{ref}}{P_{inc}} = \left|\Gamma_L\right|^2$$

This quantity is measured in practice in dB and is called return loss (RL)

$$RL\left[\mathrm{dB}\right] = -10\log\left|\Gamma_L\right|^2 = -20\log\left|\Gamma_L\right|$$

This means that when the load is mismatched, not all of the available power from the generator is delivered to the load.

Maximum power transfer

If we have

$$Z_{in} = Z_g^* \leftrightarrow R_{in} + jX_{in} = R_g - jX_g \leftrightarrow R_{in} = R_g \ ; X_{in} = -X_g$$

This condition is known as conjugate matching, and it results in maximum power transfer to the load for a fixed generator impedance. The power delivered is maximum and equal to:

$$P_{max} = \frac{|V_g|^2}{8R_g}$$

This is also the maximum available power from the generator.

Scattering parameters

Usually we use Y,Z, H or ABCD parameters to describe a linear two port network. These parameters require us to open or short a network to find the parameters. At RF and microwave frequencies it is difficult to have a proper short or open circuit. Open/short condition leads to standing wave, this can cause oscillation and destruction of device. For non-TEM propagation mode, it is not possible to define or even measure voltage and current, We can only measure power from E and H fields. At microwave frequencies the measurement of voltage or current is almost impossible, especially for non-TEM propagation mode. In microwave laboratory, it is possible to measure incident and reflected power waves using vector network analyzer.

Hence a new set of parameters (S parameters) is needed which:

- Do not need open/short condition.
- Do not cause standing wave.
- Relates to incident and reflected power waves, instead of total voltage and current.

S-parameters have the following advantages:

- Provide a complete description of the network as seen at its ports.
- Can be calculated using network analysis techniques, or measured directly with a vector network analyzer.
- Relates to familiar measurement such as reflection coefficient, gain, loss etc.
- Can compute Z, Y, ABCD or H parameters from S-parameters.
- In accord with direct measurements, and with the ideas of incident, reflected, and transmitted waves.

Scattering matrix

Consider the N-port microwave network shown in Figure 3, where V_n^+ is the amplitude of the voltage wave incident on port n and V_n^- is the amplitude of the voltage wave reflected from port n. The ports in Figure 3 may be any type of transmission line or transmission line equivalent of a single propagating waveguide mode. If one of the physical ports of the network is a waveguide supporting more than one propagating mode, additional electrical ports can be added to account for these modes.

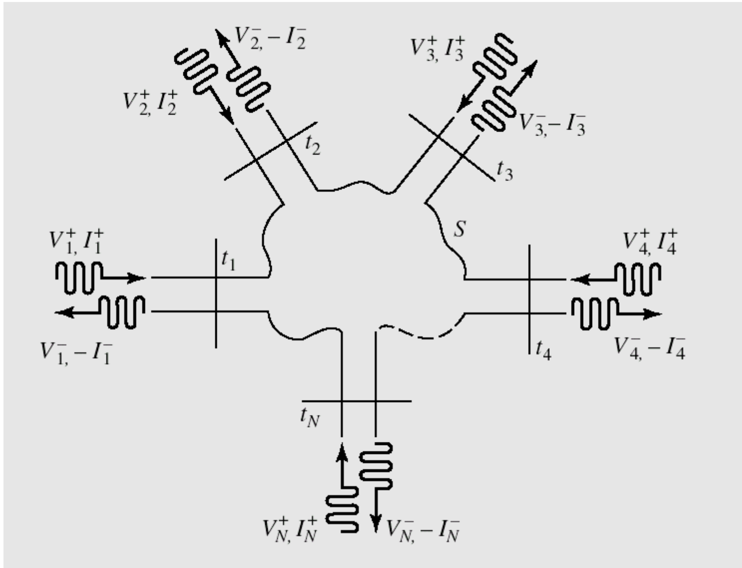


Figure 3: N−port microwave network.

At a specific point on the n^{th} port, a reference plane t_n, is defined along with voltages and currents for the incident and reflected waves. The reference planes are important in providing a phase reference for the voltage and current phasors. Now, at the n^{th} reference plane, the total voltage and current are given by:

$$V_n = V_n^+ + V_n^-$$

$$I_n = I_n^+ - I_n^-$$

The scattering matrix, or $[S]$ matrix, is defined in relation to these incident and reflected voltage waves as

$$\begin{bmatrix} V_1^- \\ V_2^- \\ \vdots \\ V_N^- \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1N} \\ S_{21} & S_{22} & \cdots & S_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ S_{N1} & S_{N2} & \cdots & S_{NN} \end{bmatrix} \cdot \begin{bmatrix} V_1^+ \\ V_2^+ \\ \vdots \\ V_N^+ \end{bmatrix}$$

Or in matrix form as

$$\left[V^{-}\right]=[S]\bullet\left[V^{+}\right]$$

A specific element of the scattering matrix can be determined as

$$S_{ij}=\left.\frac{V_i^{-}}{V_j^{+}}\right|_{V_k^{+}=0\ \mathrm{for}\ k\neq j}$$

Thus, S_{ij} is the reflection coefficient seen looking into port i when all other ports are terminated in matched loads $(V_k^{+}=0 fork\neq j)$, and S_{ij} is the transmission coefficient from port j to port i when all other ports are terminated in matched loads. Which means that all ports should be terminated in matched loads to avoid reflections. The matched load is Z_0, where all ports have the same characteristic impedance Z_0

S parameter properties

- S parameters of a network are properties only of the network itself, and are defined under the condition that all ports are matched.
- Changing the terminations or excitations of a network does not change its S parameters, but may change the reflection coefficient seen at a given port, or the transmission coefficient between two ports.
- Each S parameter, in general, is a complex quantity having a magnitude and phase. Thus, it has the form $S_{ij}=|S_i j|e^{j\theta_i j}$
- Degrees of freedom of N−port network is $2N^2$, since $[S]$ is $N\times N$ and S parameters are complex quantities.
- The network is reciprocal if $[S]$ matrix is symmetric, i.e. $S_{ij}=S_{ji}$

- The network is lossless if $[S]$ matrix is unitary:

1. The dot product of any column of $[S]$ with the conjugate of that column gives unity:

$$\sum_{k=1}^{N} S_{ki} S_{ki}^{*} = \sum_{k=1}^{N} \left| S_{ki} \right|^{2} = 1$$

2. While the dot product of any column of $[S]$ with the conjugate of a different column gives zero (orthogonal):

$$\sum_{k=1}^{N} S_{ki} S_{kj}^{*} = 0, \text{ for } i \neq j$$

A Shift in Reference Planes

Consider the N−port microwave network shown in Figure 4, where the original reference planes are assumed to be located at $z_n = 0$ or the n^{th} port, where z_n is an arbitrary coordinate measured along the transmission line feeding the n^{th} port. The scattering matrix for the network with this set of reference planes is denoted by $[S]$ Now consider a new set of reference planes defined at $z_n = \ell_n$ for the n^{th} port, and let the new scattering matrix be denoted as $[S']$. Then in terms of the incident and reflected port voltages we have that

$$\left[V^{-}\right] = [S] \cdot \left[V^{+}\right]$$

$$\left[V'^{-}\right] = [S'] \cdot \left[V'^{+}\right]$$

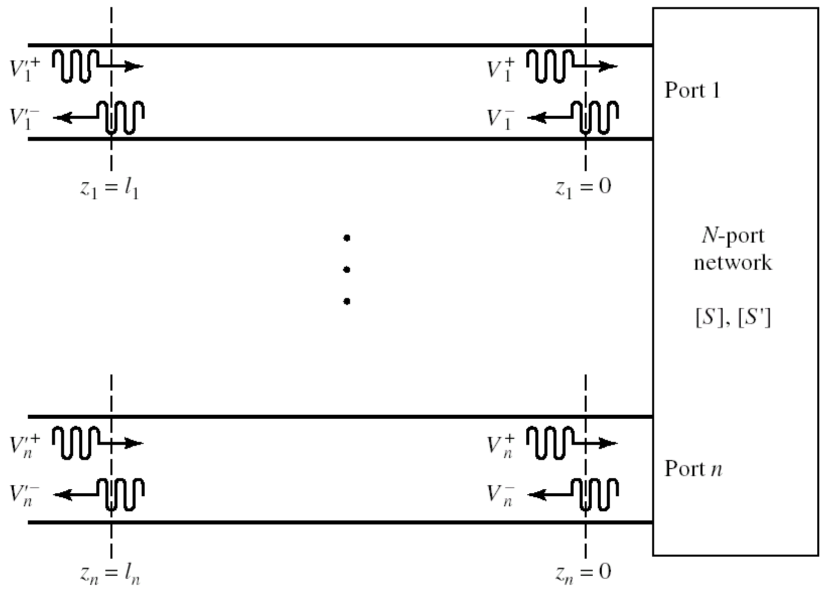


Figure 4: Shifting reference planes for an N−port microwave network.

From the theory of traveling waves on lossless transmission lines we can relate the new wave amplitudes to the original ones as

$$V'^+_n = V_n^+ e^{j\theta_n}$$

$$V'^-_n = V_n^- e^{-j\theta_n}$$

Thus we have

$$S'_{ij} = \frac{V'^-_i}{V'^+_j} = \frac{V_i^- e^{-j\theta_i}}{V_j^+ e^{j\theta_j}} = S_{ij}\, e^{-j\left(\theta_i + \theta_j\right)}$$

$$S'_{ii} = \frac{V'^-_i}{V'^+_i} = \frac{V_i^- e^{-j\theta_i}}{V_i^+ e^{j\theta_i}} = S_{ii}\, e^{-2j\theta_i}$$

Power Waves and Generalized Scattering Parameters

We previously studied the S parameters of N−port network with all ports are lossless and have the same characteristic impedance Z_0. In this case the average power delivered to a load can be expressed as:

$$P_L = P_{inc} - P_{ref} = \frac{1}{2Z_0}\left(\left|V_0^+\right|^2 - \left|V_0^-\right|^2\right)$$

Unfortunately, this result is only valid when the characteristic impedance is real, and the S parameters defined previously are valid only if all ports are lossless and have the same characteristic impedance Z_0.

To have a generalization of scattering parameters, where the characteristic impedances are different but still real, we define a new set of waves, called power waves, which have useful properties when dealing with power transfer between a generator and a load. For a characteristic impedance Z_{0n} of port n, we define the incident and reflected power waves as

$$a_n = \frac{V_n^+}{\sqrt{Z_{0n}}} \; ; \; b_n = \frac{V_n^-}{\sqrt{Z_{0n}}}$$

Then the power delivered to the load at port n can be expressed as

$$P_n = \frac{1}{2}\left|a_n\right|^2 - \frac{1}{2}\left|b_n\right|^2$$

Where $\frac{1}{2}\left|a_n\right|^2$ is the average power of the incident wave, and $\frac{1}{2}\left|b_n\right|^2$ is the average power of the reflected wave as they are measured using vector network analyzer in laboratory. Then we define the $[S]$ matrix in this case as

$$[b] = [S] \cdot [a]$$

Where a specific element of the scattering matrix can be determined as

$$S_{ij} = \left.\frac{b_i}{a_j}\right|_{a_k=0 \text{ for } k \neq j} = \left.\frac{V_i^-}{V_j^+}\sqrt{\frac{Z_{0j}}{Z_{0i}}}\right|_{V_k^+=0 \text{ for } k \neq j}$$

Problems with solutions

Problem 1

1. Find the characteristic impedance Z_1 of a lossless transmission line of length $\ell_t = \frac{\lambda}{4}$, terminated in a load $R_L = 100\Omega$, to obtain an input impedance $Z_{in} = 50\Omega$.

The transmission line of length $\ell_t = \frac{\lambda}{4}$ is a quarter−wave transformer, then:

$$Z_{in} = \frac{Z_1^{\ 2}}{Z_L} \Rightarrow Z_1 = \sqrt{Z_L \cdot Z_{in}} = 70.7\Omega$$

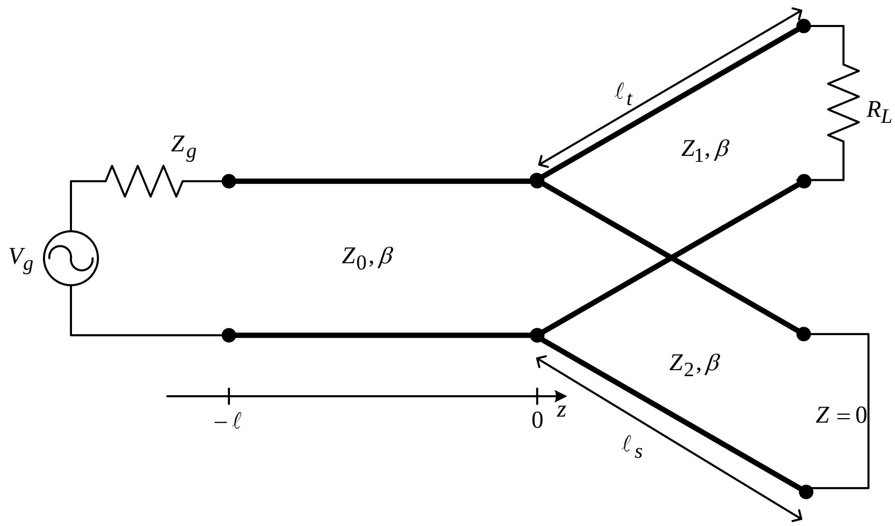
2. Find the length ℓ_s of a lossless transmission line terminated in a short circuit, of characteristic impedance $Z_2 = 50\Omega$, to obtain an input impedance $Z_{in} = j100\Omega$.

The transmission line of length ℓ_s is terminated in a short circuit, then:

$$Z_{in} = jZ_2 \tan\beta\ell_s \Rightarrow \ell_s \frac{\lambda}{2\pi} a\ \tan(2) = 0.176\lambda$$

The two transmission lines of lengths ℓ_t and ℓ_s are connected in parallel to a lossless transmission line of characteristic impedance $Z_0 = 50\Omega$ and length $\ell = \frac{3\lambda}{8}$

. This transmission line is connected to a generator of, $Vg = 10\text{V}, Z_g = 50\Omega$ as shown below.



For the transmission line of characteristic impedance $Z_0 = 50\Omega$ and length $\ell = \frac{3\lambda}{8}$, find:

- (give Z in Real and Imaginary part form, and Γ in magnitude and phase form)
- give V and I in magnitude and phase form; Powers in mW and dBm

3. Z_L, the load impedance seen by the transmission line.

$$Z_L = 50\Omega // j100\Omega = 40 + 20\Omega$$

4. The input impedance Z_{in}.

$$\boldsymbol{Z_{in}} = \boldsymbol{Z}_0 \frac{\boldsymbol{Z_L} + \boldsymbol{j}\ \boldsymbol{Z}_0 \tan\beta\ell}{\boldsymbol{Z}_0 + \boldsymbol{j}\ \boldsymbol{Z_L} \tan\beta\ell} \quad ; \quad \beta\ell = \frac{3\pi}{4}, \quad \tan\beta\ell = -1$$

$$Z_{in} = 50\frac{40+j\,20-j\,50}{50-j\;(40+j\,20)} = 50\frac{40-j\,30}{70-j\,40} = 50(0.615-j\,0.077) = 30.8-j\,3.8\Omega$$

5. The reflection coefficients: Γ_L at the load, and Γ_{in} at the input.

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} = \frac{-10+j\,20}{90+j\,20} = 0.24\angle 104°$$

$$\Gamma_{in} = \Gamma_L\, e^{-2j\beta\ell} = |\Gamma_L|\; e^{j\,(104°-270°)} = 0.24\angle -166°$$

6. The return loss RL and $VSWR$ on the line.

$$RL = -20\log(|\Gamma_L|) = 12.3\text{dB}$$

$$VSWR = \frac{1+|\Gamma_L|}{1-|\Gamma_L|} = 1.64$$

7. The distance $d_{\max}$ from the load to the first voltage maximum.

The voltage standing wave on the line has a maximum when $e^{j(\theta-2\beta\ell)} = 1$, where θ is the angle of reflection coefficient, then

$$d_{\max} = \frac{\theta}{2\beta} = 0.144\lambda$$

For the transmission line connected to the generator and terminated in Z_L :

1. Find V_{in} and I_{in} at $z = -\ell$.

$$V_{in} = V_g\,\frac{Z_{in}}{Z_{in}+Z_g} = 3.8\ \text{V}\angle 4.4°$$

$$I_{in} = \frac{V_{in}}{Z_{in}} = \frac{V_g}{Z_{in}+Z_g} = 124\ \text{mA}\angle 2.7°$$

2. Write $V(z)$ and $I(z)$ along the line. Find V_0^+, then find V_L, I_L at $z=0$.

$$V(z)=V_0^+\left(e^{-j\beta z}+\Gamma_L\, e^{j\beta z}\right)$$

$$I(z)=\frac{V_0^+}{Z_0}\left(e^{-j\beta z}-\Gamma_L\, e^{j\beta z}\right)$$

$$V_0^+=V_g\,\frac{Z_0}{Z_0+Z_g}\,\frac{e^{-j\beta\ell}}{1-\Gamma_L\Gamma_g\, e^{-2j\beta\ell}}$$

$$Z_0=Z_g \quad\Rightarrow\quad \Gamma_g=0$$

$$V_0^+=\frac{V_g}{2}e^{-j\beta\ell}=5\,\text{V}\angle-135°$$

$$V_L=V(z=0)=V_0^+\left(1+\Gamma_L\right)=4.85\ \text{V}_{\text{rms}}\angle-121°$$
$$I_L=I(z=0)=\frac{V(0)}{Z_L}=\frac{V_L}{Z_L}=108\ \text{mA}_{\text{rms}}\angle-148°$$

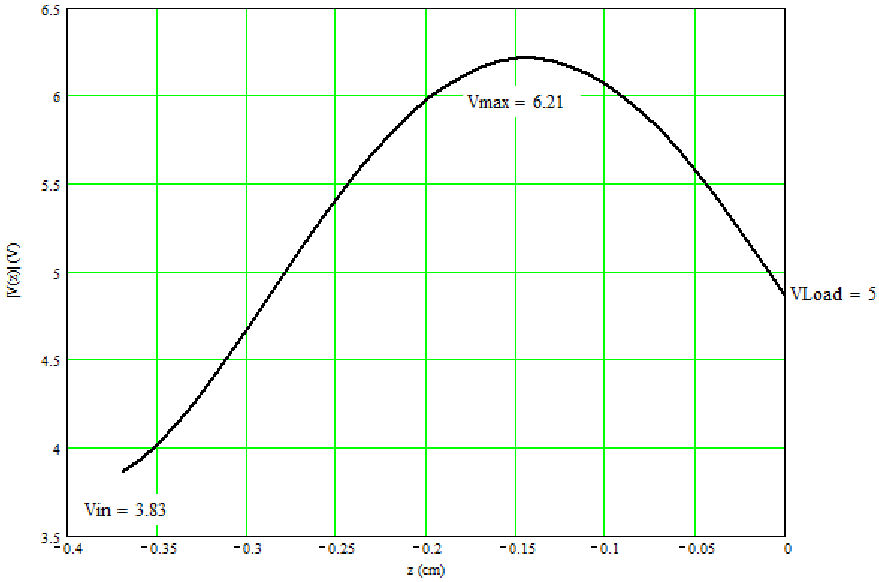
3. Compute the values of voltage maximum and minimum on the line.

$$V_{\max}=\left|V_0^+\right|\left(1+\left|\Gamma_L\right|\right)=6.2\,\text{V}$$

$$\left|\text{VL}\right|=4.851$$

$$V_{\min}=\left|V_0^+\right|\left(1-\left|\Gamma_L\right|\right)=3.8\,\text{V}$$

4. Draw the magnitude of the standing wave $\left|V(z)\right|$ along the line.



5. Compute the following powers: P_{in} delivered to the network, P_L delivered to Z_L, P_R delivered to R_L, P_s delivered by the source, the incident power P_{inc}, and the reflected power P_{ref}. Discuss the results.

$$P_{in} = \tfrac{1}{2}\Re e\left\{ V_{in} \cdot I_{in}^{*} \right\} = \tfrac{1}{2}\Re e\left\{ Z_{in} I_{in} \cdot (I_{in})^{*} \right\} = \tfrac{1}{2}\left|I_{in}\right|^{2} \Re e\left\{ Z_{in} \right\} = 237\,\text{mW} = 23.75\ \text{dBm}$$

The transmission line is lossless, thus, the power delivered to its load is:

$$P_L = P_{in}$$

Since the transmission line of length ℓ_s is short–circuited, then all the power P_L is delivered to R_L :

$$P_R = P_L$$

The power delivered by the source is:

$$P_s = \tfrac{1}{2}\Re e\left\{ V_g \cdot I_g^{*} \right\} = \tfrac{1}{2}\Re e\left\{ \left(Z_{in} + Z_g \right) I_{in} \cdot (I_{in})^{*} \right\} = \tfrac{1}{2}\left|I_{in}\right|^{2} \Re e\left\{ Z_{in} + Z_g \right\} = 618\,\text{mW} = 27.9\,\text{dBm}$$

It is clear that

$$P_s = P_{in} + P_{Z_g}$$

Where $P_{Z_g} = 382\text{mW}$ is the power dissipated in Z_g .

The incident power is:

$$P_{inc} = \frac{\left|V_0^+\right|^2}{2Z_0} = 250 \text{ mW} = 24 \text{ dBm}$$

And the reflected power is:

$$P_{ref} = \frac{\left|V_0^+\right|^2}{2Z_0} \cdot \left|\Gamma_L\right|^2 = P_{inc} \cdot \left|\Gamma_L\right|^2 = 14.7 \text{ mW} = 11.7 \text{ dBm}$$

Discussion: the calculated powers are related by the following relationships:

$$P_{in} = P_L = P_R = P_{inc} - P_{ref} = P_s - P_{Z_g}$$

Problem 2

Consider a lossless transmission line terminated in a load impedance Z_L . If the magnitude of the voltage standing wave has: a voltage maximum of $7.75 V_{rms}$ and a voltage minimum of $2.25 V_{rms}$; a distance between two successive minima of 3.1cm; a distance of the first voltage minimum from the load of 0.88cm, $Z_0 = 100\Omega$, and $\ell = 5.1\text{cm}$,

Use Smith chart to find:

(give Z and Y in Real and Imaginary part form, and Γ in magnitude and phase form)

1. The load impedance Z_L, the load admittance Y_L, the input impedance Z_{in}, and the input admittance Y_{in}.

The distance between two successive minima is $\frac{\lambda}{2} = 3.1\text{cm} \Rightarrow \lambda = 6.2\text{cm}$.

The physical length of the line is $\ell = 5.1\text{cm} = 0.823\lambda$.

The distance of first voltage minimum from load is $\ell_{\max} = 0.88\text{cm} = 0.142\lambda$.

$$VSWR = \frac{V_{\max}}{V_{\min}} = 3.44$$

Then from Smith chart we find:

$$Z_L = z_L \cdot Z_0 = 100(0.65 - j) = 65 - j100 \ \Omega$$
$$Y_L = \frac{y_L}{Z_0} = \frac{0.46 + j\ 0.7}{100} = 4.6 + j\ 7\ mS$$
$$Z_{in} = z_{in} \cdot Z_0 = (1.2 + j1.4) \times 100 = 120 + j\ 140\ \Omega$$
$$Y_{in} = \frac{y_{in}}{Z_0} = \frac{0.35 - j\ 0.41}{100} = 3.5 - j\ 4.1\ \text{mS}$$

2. The reflection coefficients: Γ_L at the load, and Γ_{in} at the input. The return loss RL.

$$\Gamma_L = 0.55\angle -78^\circ$$
$$\Gamma_{in} = 0.55\angle 50^\circ$$
$$RL = 5.2\ \text{dB}$$

3. The distance d_{in} from the input to obtain $Z = (30 - j\ 11)\ \Omega$.

$$d_{in} = (0.5\lambda - 0.319\lambda) + 0.02\lambda = 0.2\lambda = 1.24\ \text{cm}$$

4. The distances $d_{\min}$ and $d_{\max}$ from the input to the voltage minima and maxima.

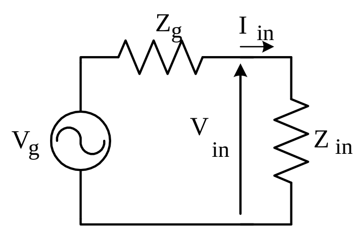
$$d_{\min 1} = 0.5\lambda - 0.319\lambda = 0.181\lambda = 1.12 \text{ cm}$$
$$d_{\min 2} = d_{\min 1} + 0.5\lambda = 0.681\lambda = 4.22 \text{ cm}$$
$$d_{\max 1} = d_{\min 1} + 0.25\lambda = 0.431\lambda = 2.67 \text{ cm}$$

A generator is connected to the transmission line with $V_g = 10\text{V}_{\text{rms}}$; $Z_g = 100\ \Omega$ Find:

(give V and I in magnitude and phase form; the Power in mW and dBm)

1. V_{in} , I_{in} . Write $V(z)$ and $I(z)$ along the line. Find V_0^+ , then find V_L , I_L

The equivalent circuit seen from the generator is the following:



Then we can write:

$$V_{in} = V_g \frac{Z_{in}}{Z_{in} + Z_g} = 7.1 \text{ V}_{\text{rms}} \angle 17°$$

$$I_{in} = \frac{V_{in}}{Z_{in}} = \frac{V_g}{Z_{in} + Z_g} = 38.3 \text{ mA}_{\text{rms}} \angle -33°$$

$$V(z) = V_0^+ \left(e^{-j\beta z} + \Gamma_L e^{j\beta z}\right)$$

$$I(z) = \frac{V_0^+}{Z_0} \left(e^{-j\beta z} - \Gamma_L e^{j\beta z}\right)$$

$$V_0^+ = V_g \frac{Z_0}{Z_0 + Z_g} \frac{e^{-j\beta\ell}}{1 - \Gamma_L \Gamma_g e^{-2j\beta\ell}}$$

$$Z_0 = Z_g \quad \Rightarrow \quad \Gamma_g = 0$$

$$V_0^+ = \frac{V_g}{2} e^{-j\beta\ell} = 5 \text{ V}_{\text{rms}} \angle -116°$$

$$V_L = V(z=0) = V_0^+ \left(1 + \Gamma_L\right) = 6.2 \text{ V}_{\text{rms}} \angle -142°$$

$$I_L = I(z=0) = \frac{V(0)}{Z_L} = \frac{V_L}{Z_L} = 51.8 \text{ mA}_{\text{rms}} \angle -85°$$

2. The power P_{in} delivered to the network, the power P_L delivered to Z_L, and the power P_s delivered by the source. Discuss the results.

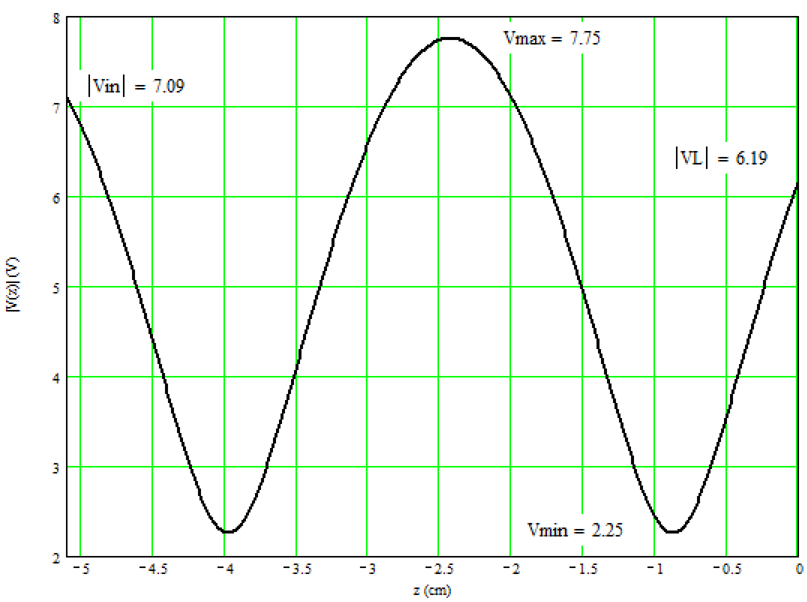
$$P_L = P_{in} = \Re e\left\{ V_L \cdot I_L^{*} \right\} = \left|I_L\right|^2 \Re e\left\{Z_L\right\} = 174.4 \ \text{mW} = 22.4 \ \text{dBm}$$
$$P_s = \Re e\left\{ V_g \cdot I_g^{*} \right\} = \Re e\left\{ V_g \cdot (I_{in})^{*} \right\} = 321.2 \ \text{mW} = 25.1 \ \text{dBm}$$

$P_L = P_{in}$ because the line is lossless. The difference $P_s - P_{in} = P_{R_g}$ is the power dissipated in R_g, the real part of Z_g.

3. P_{inc}, the incident power, and P_{ref}, the reflected power on the line. Refind *RL*.

$$P_{inc} = \frac{\left|V_0^{+}\right|^2}{Z_0} = 250 \ \text{mW} = 24 \ \text{dBm}$$
$$P_{ref} = P_{inc} \cdot \left|\Gamma_L\right|^2 = 75.6 \ \text{mW} = 18.8 \ \text{dBm}$$
$$RL = P_{inc} - P_{ref} = 5.2 \ \text{dBm}$$

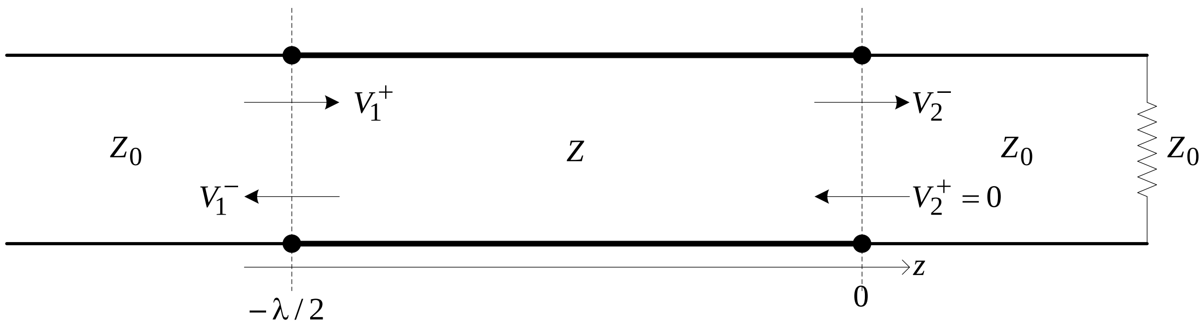
4. Draw the magnitude of the standing wave $\left|V(z)\right|$ along the line.



Problem 3: $[S]$ Matrix

- Derive the scattering matrix $[S]$ for a lossless transmission line of characteristic impedance Z and length $\ell = \lambda/2$, relative to a system impedance Z_0.

We have the following circuit to find S_{11} and S_{21}:



The transmission line is a lossless, symmetric and reciprocal network, thus:

$$S_{11} = S_{22} \qquad (Symmetry)$$
$$S_{21} = S_{12} \qquad (Reciprocity)$$

The length of the transmission line is $\ell = \lambda/2$, thus:

$$Z_{in} = Z_L = Z_0 \quad \Rightarrow \quad S_{11} = 0$$
$$S_{11} = 0 \quad \Rightarrow \quad V_1^- = 0 \quad \Rightarrow \quad V_1 = V_1^+$$
$$V_2^+ = 0 \quad \Rightarrow \quad V_2 = V_2^-$$
$$S_{21} = \left.\frac{V_2^-}{V_1^+}\right|_{V_2^+=0} = \frac{V_2}{V_1}$$
$$V(z) = V_0^+\left(e^{-j\beta z} + \Gamma_L e^{j\beta z}\right)$$
$$V_1 = V(-\lambda/2) = V_0^+(-1-\Gamma_L) = -V_0^+(1+\Gamma_L)$$
$$V_2 = V(0) = V_0^+(1+\Gamma_L)$$

Thus: $S_{21} = -1$

This result is expected and means that the travelling wave V_1^+ at the input of the line and the traveling wave V_2^- in the same direction at the load are out of phase because the length is $\ell = \lambda / 2$.

Then the scattering matrix $[S]$, relative to a system impedance Z_0, for a lossless transmission line with characteristic impedance Z and length $\ell = \lambda / 2$ is:

$$[S] = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

- A three−port network has the following $[S]$ matrix in a system impedance Z_0.

$$[S] = \frac{-j}{\sqrt{2}} \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

  - Is this network matched? lossless? reciprocal?

The network is matched, because $S_{ii} = 0$, and reciprocal because $S_{ij} = S_{ji}$

The network is lossless if $[S]$ is unitary. Taking the first column:

$$|S_{11}|^2 + |S_{21}|^2 + |S_{31}|^2 = 1$$

Taking the second column:

$$|S_{21}|^2 + |S_{22}|^2 + |S_{23}|^2 = 0.5 \neq 1$$

Thus the network is not lossless.

Chapter Two:

RF and Microwave Technologies and Active Devices

Keywords

Monolithic Microwave Integrated Circuits (MMICs) System-on-a-Chip (SoC), Hybrid MICs, bandwidth, octave and decade, narrowband (narrow bandwidth), wideband (wide bandwidth), gallium arsenide (GaAs), Silicon (Si), indium phosphide (InP), or silicon germanium (SiGe), gallium nitride (GaN) gallium aluminum arsenide (GaAlAs), gallium indium arsenide (GaInAs), aluminum gallium nitride (AlGaN), Varactor diodes, Schottky diodes, PIN diodes, field effect transistors (FETs), bipolar junction transistors (BJTs), heterojunction bipolar transistors (HBTs), metal semiconductor field effect transistor (MESFET), metal oxide semiconductor field effect transistor (MOSFET), high electron mobility transistor (HEMT).

Abstract

This chapter is a concise overview of RF and microwave semiconductor devices and technologies including Monolithic Microwave Integrated Circuits (MMICs), system design convergence, system-on-a-chip (SoC), details of the properties and characteristics of RF and microwave semiconductor devices. Devices included are: Varactors, Schottky diodes, PIN diodes, BJTs, HBTs, MOSFETs, MESFETs, and HEMTs.

Learning Objectives

In this chapter, student will be able to:

- Read and understand datasheets of different RF and microwave semiconductor devices and system-on-a-chip.
- Recognize different RF and microwave semiconductor devices and technologies.

Introduction

In this chapter we will briefly overview RF and microwave semiconductor devices and technologies including Monolithic Microwave Integrated Circuits (MMICs), system design convergence, system−on−a−chip (SoC), details of the properties and characteristics of RF and microwave semiconductor devices. Devices included in the text are: Varactors, Schottky diodes, PIN diodes, BJTs, HBTs, MOSFETs, MESFETs, and HEMTs.

This overview help student to well read and understand datasheets of different RF and microwave semiconductor devices and system−on−a−chip (diodes, transistors, amplifiers, oscillators, mixers, synthesizers, RF Front−End receivers (RF−FE), …). Nowadays, the key in wireless system design is well choosing different components, after determining specifications from system−level analysis work.

Monolithic Microwave Integrated Circuits

Following the lead from integrated circuits (IC) at lower frequencies, Monolithic Microwave Integrated Circuits (MMICs) combine transmission lines, active devices (diodes and transistors), and other components on a semiconductor substrate. The first single−function MMICs were developed in the late 1960s, but more sophisticated circuits and subsystems, such as multistage amplifiers, transmit/receive radar modules, front ends for wireless products, and many other circuits, are now being fabricated as MMICs.

The trend of any maturing electronic technology is toward smaller size, lighter weight, lower power requirements, lower cost, and increased complexity. Microwave technology has been moving in this direction for the last 10–30 years with the development of microwave integrated circuits (MICs). This technology strives to replace bulky and expensive waveguide and coaxial components with small and inexpensive planar components, and is analogous to the digital integrated circuitry that has led to the rapid increase in sophistication of computer systems.

Microwave integrated circuitry can incorporate transmission lines, discrete resistors, capacitors, and inductors, as well as active devices such as diodes and transistors. MIC technology has advanced to the point where complete microwave subsystems, such as RF−FE receiver and radar transmit/receive modules, can be integrated on a chip that is only a few square millimeters in size.

There are two distinct types of microwave integrated circuits. Hybrid MICs have one layer of metallization for conductors and transmission lines, with discrete components (resistors, capacitors, integrated circuit chips, transistors, diodes, etc.) bonded to the substrate. In a thin−film hybrid MIC, some of the simpler components are deposited on the substrate. Hybrid MICs were first developed in the 1960s, and still provide a very flexible and cost−effective means for circuit implementation. Monolithic microwave integrated circuits (MMICs) are a more recent development, where the active and passive circuit elements are grown on the substrate. The substrate is a semiconductor material, and several layers of metal, dielectric, and resistive films are used.

System Design Convergence

Convergence in the wireless worlds of communications, computing, global positioning systems (GPS), and consumer electronic devices is an irresistible trend today. The growing complexity of today's wireless products results largely from the rapid convergence of traditionally separate technologies. One inevitable consequence of the convergence is the impending integration of voice, data, image, video, music, the Internet, instant messaging, home automation, and GPS.

The design task in this converging world is tough to do because of the merging of once−distinct systems. The best way to overcome the challenges caused by the converging technologies is to successfully incorporate RF analog−digital, hardware−software, system−on−a−chip (SoC), and printed−circuit−board (PCB) designs. Increased complexity and the shrinking size of wireless products have resulted in designing and using highly integrated SoCs to reduce number of discrete devices on the PCBs as possible.

Definitions

Bandwidth

Probably no word is used more often in the world of RF than the word bandwidth. Bandwidth is a way of describing a range of frequencies. It equals the difference between the highest frequency and the lowest frequency of the device or application, and therefore two frequencies are required to define a bandwidth. For instance, if a particular device can accommodate all frequencies between 75 MHz and 125 MHz, it has a bandwidth of 50 MHz (125 MHz−75 MHz).

The bandwidth of a signal provides a measure of the extent of the significant spectral content (frequencies) of the signal for positive frequencies. When the signal is strictly band limited, the bandwidth is well defined. For example, the sinc pulse in Figure 1 has a bandwidth equal to W. However, when the signal is not strictly band limited, which is generally the case, we encounter difficulty in defining the bandwidth of the signal. The difficulty arises because the meaning of the word "significant" attached to the spectral content of the signal is mathematically imprecise. Consequently, there is no universally accepted definition of bandwidth.

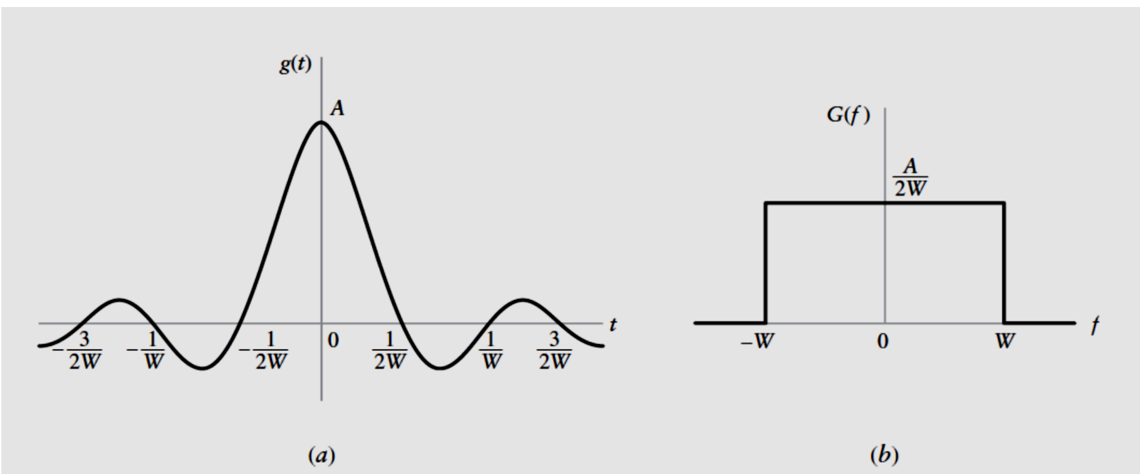


Figure 1. (a) Sinc pulse $g(t)$ (b) Fourier transform $G(f)$.

Nevertheless, there are some commonly used definitions for bandwidth. The formulation of each definition depends on whether the signal is low–pass or band–pass. A signal is said to be low–pass if its significant spectral content is centered around the origin $f = 0$ (DC component of the signal). A signal is said to be band–pass if its significant spectral content is centered around f_c where f_c is a constant frequency. Figure 2 shows a spectrum of a low–pass and a band–pass signal.

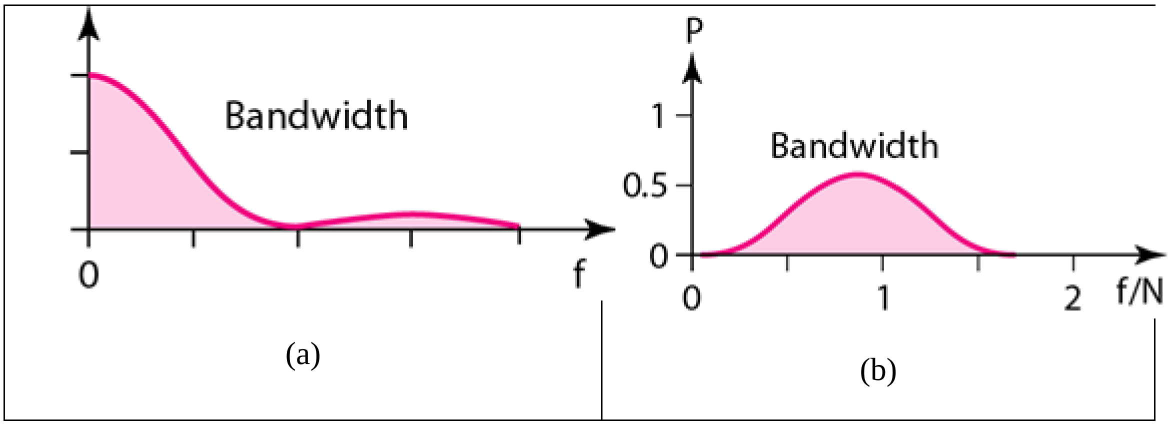


Figure 2. (a) Spectrum of a low–pass signal. (b) Spectrum of a band–pass signal.

When the spectrum of a signal is symmetric with a main lobe bounded by well–defined nulls (i.e., frequencies at which the spectrum is zero), we may use the main lobe as the basis for defining the bandwidth of the signal. The rationale for doing so is that the main spectral lobe contains the significant portion of the signal energy. If the signal is low–pass, the bandwidth is defined as one half the total width of the main spectral lobe, since only one half of this lobe lies inside the positive frequency region. For example, a rectangular pulse of duration T seconds has a main spectral lobe of total width $(2/T)$ hertz centered at the origin, as depicted in Figure 3(b).

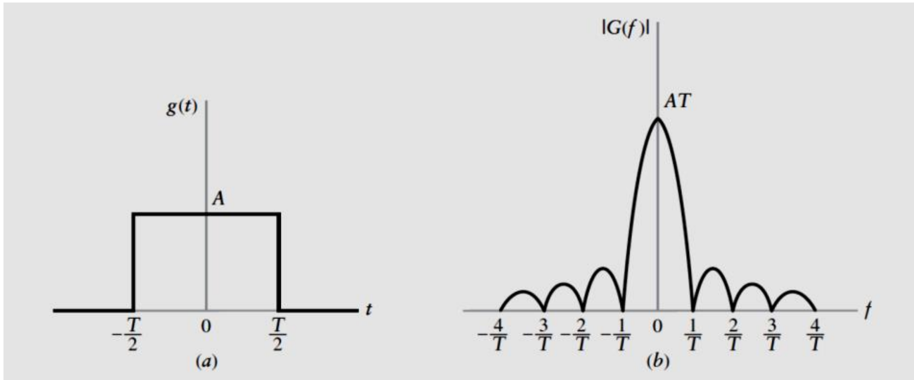


Figure 3: (a) Rectangular pulse. (b) Amplitude spectrum.

Accordingly, we may define the bandwidth of this rectangular pulse as $(1/T)$ hertz. If, on the other hand, the signal is band−pass with main spectral lobes centered around f_c, where f_c is large, the bandwidth is defined as the width of the main lobe for positive frequencies. This definition of bandwidth is called the null−to−null bandwidth. For example, an RF pulse of duration T seconds and frequency f_c has main spectral lobes of width $(2/T)$ hertz centered around $\pm f_c$ as depicted in Figure 4.

Hence, we may define the null−to−null bandwidth of this RF pulse as $(2/T)$ hertz. On the basis of the definitions presented here, we may state that shifting the spectral content of a low−pass signal by a sufficiently large frequency has the effect of doubling the bandwidth of the signal. Such a frequency translation is attained by using the process of modulation, which was discussed in detail in the course “Analog Communications”.

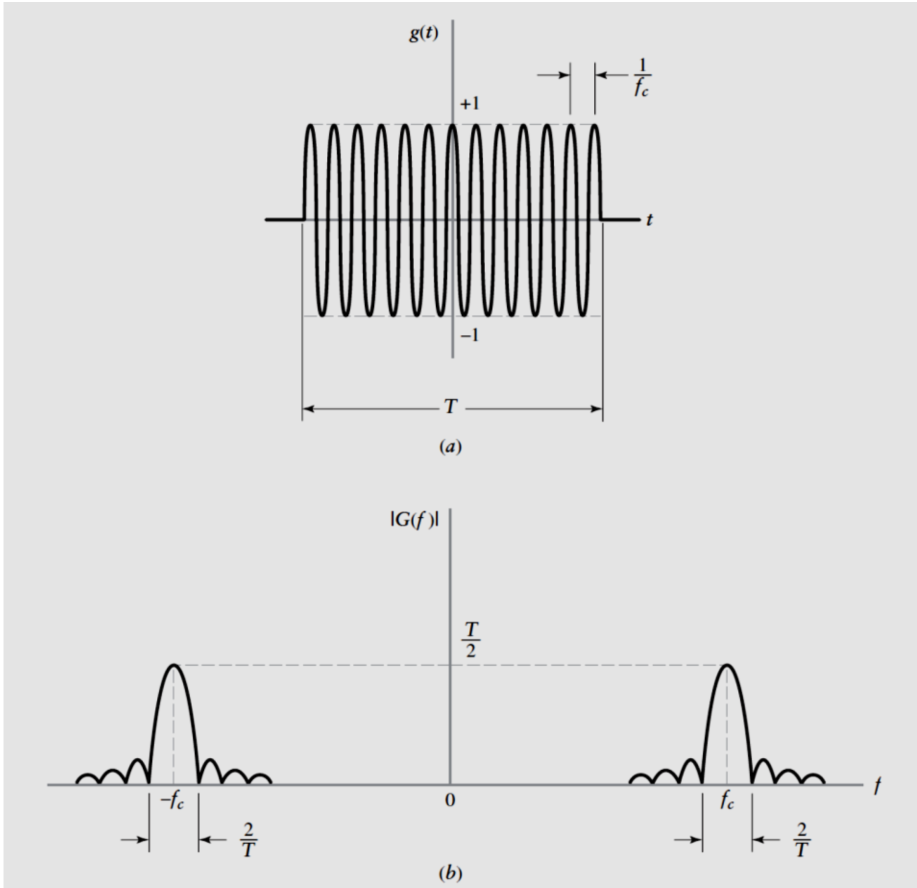


Figure 4: (a) RF pulse of unit amplitude and duration T. (b) Amplitude spectrum.

Another popular definition of bandwidth is the **3−dB bandwidth**. Specifically, if the signal is low−pass, the 3−dB bandwidth is defined as the separation between zero frequency, where the amplitude spectrum attains its peak value, and the positive frequency at which the amplitude spectrum drops to $1/\sqrt{2}$ of its peak value. If, on the other hand, the signal is band−pass, centered at $\pm f_c$, the 3−dB bandwidth is defined as the separation (along the positive frequency axis) between the two frequencies at which the amplitude spectrum of the signal drops to $1/\sqrt{2}$ of the peak value at f_c. The 3−dB bandwidth has an advantage in that it can be read directly from a plot of the amplitude spectrum. However, it has a disadvantage in that it may be misleading if the amplitude spectrum has slowly decreasing tails.

Octaves and Decades

There are two other important descriptors of bandwidth: octave and decade. Octave and decade come from the world of logarithms in which octave means twice as big and decade means ten times as big. If the upper frequency of a device is twice as big as the lower frequency, then the device has an octave bandwidth. For instance, a device which operates from 100 MHz to 200 MHz has an octave bandwidth. The same is true for a device whose lower frequency is 1.2 GHz and upper frequency is 2.4 GHz. To really confuse you, if the lower frequency of a device is 100 MHz and the upper frequency is 300 MHz, the device has a two octave bandwidth. Any device which has a bandwidth greater than one octave is said to have a multioctave bandwidth (as if it could be anything else).

If the upper frequency of a device is ten times the lower frequency, then the device has a decade bandwidth. An example of a decade bandwidth is a device which operates from 100 MHz to 1000 MHz (1 GHz).

Sometimes bandwidth is expressed in terms of a percentage. In this case, the bandwidth is simply divided by the average of the upper and lower frequencies. A simple example will explain everything.

Example

Calculating percentage bandwidth.

If a device can accommodate all frequencies between 75 MHz and 125 MHz, what is its percentage bandwidth?

First, you calculate the actual bandwidth. As noted above, the bandwidth for this example is

$$125\ \text{MHz} - 75\ \text{MHz} = 50\,\text{MHz}$$

Next, you calculate the average of the two frequencies. In this case it is

$$\frac{125\ \text{MHz} + 75\ \text{MHz}}{2} = 100\ \text{MHz}$$

Finally, you divide the bandwidth by the average frequency and multiply by 100%.

$$\frac{50\,\text{MHz}}{100\ \text{MHz}} \times 100\% = 50\%$$

A device which operates from 75 MHz to 125 MHz has a 50% bandwidth.

Wideband And Narrowband

Why is all this important? It is important because all RF components are classified as either narrowband (meaning narrow bandwidth) or wideband (meaning wide bandwidth). There are no hard and fast rules for the separation between narrowband and wideband, so I will make one up. If the bandwidth of a component is less than 50%, it is narrowband. If it is greater than 50%, it is wideband. Ok, so where is all this leading? Here is the key: The wider the bandwidth of a component, the more frequencies it can accommodate, but the more it costs and the worse it performs. For instance, a narrowband passive component might have 1 dB of insertion loss (good), where an identical wideband passive component might have 3 or 4 dB of insertion loss (bad). The trick in designing an RF circuit is to get away with the narrowest bandwidth device possible while still accommodating all the frequencies required. As an example, cellular phone conversations -from the phone to the base station- cover the frequency range 824-849 MHz. The most intelligent designs cover just this frequency range and no more. An interesting thing to note is that narrowband and wideband devices are manufactured entirely different, which is why in the RF industry there are companies who specialize in either narrowband or wideband products.

Diodes and applications

Diodes are active devices used in RF and microwave circuits with different types. A diode is a two-terminal semiconductor device having a nonlinear V–I relationship as shown in Figure 5. This nonlinearity can be exploited for the useful functions of signal detection, demodulation, switching, frequency multiplication, and oscillation. RF and microwave diodes can be packaged as axial or beam lead components or as surface-mountable chips, or be monolithically integrated with other components on a single semiconductor substrate. We first consider detector diodes and circuits, then discuss PIN diodes and control circuits, varactor diodes, and a summary of other types of diodes.

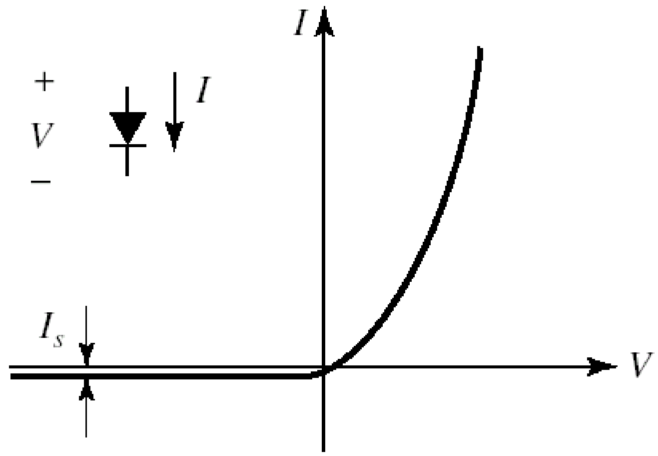


Figure 5: V–I characteristics of a Schottky diode.

Schottky Diodes and Detectors

The classical p–n junction diode commonly used at low frequencies has a relatively large junction capacitance that makes it unsuitable for high frequency application. The Schottky barrier diode, however, relies on a semiconductor–metal junction that results in a much lower junction capacitance, allowing operation at higher frequencies. Commercially available microwave Schottky diodes generally use n–type gallium arsenide (GaAs) material, while lower frequency versions may use n–type silicon. Schottky diodes are often biased with a small DC forward current, but can be used without bias.

The primary application of Schottky diodes is in frequency conversion of an input signal. Figure 6 illustrates the three basic frequency conversion operations of rectification (conversion to DC), detection (demodulation of an amplitude–modulated signal), and mixing (frequency translation/conversion).

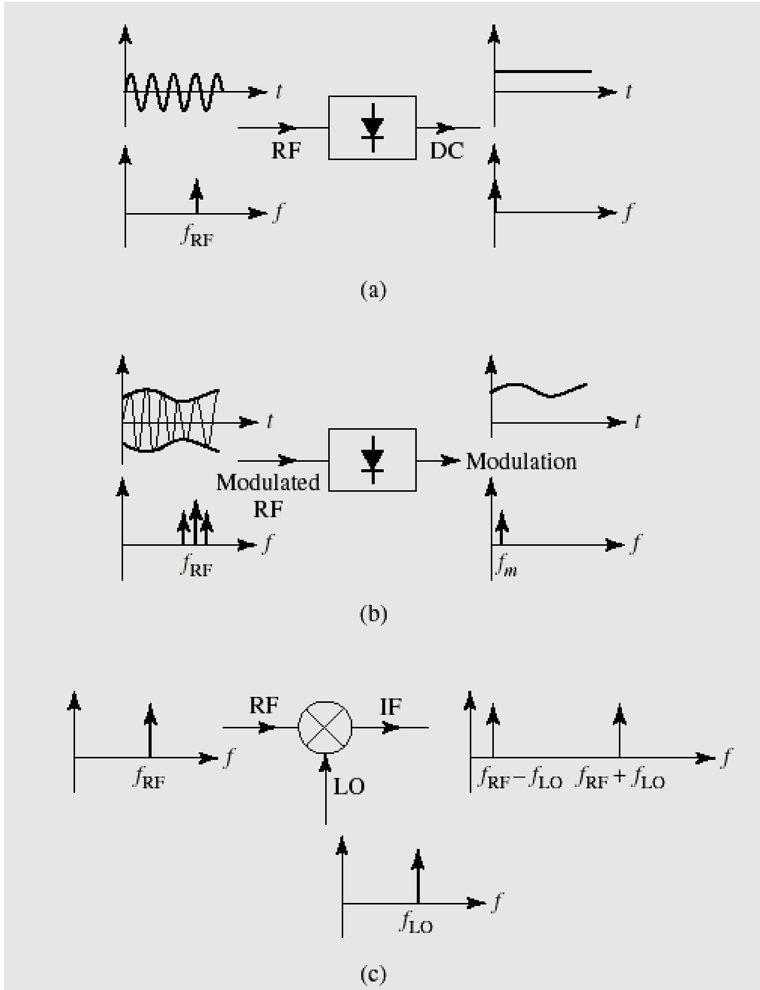


Figure 6: Basic frequency conversion operations of rectification, detection, and mixing. (*a*) Diode rectifier. (*b*) Diode detector. (*c*) Mixer.

The AC characteristics of a diode involve reactive effects due to the structure and packaging of the diode. A typical equivalent circuit for an RF diode is shown in Figure 7.

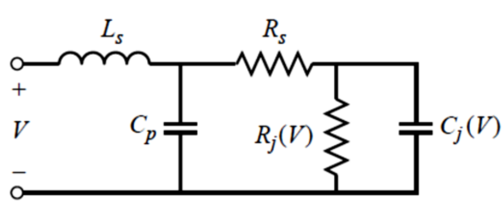


Figure 7: Equivalent AC circuit model for a Schottky diode.

The leads or contacts of the diode package are modeled as a series inductance, L_s, and shunt capacitance, C_p. The series resistor, R_s accounts for contact and current-spreading resistance. The junction capacitance, C_j, and the junction resistance, R_j, are bias dependent. Table 1 lists some parameters for a few commercially available Schottky diodes.

Schottky Diode	I_s (A)	R_s (Ω)	C_j (pF)	L_s (nH)	C_p (pF)
Skyworks SMS1546	3×10^{-7}	4	0.38	1.0	0.07
Skyworks SMS7630	5×10^{-6}	20	0.14	0.05	0.005
Avago HSMS2800	3×10^{-8}	30	1.6	—	—
Macom MA4E2054	3×10^{-8}	11	0.1	—	0.11

Table 1: Parameters for Some Commercial Schottky Diodes

Detailed technical data for surface mount RF schottky barrier diodes HSMS280x series and surface mount microwave schottky detector diodes HSMS280x series from AVAGO Technologies are presented in attached PDF files (Avago-HSMS-280x.pdf and Avago-HSMS-286x).

PIN Diodes and Control Circuits

Switches are used extensively in microwave systems for directing signal or power flow between components. Switches can also be used to construct other types of control circuits, such as phase shifters and attenuators. Mechanical switches can be made in waveguide or coaxial form, and can handle high powers but are bulky and slow. PIN diodes, however, can be used to construct an electronic switching element easily integrated with planar circuitry and capable of high-speed operation, such as used in reconfigurable microstrip filters. Switching speeds typically range from 1 to 10 μs, although speeds as fast as 20 ns are possible with careful design of the diode driving circuit. PIN diodes can also be used as power limiters, modulators, and variable attenuators.

PIN diode characteristics: A PIN diode contains an intrinsic (lightly doped) layer between the p and n semiconductor layers. When reverse biased, a small series junction capacitance leads to a relatively high diode impedance, while a forward bias current removes the junction capacitance and leaves the diode in a low−impedance state. These characteristics make the PIN diode a useful RF switching element. Equivalent circuits for the forward− and reverse−biased states are shown in Figure 8.

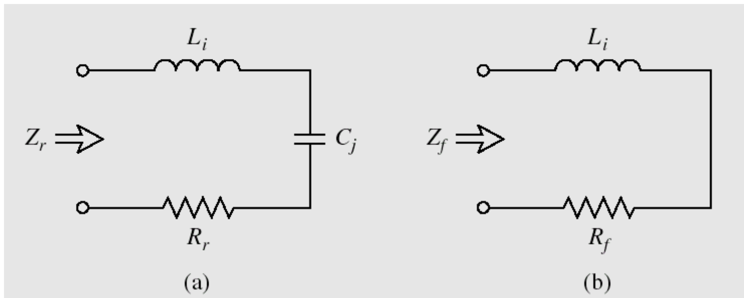


Figure 8: Equivalent circuits for the ON and OFF states of a PIN diode. (*a*) Reverse bias (OFF) state. (*b*) Forward bias (ON) state

The parasitic inductance, L_i, is typically less than 1 nH. The reverse resistance, R_r is usually small relative to the series reactance due to the junction capacitance and is often ignored. The forward bias current is typically 10–30 mA, and the reverse bias voltage is typically 10–60 V. The bias voltages must be applied to the diode with RF chokes and DC blocks for isolation from the RF signal. Table 2 lists parameters for some commercially available PIN diodes.

PIN Diode	R_f (Ω)	C_j (pF)
ASI 8001	3.0	0.03
Skyworks DSG9500	4.0	0.025
Infineon BA592	0.36	1.4
Microsemi UM9605	1.5	0.5

Table 2: Parameters for some commercial pin diodes

Detailed technical data for surface mount RF PIN low distortion attenuator diodes HSMP−381x, 481x series and PIN diode diversity switch HSMP−389D from AVAGO Technologies are presented in attached PDF files (Avago−HSMP−381x−481x.pdf and Avago−HSMP−389D).

Varactor Diodes

We have seen that a PIN diode has a junction capacitance that can be switched on or off with bias voltage. This effect can be enhanced by tailoring the size and doping profile of the intrinsic layer of the diode to provide a desired junction capacitance versus junction voltage (C vs. V) behavior when reverse biased. Such a device is called a varactor diode, and it produces a junction capacitance that varies smoothly with bias voltage, thus providing an electrically adjustable reactive circuit element.

One of the most common applications of varactor diodes is to provide electronic frequency tuning of the local oscillator in a multichannel receiver, such as those used in cellular telephones, wireless local area network radios, and television receivers. This is accomplished by using a varactor diode in the resonant circuit of a transistor oscillator, and controlling the DC reverse bias voltage applied to the diode. The nonlinearity of varactor diodes also makes them useful for frequency multipliers. Varactor diodes are generally made from silicon for RF applications, and gallium arsenide GaAs for microwave applications.

A simplified equivalent circuit for a reverse-biased varactor diode is shown in Figure 9. In the equivalent circuit, R_s is the series junction and contact resistance, typically on the order of a few ohms. A typical GaAs varactor diode may have a junction capacitance C_j that varies from about 0.1 to 2.0 pF as the bias voltage ranges from -20 to 0 V. Parasitic reactances due to the diode package should be included in a realistic design.

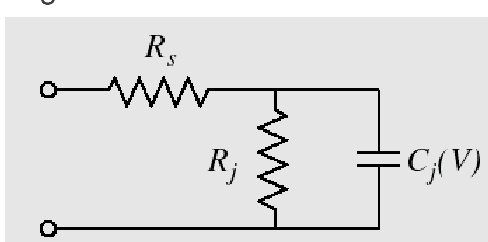


Figure 9: Equivalent circuit of a reverse biased varactor diode.

Detailed technical data for varactor diode NTE617 from NTE Electronics Inc. is presented in attached PDF files (NTE617.pdf).

Bipolar Junction Transistors

Transistors are three−terminal semiconductor devices, and can be categorized as either junction transistors or field effect transistors (FETs). Junction transistors include bipolar junction transistors (BJTs) that use a single semiconductor material (usually silicon), and heterojunction bipolar transistors (HBTs) that use compound semiconductors. Both npn and pnp configurations are possible, but most RF junction transistors are usually of the npn type due to higher electron mobility at higher frequencies.

Bipolar Junction Transistor

RF bipolar junction transistors (BJTs) are usually made using silicon (Si), and this transistor is one of the oldest and most popular active RF devices in use today because of its low cost and good operating performance in terms of frequency range, power capacity, and noise characteristics. Silicon junction transistors are useful for amplifiers up to the range of 2–10 GHz, and in oscillators up to about 20 GHz. Bipolar transistors typically have very low $1/f$ Flicker noise characteristics, making them well suited for oscillators with low−phase noise.

BJTs are sometimes preferred over FETs at frequencies below about 2–4 GHz because of higher gain and lower cost, and the possibility of biasing with a single power supply. Bipolar transistors are subject to shot noise as well as thermal noise effects, so their noise figure is not as good as that of FETs. The BJT is current driven, with the base current modulating the collector current. The upper frequency limit of the bipolar transistor is controlled primarily by the base length, which is typically on the order of 0.1 μm. A small−signal equivalent circuit model for an RF bipolar transistor is shown in Figure 10 for a common emitter configuration. This model, known as the hybrid−π model, is popular because of its similarity to the equivalent circuit of a FET, and because of its utility in circuit analysis. This model does not include parasitic resistances and inductances due to the base and emitter leads.

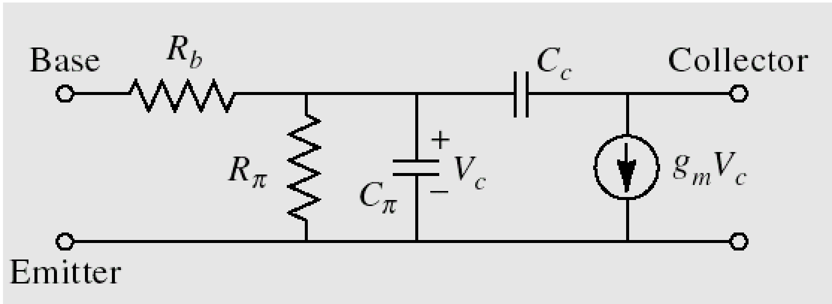


Figure 10: Simplified hybrid−π equivalent circuit for a microwave bipolar transistor in the common emitter configuration.

In many cases the capacitor, C_c between the base and collector in the hybrid−π model, has a relatively small value and may be ignored. This has the effect of making, implying that power only flows in one direction through the device (from port 1 to port 2); such a device is called unilateral. This approximation is often used to simplify analysis.

In practice, scattering parameters, measured under typical operating conditions, are supplied by the device manufacturer. Table 3 shows scattering parameters for a typical RF silicon junction transistor in a common emitter configuration. Note that there are relatively large mismatches at the base (port 1) and the collector (port 2), and that the gain (given roughly by $|S_{21}|$) drops quickly with an increase in frequency. Also note that $|S_{21}|$ is relatively small (particularly at low frequencies), making the device approximately unilateral.

Frequency (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
0.1	$0.78\angle-33°$	$0.03\angle71°$	$12.7\angle155°$	$0.93\angle-17°$
0.5	$0.46\angle-113°$	$0.08\angle52°$	$6.3\angle104°$	$0.53\angle-38°$
1.0	$0.38\angle-158°$	$0.11\angle54°$	$3.5\angle80°$	$0.40\angle-43°$
2.0	$0.40\angle157°$	$0.19\angle56°$	$1.9\angle52°$	$0.33\angle-63°$
4.0	$0.52\angle117°$	$0.38\angle45°$	$1.1\angle14°$	$0.33\angle-127°$

Table 3: Scattering Parameters for an NPN Silicon BJT

(NEC NE 58219 $V_{ce} = 5.0\ V$, $I_c = 5.0\ mA$, common emitter)

The equivalent circuit of Figure 10 can be used to estimate the upper frequency limit, f_T, defined as the threshold frequency where the short−circuit current gain of the transistor is unity.

$$f_T = \frac{g_m}{2\pi C_\pi}$$

Heterojunction Bipolar Transistor

The operation of a heterojunction bipolar transistor (HBT) is essentially the same as that of a BJT, but an HBT has a base−emitter junction made from a compound semiconductor material such as GaAs, indium phosphide (InP), or silicon germanium (SiGe), often in conjunction with thin layers of other materials (e.g., aluminum). This structure offers much improved performance at high frequencies. Some HBTs can operate at frequencies exceeding 100 GHz, and recent developments with HBTs using SiGe have demonstrated that these devices are useful in low−cost circuits operating at frequencies of 60 GHz or higher.

Since the HBT is similar in structure and operation to the BJT, the equivalent circuit model of Figure 10 can be used for both transistor types. As with BJTs, equivalent circuit models may have limited applicability when attempting to model HBTs over a range of operating conditions, so scattering parameter data, measured for a particular bias point, may be more useful. Table 4 gives the scattering parameters at several frequencies for a popular microwave HBT. Observe that $|S_{21}|$ decreases much less rapidly with frequency when compared with the BJT of Table 4. The device also is seen to be approximately unilateral, as $|S_{21}|$ is relatively small.

Frequency (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
1.0	$0.91\angle-44°$	$0.06\angle68°$	$3.92\angle149°$	$0.93\angle-17°$
2.0	$0.75\angle-86°$	$0.10\angle46°$	$3.39\angle120°$	$0.79\angle-31°$
4.0	$0.59\angle-144°$	$0.11\angle29°$	$2.18\angle82°$	$0.64\angle-43°$
6.0	$0.54\angle176°$	$0.11\angle34°$	$1.64\angle57°$	$0.58\angle-53°$

Table 4: Scattering Parameters for a SiGe HBT

(Infineon BFP640F, $V_{ce} = 2.0\ V$, $I_c = 1.2\ mA$, common emitter)

High levels of monolithic integration are easy and inexpensive with SiGe HBTs, so this technology is proving to be very useful for low-cost millimeter wave circuits for both defense and commercial applications.

Field Effect Transistors

In contrast to BJTs, field effect transistors (FETs) are monopolar, with only one carrier type (holes or electrons) providing current flow through the device: n-channel FETs employ electrons, while p-channel devices use holes. In addition, while a BJT is a current-controlled device, an FET is a voltage-controlled device, having a source-to-drain characteristic that is similar to that of a voltage-dependent variable resistor.

Field effect transistors can take many forms, including the MESFET (metal semiconductor FET), the MOSFET (metal oxide semiconductor FET), the HEMT (high electron mobility transistor), and the PHEMT (pseudomorphic HEMT). FET transistor technology has been under continuous development for more than 50 years—the first junction FETs were developed in the 1950s, while the HEMT was proposed in the early 1980s. GaAs MESFETs are among the most commonly used transistors for microwave and millimeter wave applications, being usable at frequencies up to 60 GHz or more. Even higher operating frequencies can be obtained with GaAs HEMTs. GaAs MESFETs and HEMTs are especially useful for low-noise amplifiers since these transistors have lower noise figures than any other active devices. Recently developed gallium nitride (GaN) HEMTs are very useful for high power RF and microwave amplifiers. CMOS FETs are increasingly being used for RF integrated circuits, offering high levels of integration at low cost and low power requirements, for commercial wireless applications. Table 5 summarizes the performance characteristics of some of the most popular microwave transistors.

Device	BJT	HBT	CMOS	MESFET	HEMT	HEMT
Semiconductor	Si	SiGe	Si	GaAs	GaAs	GaN
Frequency range (GHz)	10	30	20	60	100	10
Typical gain (dB)	10–15	10–15	10–20	5–20	10–20	10–15
Noise figure (dB)	2.0	0.6	1.0	1.0	0.5	1.6
(frequency, GHz)	(2)	(8)	(4)	(10)	(12)	(6)
Power capacity	High	Medium	Low	Medium	Medium	High
Cost	Low	Medium	Low	Medium	High	Medium
Single-polarity supply	Yes	Yes	Yes	No	No	No

Table 5: Performance characteristics of microwave transistors

Metal Semiconductor Field Effect Transistor

One of the most important developments in microwave technology has been the GaAs metal semiconductor field effect transistor (MESFET), as this device permitted the first practical solid−state implementation of amplifiers, oscillators, and mixers at microwave frequencies, leading to key applications in radar, GPS, remote sensing, and wireless communications. GaAs MESFETs can be used at frequencies well into the millimeter wave range, with high gain and low noise figure, often making them the device of choice for hybrid and monolithic integrated circuits at frequencies above 10 GHz.

Figure 11 shows the cross section of a typical n−channel GaAs MESFET. The gate junction is formed as a Schottky barrier. The desirable gain and noise features of this transistor are a result of the higher electron mobility of GaAs compared to silicon, and the absence of shot noise. The device is biased with a drain−to−source voltage, V_{ds}, and a gate−to−source voltage, V_{gs}. In operation, electrons are drawn from the source to the drain by the positive V_{ds} supply voltage. An applied signal voltage on the gate then modulates these majority electron carriers, producing voltage amplification. The maximum frequency of operation is limited by the gate length; present FETs have gate lengths on the order of 0.2–0.6 μm, with corresponding upper frequency limits of 100 to 50 GHz.

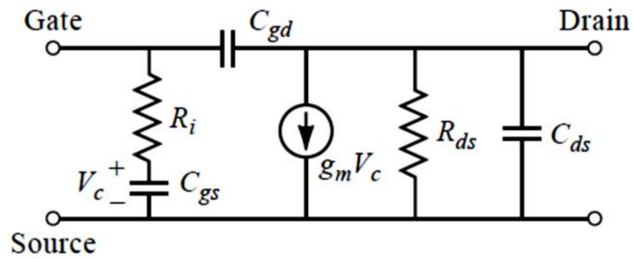


Figure 11: Small−signal equivalent circuit for a microwave FET in the common−source configuration.

This model does not include package parasitics, which typically introduce small series resistances and inductances at the three terminals due to ohmic contacts and bonding leads. The dependent current generator $g_m V_c$ depends on the voltage across the gate−to−source capacitor C_{gs}, leading to a value of $|S_{21}| > 1$ under normal operating conditions (where port 1 is at the gate, and port 2 is at the drain). The reverse signal path, given by S_{21} is due solely to the capacitance C_{gd} As seen from the above data, this is typically a very small capacitor, which can often be ignored in practice. In this case, $S_{12} = 0$, and the device is unilateral. The scattering parameters for a typical GaAs MESFET are given in Table 6.

Frequency (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
1.0	$0.97\angle -28°$	$0.04\angle 72°$	$3.82\angle 154°$	$0.70\angle -19°$
2.0	$0.90\angle -55°$	$0.08\angle 54°$	$3.56\angle 129°$	$0.65\angle -37°$
4.0	$0.72\angle -103°$	$0.12\angle 28°$	$2.91\angle 86°$	$0.53\angle -68°$
8.0	$0.52\angle 179°$	$0.14\angle -1°$	$2.0\angle 20°$	$0.42\angle -129°$
12.0	$0.49\angle 103°$	$0.17\angle -19°$	$1.5\angle -38°$	$0.44\angle 170°$

Table 6: Scattering parameters for an n−channel GaAs MESFET (NEC NE76184A, $V_{DS} = 3.0\ V$, $I_D = 10.0\ mA$, common source)

Metal Oxide Semiconductor Field Effect Transistor

The silicon metal oxide semiconductor field effect transistor (MOSFET) is the most common type of FET, being used extensively in analog and digital integrated circuits.

MOSFETs can be used at frequencies into the UHF range, and can provide powers of several hundred watts when devices are packaged in parallel. Laterally diffused MOSFETs (LDMOS) have direct grounding of the source, and can operate at low microwave frequencies with high powers. These devices are commonly used for high-power transmitters for cellular base stations at 900 and 1900 MHz. High-density integrated circuits typically use complementary MOS (CMOS), where both n-channel and p-channel devices are used. This technology is very mature, and has the advantages of low power requirements and low unit cost. Most RF and microwave MOSFETs use n-channel silicon devices, although GaN devices are possible. The small-signal equivalent circuit for a MOSFET is the same as that of the MESFET. Scattering parameters are available for most nMOS devices intended for high-frequency applications.

High Electron Mobility Transistor

The high electron mobility transistor (HEMT) is a heterojunction FET, meaning that it does not use a single semiconductor material, but instead is constructed with several layers of compound semiconductor materials. These may include transitions between gallium aluminum arsenide (GaAlAs), GaAs, gallium indium arsenide (GaInAs), and similar compounds. These structures result in high carrier mobility—about twice that found in a standard MESFET. GaAs HEMTs can operate at frequencies above 100 GHz.

The relatively complicated structure of the HEMT requires sophisticated fabrication techniques, leading to a relatively high cost. The HEMT is also referred to in the literature as a MODFET (modulation-doped FET), a TEGFET (two-dimensional electron gas FET), and an SDFET (selectively doped FET).

A relatively new type of HEMT uses GaN and aluminum gallium nitride (AlGaN) on a silicon or SiC substrate. GaN HEMTs operate with drain voltages in the range of 20–40 V, and can deliver powers up to 100 W at frequencies in the low microwave range, making these devices popular for high−power transmitters. The equivalent circuit model of Figure 11 can also be used for HEMTs, and the DC bias characteristics of a HEMT are similar to those of the MESFET. Table 11.8 gives the scattering parameters for a medium power GaN HEMT.

Frequency (GHz)	S_{11}	S_{12}	S_{21}	S_{22}
0.5	$0.96\angle 180°$	$0.007\angle -16°$	$3.67\angle 68°$	$0.72\angle -174°$
1.0	$0.95\angle 172°$	$0.008\angle -35°$	$2.03\angle 44°$	$0.78\angle -172°$
2.0	$0.78\angle 153°$	$0.014\angle -83°$	$2.09\angle -17°$	$0.91\angle -174°$
4.0	$0.88\angle -51°$	$0.008\angle 79°$	$0.84\angle 88°$	$0.88\angle 171°$

Table 7: Scattering parameters for a GaN HEMT

(Cree CGH21120, $V_{DD} = 328\ V$, $I_D = 500mA$, common source)

Chapter Three:

Noise and Nonlinear Distortion in RF and Microwave systems

Keywords

Thermal noise, white noise, noise power, noise figure, equivalent noise temperature, signal–to–noise ratio (SNR), nonlinear distortion, gain compression, 1 dB compression point $P_{1\text{dB}}$, minimum detectable signal (MDS), receiver sensitivity, linear dynamic range (LDR), intermodulation (IM) products, third–order intercept point IP_3 spurious free dynamic range (SFDR).

Abstract

In this chapter, we will discuss the effect of noise on the performance of RF and microwave receiver systems, and the characterization of components in terms of noise figure, including the effect of impedance mismatch. The additional noise-related topics of transistor amplifier noise figure, oscillator phase noise, and antenna noise temperature will be discussed in later chapters.

We will also discuss in this chapter the related topics of nonlinear distortion (gain compression, harmonic distortion, intermodulation distortion), and dynamic range. These can have important limiting effects when large signal levels are present in mixers, amplifiers, and other components that use nonlinear devices such as diodes and transistors.

Learning Objectives

In this chapter, student will be able to

- Recall the concept of noise power, noise figure, and equivalent noise temperature.
- Identify different types of nonlinear distortion: gain compression, harmonic distortion, intermodulation distortion.
- Recall the concept of minimum detectable signal, and receiver sensitivity.
- Identify third–order intercept point of a nonlinear device.
- Recall the concept of linear dynamic range and spurious free dynamic range.
- Apply acquired techniques and rules to evaluate the noise figure of a receiver system.
- Apply acquired techniques and rules to evaluate the third–order intercept point of a receiver system.
- Apply acquired techniques and rules to evaluate the linear and spurious free dynamic ranges of a receiver system.

Introduction

The effect of noise is critical to the performance of wireless communications systems, such as mobile communications systems, RF and microwave systems, microwave links, radar, and remote sensing systems. Noise ultimately determines the threshold for the minimum signal that can be reliably detected by a receiver, which is called sensitivity that is related to the signal to noise ratio at the output of the receiver. For instance, it is crucial to account for noise in order to predict the reliability of coverage provided by any mobile cellular system.

Noise power in a receiver will be introduced from the external environment through the receiving antenna, as well as generated internally by the receiver circuitry. In this chapter, we will study the sources of noise in RF and microwave systems, and the characterization of components in terms of noise temperature and noise figure, including the effect of impedance mismatch. The additional noise–related topics of transistor amplifier noise figure, oscillator phase noise, and antenna noise temperature will be discussed in later chapters.

We will also discuss in this chapter the effects of nonlinearity of receiver components, such as compression, harmonic distortion, intermodulation distortion, and dynamic range. These can have important limiting effects when large signal levels are present in mixers, amplifiers, and other components that use nonlinear devices such as diodes and transistors.

Sources of Receiver Noise

The receiver encounters two types of noise: the noise picked up by the antenna and the noise generated by the receiver.

The noise picked up by the antenna includes sky noise, earth noise, atmospheric (or static) noise, galactic noise, and man–made noise, as shown in Figure 1.

- The sky noise has a magnitude that varies with frequency and the direction to which the antenna is pointed.
- Static or atmospheric noise is due to a flash of lightning somewhere in the world. The lightning generates an impulse noise that has the greatest magnitude at 10 kHz and is negligible at frequencies greater than 20MHz.
- Galactic noise is produced by radiation from distant stars. It has a maximum value at about 20 MHz and is negligible above 500MHz.
- Man–made noise includes many different sources. For example, when electric current is switched on or off, voltage spikes will be generated. These transient spikes occur in electronic or mechanical switches, vehicle ignition systems, light switches, motors, and so on. Electromagnetic radiation from communication systems, broadcast systems, radar, and power lines is everywhere, and the undesired signals can be picked up by a receiver. The interference is always present and could be severe in urban areas.

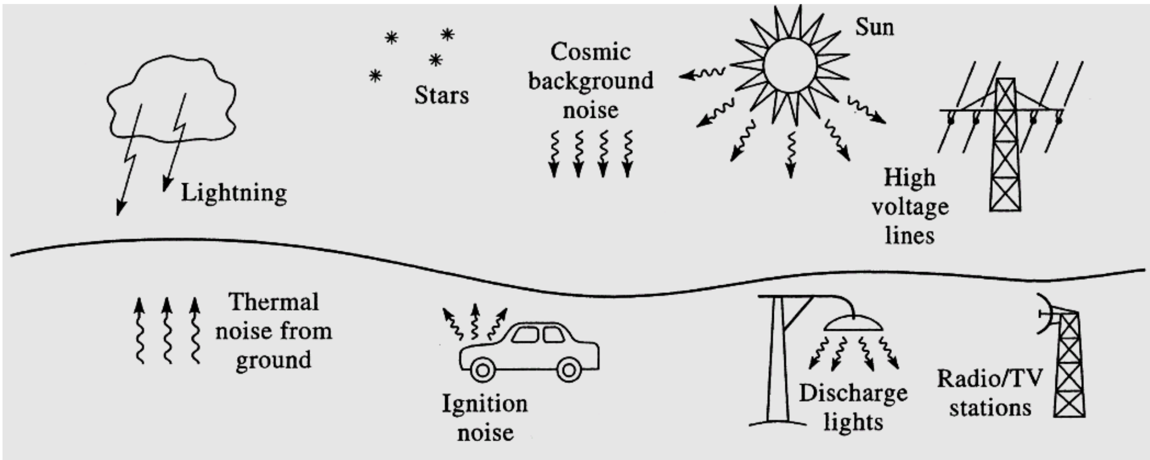


Figure 1: Natural and manmade sources of background noise.

In addition to the noise picked up by the antenna, the receiver itself adds further noise to the signal from its amplifier, filter, mixer, oscillator, and detector stages. If all these components were assumed linear, the output signal level will be directly proportional to the input signal level, and if they were assumed deterministic, the output signal will be predictable from the input signal.

In reality no component can perform in this way over an unlimited range of input/output signal levels. In practice, however, there is usually a range of signal levels over which such assumptions are approximately valid; this range is called the dynamic range of the component. As an example, consider a realistic amplifier having a power gain G, as shown in Figure 2. If the amplifier were ideal, the output power would be related to the input power as $P_{out} = GP_{in}$ and this relation would hold true for any value of P_{in}.

Thus, $P_{in} = 0$, we would have $P_{out} = 0$, and if $P_{in} = 10^6\,\mathrm{W}$ and $G = 10$ dB, we would have $P_{in} = 10^7\,\mathrm{W}$. Neither of these results would actually occur in practice, however. Because of noise generated by the amplifier itself, some nonzero noise power will always be delivered by the amplifier, even when the input power is zero. At the other extreme, very high input power will cause the amplifier to fail. Thus, the actual relation between the output and input power will be as shown in Figure 2. At very low input power levels, the output will be dominated by the noise generated by the amplifier. This level is often called the noise floor of the component or system; typical values may range from -80 to -140 dBm over the bandwidth of the system, with the lowest values being obtained with thermally cooled components. Above the noise floor, the amplifier will have a range of input power for which $P_{out} = GP_{in}$ is closely approximated. This is the usable dynamic range of the component. At the upper end of this range, the output will begin to saturate, meaning that the output power no longer increases linearly as the input power increases. Excessive input power will lead to failure of the amplifier.

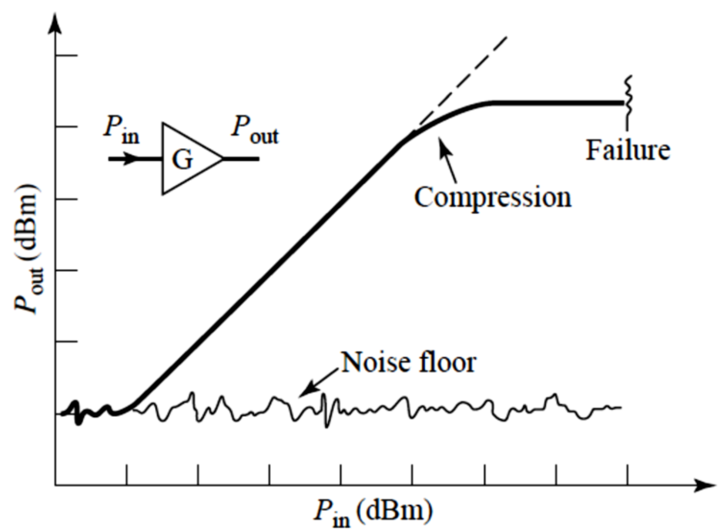


Figure 2: Dynamic range of a realistic amplifier.

The noise that occurs in a receiver acts to mask weak signals and to limit the ultimate sensitivity of the receiver. In order for a signal to be detected, it should have a strength much greater than the noise floor of the system.

Various types of noise that is generated internally in receiver components are:

Thermal noise is the most basic type of noise, being caused by thermal vibration of bound charges. It is also known as Johnson or Nyquist noise.

Shot noise is due to random fluctuations of charge carriers in an electron tube or solid-state device.

Flicker noise occurs in solid-state components and vacuum tubes. Flicker noise power varies inversely with frequency, and so is often called $1/f$ .noise. The noise is important from 1 Hz to 1 MHz. Beyond 1 MHz, the thermal noise is more noti $1/f$ ceable.

The characterization of noise effects in receiver systems in terms of noise figure will apply to all types of noise, regardless of the source, as long as the spectrum of the noise is relatively flat over the bandwidth of the system. Noise with a flat frequency spectrum is called white noise.

The quality of the output signal from the receiver for its intended purpose is expressed in terms of its signal-to-noise ratio (SNR):

$$SNR = \frac{\text{wanted signal power}}{\text{unwanted noise power}}$$

Noise Power

A noise source at absolute temperature T in degrees kelvin (K), produces a maximum available noise power given by

$$P_N = kTB$$

Where B is the bandwidth which in practice is usually limited by the passband filter of the receiver, and $k = 1.380 \times 10^{(-23)} J / K$ is Boltzmann's constant.

Note that this noise power is:

- Independent of frequency; such a noise source has a power spectral density that is constant with frequency, and is an example of a **white noise** source.
- Directly proportional to the bandwidth, which in practice is usually limited by the passband of the RF or microwave system.
- Independent white noise sources can be treated as Gaussian−distributed random variables, so the noise powers (variances) of independent noise sources are **additive**.

Noise Figure

Noise figure is a figure of merit quantitatively specifying how noisy a component or system is. It is a measure of the degradation in the signal−to−noise ratio between the input and output of the component. The signal−to−noise ratio is the ratio of desired signal power to undesired noise power, and so is dependent on the signal power.

The noise factor of a two−port network is defined as:

$$F = \frac{SNR \text{ at input}}{SNR \text{ at output}} = \frac{S_i / N_i}{S_o / N_o}$$

By definition, the input noise power is assumed to be the noise power resulting from a matched load at $T_0 = 290\,\mathrm{K}$ (room temperature); that is $N_i = kT_0B$, since the noise factor of a component should be independent of the input noise.

When noise and a desired signal are applied to the input of a noiseless network, both noise and signal will be attenuated or amplified by the same factor, so that the signal−to−noise ratio will be unchanged; that is $F = 1$.

When the network is noisy, the output noise power will be increased more than the output signal power, so that the output signal−to−noise ratio will be reduced; that is $F > 1$. The noise factor F , , is a measure of this reduction in signal−to−noise ratio.

The noise figure is simply the noise factor converted in decibel notation:

$$NF = 10 \log F$$

The equivalent noise temperature is defined as

$$T_e = (F - 1)T_0$$

Therefore:

$$F = 1 + \frac{T_e}{T_0}$$

Note that T_e is not the physical temperature. If the network were noiseless, T_e would be zero, giving $F = 1$, or 0 dB.

Example 1: Added noise by active device

Consider a two-port network with a gain (or loss) G as shown in figure 3. We have

$$S_o = GS_i$$

Note that $N_o \neq GN_i$; instead, the output noise $N_o = GN_i +$ noise added by the network. The noise added by the network is $N_n = N_o - GN_i$.

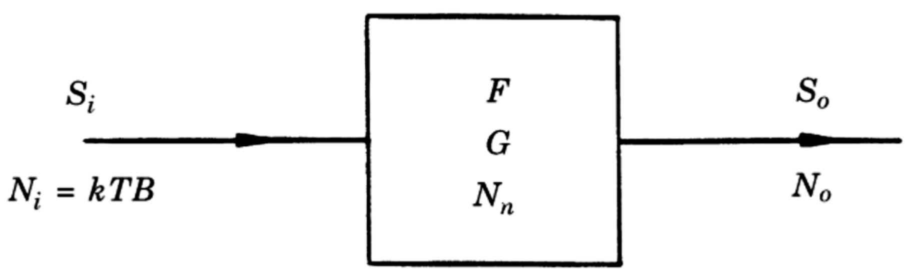


Figure 3: two-port network with gain (or loss) G and added noise power N_n

The noise factor is:

$$F = \frac{S_i / N_i}{S_o / N_o} = \frac{S_i}{GS_i} \frac{N_o}{N_i} = \frac{N_o}{GN_i}$$

Therefore:

$$N_o = FGN_i$$

Where N_o and N_i are measured in Watts. In dB:

$$N_o\left[\text{dBm}\right] = NF + G\left[\text{dB}\right] + N_i\left[\text{dBm}\right]$$

This implies that the input noise N_i (in dBm) is raised by the noise figure NF and the gain (in dB) of the two-port network.

But $N_i = kT_0B$, and:

$$kT_0 = 1.380\times10^{-23}\times290 = 4.0\times10^{-21}\ \text{J} = 4.0\times10^{-18}\ \text{mW/Hz}$$

$$10\ \log kT_0 = -174\,\text{dBm/Hz}$$

kT_0 is the spectral density of the noise power at room temperature T_0. Therefore,

$$N_o\left[\text{dBm}\right] = NF + G\left[\text{dB}\right] + 10\ \log kT_0 + 10\ \log B$$

$$N_o\left[\text{dBm}\right] = -174 + NF + G\left[\text{dB}\right] + 10\ \log B$$

Example 2: Added noise by passive lossy device

An important special case occurs in practice for a two−port network consisting of a passive, lossy component, such as an attenuator or lossy transmission line, held at a physical temperature T .

Consider such a network with a matched source resistor R that is also at temperature T , as shown in Figure 4. The power gain, G , of a lossy network is less than unity; the loss factor, L , can be defined as $L = 1/G > 1$.

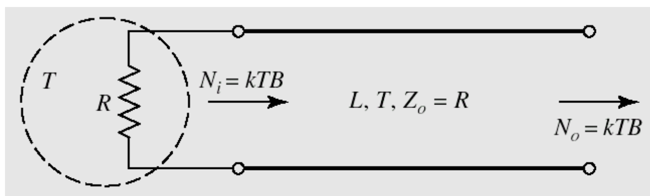


Figure 4: lossy line or attenuator with loss L and temperature T .

Because the entire system is in thermal equilibrium at the temperature T , and has a driving point impedance of R , the output noise power must be $N_o = kTB$. However, we can also think of this power as coming from the source resistor (attenuated by the lossy line), and

$$N_o = kTB = GN_i + GN_n = GkTB + GN_n$$

Where N_n is the noise generated by the line, as if it appeared at the input terminals of the line. Therefore:

$$N_n = \frac{1-G}{G} kTB = (L-1) kTB$$

This result shows that the lossy line has an equivalent noise temperature (referred to the input) given by

$$N_n = kT_e B = \frac{1-G}{G} kTB = (L-1)kTB \to T_e = \frac{1-G}{G} T = (L-1)T$$

Then the noise factor is

$$F = 1 + \frac{T_e}{T_0} = 1 + (L-1)\frac{T}{T_0}$$

If the line is at temperature T_0, then $F = L$. For instance, a 6 dB attenuator at room temperature T_0 has a noise figure of $F = 6\text{dB}$

Important result

If a passive two port network with loss factor L is at temperature T_0, then $F = L$.

Example 3: Noise figure of cascaded system

For the two-element cascaded system shown in Figure 5, prove that the overall noise factor

$$F = F_1 + \frac{F_2 - 1}{G_1}$$

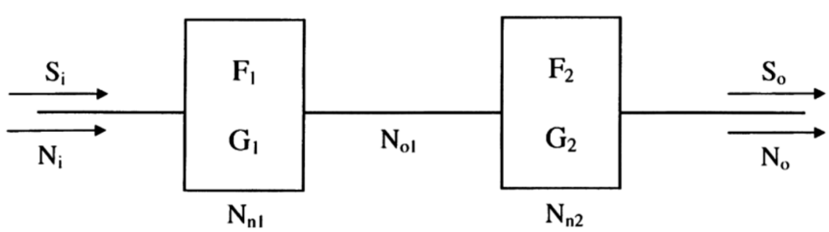


Figure 5: Two-element cascaded system.

Solution

From example 1:

$$N_o = F_{12} G_{12} k T_0 B ; N_{o1} = F_1 G_1 k T_0 B$$

Where $F = F_{12}$ is the required overall noise factor and $G_{12} = G_1 G_2$ is the overall gain.

By definition, the added power noise by the element 2 is

$$N_{n2} = N_o - G_2 k T_0 B = F_2 G_2 k T_0 B - G_2 k T_0 B = (F_2 - 1) G_2 k T_0 B$$

Therefore:

$$N_o = G_2 N_{o1} + N_{n2}$$

$$N_o = G_2 F_1 G_1 k T_0 B + (F_2 - 1) G_2 k T_0 B$$

The equivalent system of the Two−element cascaded system has a noise factor F_{12} Then:

$$N_o = F_{12} G_{12} k T_0 B$$

Overall:

$$F_{12} G_{12} k T_0 B = G_2 F_1 G_1 k T_0 B + (F_2 - 1) G_2 k T_0 B$$

$$F = F_{12} = F_1 + \frac{F_2 - 1}{G_1}$$

Cascaded System

The proof can be generalized to find the noise factor of a cascaded system of n elements:

$$F = F_1 + \frac{F_2 - 1}{G_1} + \frac{F_3 - 1}{G_1 G_2} + \cdots$$

The overall equivalent noise temperature of n−element system is

$$T_e = T_{e1} + \frac{T_{e2}}{G_1} + \frac{T_{e3}}{G_1 G_2} + \cdots$$

Problem 1: Noise Analysis of a Wireless Receiver

The block diagram of a wireless receiver front−end is shown in Figure 6. Assume the system is at temperature T_0 with a characteristic impedance of 50Ω, and an IF bandwidth of 10 MHz.

1. Compute the overall noise figure of this subsystem.
2. If the input noise power from a feeding antenna is $N_i = kT_A B$, where $T_A = 150\ K$, find the output noise power in dBm.
3. If we require a minimum signal−to−noise ratio of 20 dB at the output of the receiver, what is the minimum signal voltage that should be applied at the receiver input?

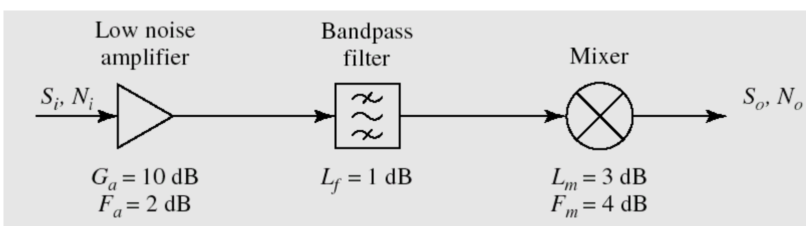


Figure 6: block diagram of a wireless receiver front−end for problem 1.

Solution

1. We first perform the required conversions from dB to numerical values:

$G_a = 10\,\text{dB} = 10$	$G_f = -1\,\text{dB} = 0.79$	$G_m = -3\,\text{dB} = 0.5$
$F_a = 2\,\text{dB} = 1.58$	$F_f = 1\,\text{dB} = 1.26$	$F_m = 4\,\text{dB} = 2.51$

Next, we use the noise factor of a cascaded system to find the overall noise figure of the receiver front-end system:

$$F = F_a + \frac{F_f - 1}{G_a} + \frac{F_m - 1}{G_a G_f}$$

$$F = 1.58 + \frac{1.26 - 1}{10} + \frac{2.51 - 1}{10 \times 0.79} = 1.80$$

$$NF = 10\,\log 1.80 = 2.55\,\text{dB}$$

Note: If we rearrange the cascaded system by putting the bandpass filter first and the low-noise amplifier next, we find

$$F = F_f + \frac{F_a - 1}{G_f} + \frac{F_m - 1}{G_a G_f}$$

$$F = 1.26 + \frac{1.58 - 1}{0.79} + \frac{2.51 - 1}{10 \times 0.79} = 2.19$$

$$NF = 10\,\log 2.19 = 3.40\,\text{dB}$$

We notice that the noise figure is higher.

Important note on cascaded system

This result shows that the noise characteristics of a cascaded system are dominated by the characteristics of the first stage since the effect of the second stage is reduced by the gain of the first (assuming $G_1 > 1$). **Thus, for the best overall system noise performance, the first stage should have a low noise figure and at least moderate gain.** Expense and effort should be devoted primarily to the first stage, as opposed to later stages, since later stages have a diminished impact on the overall noise performance.

1. The best way to compute the output noise power is to use noise temperatures. The equivalent noise temperature of the overall system is

$$T_e = (F-1)T_0 = (1.80-1)\times 290 = 232\,\mathrm{K}$$

The overall gain of the system is

$$G = G_a G_f G_m = 3.95$$

Then we can find the output noise power as

$$N_o = GN_i + N_n = GkT_A B + GkT_e B = Gk\left(T_A + T_e\right)B$$

$$N_o = 3.95\left(1.380\times 10^{-23}\right)\left(150+232\right)\left(10\times 10^{6}\right) = 2.08\times 10^{-13}\,\mathrm{W}$$

$$N_o \approx 2\times 10^{-10}\,\mathrm{mW} = -97\,\mathrm{dBm}$$

Important Note: It may be tempting to compute the output noise power from the definition of the noise figure, as

$$N_o = FGN_i = FGkT_A B$$

$$N_o = 1.80 \times 3.95 \times \left(1.380 \times 10^{-23}\right)(150)\left(10 \times 10^6\right) = 1.47 \times 10^{-13}\ \text{W}$$

This is an **incorrect result!** The reason for the disparity with the earlier result is that the definition of noise figure assumes an input noise level of $\boldsymbol{kT_0B}$, while this problem involves an input noise of kT_AB, with $T_A \neq T_0$.

This is a common error, and suggests that when computing absolute noise power it is often safer to use noise temperatures to avoid this confusion.

1. For an output $\dfrac{S_o}{N_o} = 20\,\text{dB} = 100$, the input signal power must be

$$S_i = \frac{S_o}{G} = \frac{S_o}{N_o}\frac{N_o}{G} = 100 \times \frac{2.08 \times 10^{-13}}{3.95} = 5.27 \times 10^{-12}\ \text{W}$$

For a 50Ω system impedance, this corresponds to an input signal voltage of

$$V_i = \sqrt{Z_0 S_i} = \sqrt{50\left(5.27 \times 10^{-12}\right)} = 1.62 \times 10^{-5}\ \text{V} = 16.2\,\mu\text{V}\,(\text{rms}).$$

Noise Figure of a Mismatched Amplifier

We will consider the effect of an input impedance mismatch on the noise figure of an amplifier. As shown in Figure 7, the amplifier, when matched, has a gain G, a noise factor F, and a bandwidth B. The amplifier output is matched, but there is an impedance mismatch at the input represented by the reflection coefficient, Γ.

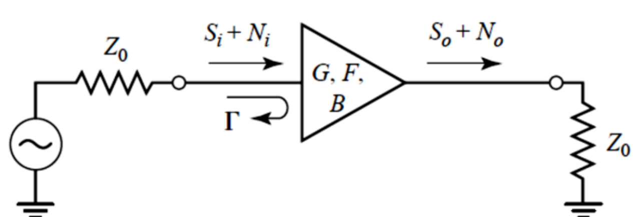


Figure 7: A noisy amplifier with an impedance mismatch at its input.

For an applied signal power S_i , the output signal power is

$$S_o = G\left(1-|\Gamma|^2\right)S_i$$

Since we are dealing with noise figure, let the input noise power to the amplifier be kT_0B . Then the output noise power from the amplifier (referenced to the input) is given by:

$$N_o = G\left(1-|\Gamma|^2\right)kT_0B + Gk\left(F-1\right)T_0B$$

The overall noise figure, F_m , of the mismatched amplifier is:

$$F_m = \frac{S_i/N_i}{S_o/N_o} = \frac{S_i}{S_o}\frac{N_o}{N_i} = \frac{1}{G\left(1-|\Gamma|^2\right)}\frac{G\left(1-|\Gamma|^2\right)kT_0B + Gk\left(F-1\right)T_0B}{kT_0B}$$

$$F_m = 1 + \frac{F-1}{1-|\Gamma|^2}$$

Observe the limiting case that $F_m = F$ when $\Gamma = 0$ (no mismatch), and that this is the minimum noise figure that can be achieved since F_m increases as the mismatch increases. This result demonstrates that good noise figure requires good impedance matching.

Nonlinear Distortion

In this section we will present the nonlinearity effect of active devices in general, and how they affect the sensitivity and dynamic range. This will be useful for understanding nonlinear parameters of amplifiers, mixers, and wireless receivers.

For input signal of frequency f , a linear device produces a sinusoidal output signal of the same frequency f . This frequency is called the fundamental. A nonlinear device produces a composite output signal containing the fundamental and the harmonics. A linear device is an electronic circuit in which, any steady−state output of the circuit (the current through any component, or the voltage between any two points) is also sinusoidal with frequency f .

We have seen that thermal noise is generated by any lossy component. Since all realistic components have at least a small loss, the ideal linear component does not exist in practice because all realistic devices are nonlinear at very low signal levels due to noise effects. In addition, practical components may also become nonlinear at high signal levels. In the case of active devices, such as diodes and transistors, this may be due to effects such as gain compression or the generation of spurious frequency components due to device nonlinearities, but all devices ultimately fail at very high power levels, as shown in figure 2. In either case, these effects set a minimum and maximum realistic power range, or dynamic range, over which a given component or network will operate as desired.

Gain Compression

A typical amplifier response is shown in Figure 8. For an ideal linear amplifier a plot of the output power versus input power would be a straight line with a slope of unity, and the power gain of the amplifier given by the ratio of the output power to the input power:

$$G = \frac{P_{out}\left[\mathrm{W}\right]}{P_{in}\left[\mathrm{W}\right]}$$

Or

$$G\left[\mathrm{dB}\right] = P_{out}\left[\mathrm{dBm}\right] - P_{in}\left[\mathrm{dBm}\right]$$

The amplifier response of Figure 8 tracks the ideal response over a limited range, then begins to saturate, resulting in reduced gain. This effect is called **gain compression**, or saturation.

To quantify the linear operating range of the amplifier, we define the 1 dB compression point as the power level for which the output power (solid line) has decreased by 1 dB from the ideal linear characteristic (dotted line). This power level is usually denoted by $P_{1\mathrm{dB}}$, and can be stated in terms of either input power $(IP_{1\mathrm{dB}})$ or output power $(OP_{1\mathrm{dB}})$.

The 1 dB compression point $P_{1\mathrm{dB}}$ is typically given as the larger of these two options, so for amplifiers $P_{1\mathrm{dB}}$ is usually specified as an output power $(OP_{1\mathrm{dB}})$, while for mixers $P_{1\mathrm{dB}}$ is usually specified in terms of input power $(IP_{1\mathrm{dB}})$.

The relation between a compression point referenced at the input versus the output is given as, in dB:

$$OP_{1\mathrm{dB}} = IP_{1\mathrm{dB}} + G\left[\mathrm{dB}\right] - 1\mathrm{dB}$$

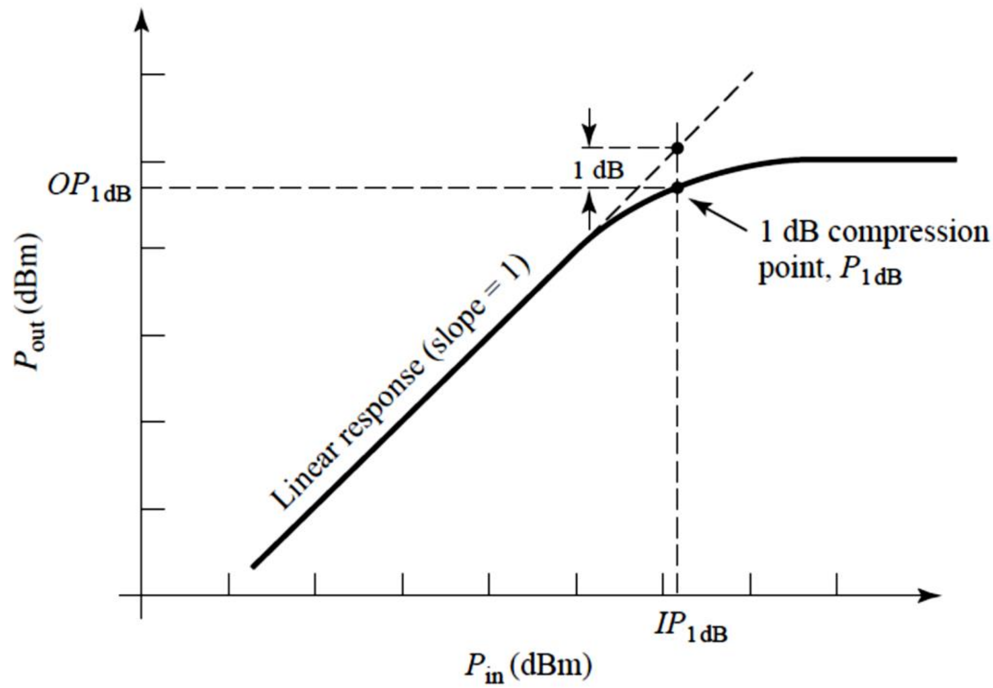


Figure 8: Definition of the 1 dB compression point for a nonlinear amplifier

In the linear region for an amplifier or a receiver,

$$P_{out}\left[\mathrm{dBm}\right]=P_{in}\left[\mathrm{dBm}\right]+G\left[\mathrm{dB}\right]$$

For a lossy mixer with a conversion loss L_c ,

$$P_{out}\left[\mathrm{dBm}\right]=P_{in}\left[\mathrm{dBm}\right]-L_c\left[\mathrm{dB}\right]$$

Minimum Detectable Signal

The minimum detectable signal (MDS) is the input that gives an output signal−to−noise ratio of 0 dB (where the signal power is equal to the noise power),

For $N_i = kT_0B$, and a receiver noise figure NF ,

$$\text{MDS}[\text{dBm}] = -174 + NF + 10\log B$$

Receiver Sensitivity

The sensitivity of a wireless receiver is defined as the smallest input RF signal power that can be processed to develop a minimum signal−to−noise ratio SNR_{min} for achieving a required error rate (BER−Bit Error Rate or FER−Frame Error Rate) by the system. The sensitivity is defined as SNR_{min} dB above the MDS and is given by

$$S_{min}[\text{dBm}] = \text{MDS}[\text{dBm}] + SNR_{min}$$

$$S_{min}[\text{dBm}] = -174 + NF + 10\log B + SNR_{min}$$

Linear Dynamic Range

We can define dynamic range in a general sense as the operating range for which a component or system has desirable characteristics. For a power amplifier this may be the power range that is limited at the low end by noise and at the high end by the compression point. This is essentially the linear operating range for the amplifier, and is called the linear dynamic range (LDR).

We can find the linear dynamic range LDR of the amplifier in Figure 9 as the ratio of $P_{1\text{dB}}$, the 1 dB compression point, to the noise level of the component. In dB, this can be written in terms of output powers as:

$$\text{LDR}[\text{dB}] = OP_{1\text{dB}} - N_{\text{o}}$$

The linear dynamic range of a receiver is defined in terms of sensitivity signal level. This definition is more appropriate for a receiver system rather than an individual component, as it depends on factors external to the component itself, such as the type of modulation used, the recommended system $SNR_{\min}$, effects of error–correcting coding, and related factors. In dB, this can be written in terms of input powers as

$$\mathrm{LDR}\left[\mathrm{dB}\right] = IP_{1\mathrm{dB}} - S_{min}$$

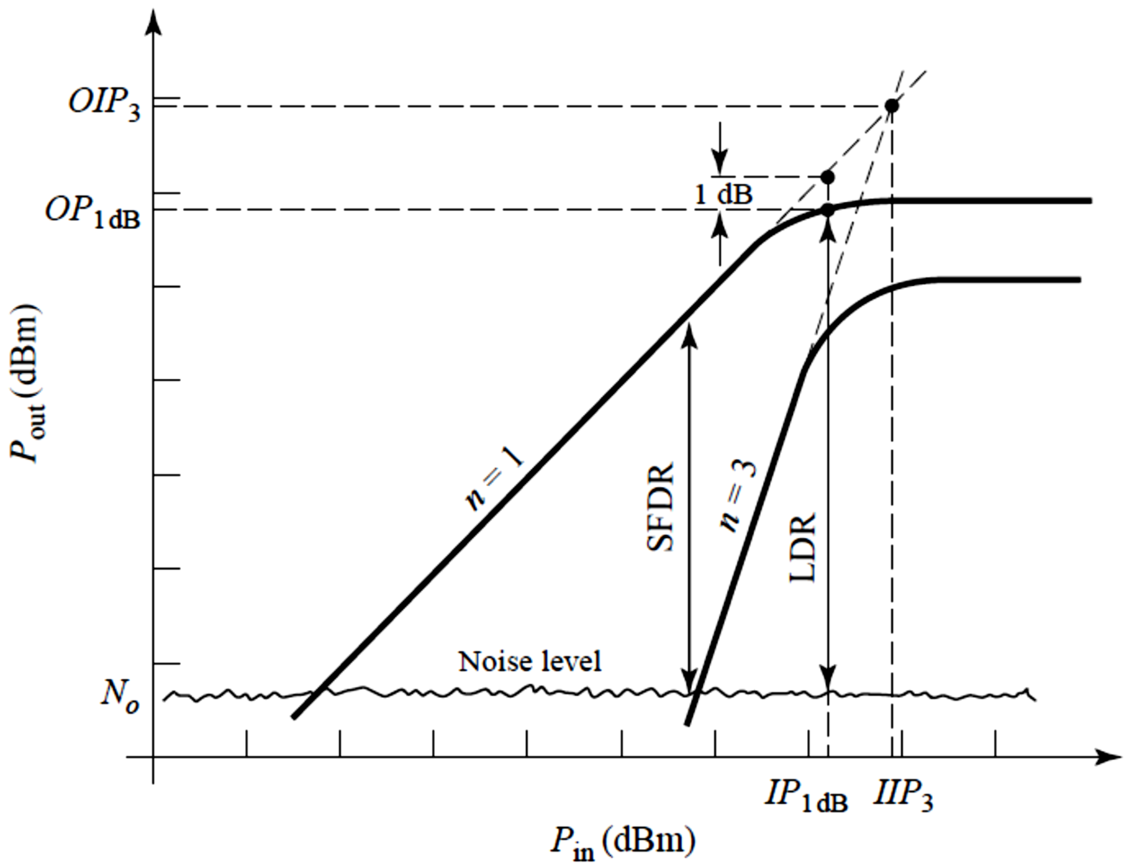


Figure 9: linear dynamic range (LDR) and spurious free dynamic range (SFDR).

Example 4: Receiver dynamic range

A receiver operating at room temperature has a noise figure of 5.5 dB and a bandwidth of 2 GHz. The input 1−dB compression point is 10 dBm. Calculate the sensitivity and dynamic range for $SNR_{min} = 3\text{dB}$.

Solution

$$S_{min} = -174 + NF + 10\ \log B + SNR_{min}$$

$$S_{min} = -174 + 5.5 + 10\ \log(2 \times 10^9) + 3$$

$$S_{min} = -174 + 5.5 + 93 + 3 = -72.5\ \text{dBm}$$

$$\text{LDR} = IP_{1\text{dB}} - S_{min}$$

$$\text{LDR} = 10 - (-72.5) = 82.5\ \text{dB}$$

Third−Order Intercept Point and Intermodulation

For a single input frequency, or tone, ω_0, the output will in general consist of harmonics of the input frequency of the form $n\omega_0$, for $n = 0, 1, 2, \ldots$. Often these harmonics lie outside the passband of the amplifier and so do not interfere with the desired signal at frequency ω_0. **The situation is different, however, when the input signal consists of two closely spaced frequencies.**

Intermodulation (IM) products

When two signals, or tones, at frequencies f_1 and f_2 are applied to a nonlinear device, they generate Intermodulation (IM) products according to $mf_1 + nf_2$ with $, n = 0, \pm1, \pm2, \cdots$. The order of a given product is defined as $|m| + |n|$. These IM products may be the second−order $f_1 \pm f_2$ products, third−order $2f_1 \pm f_2$, $2f_2 \pm f_1$ products, and so on.

The two–tone third–order IM products are of primary interest since they tend to have frequencies that are within the passband of the first IF stage. Consider a mixer or receiver as shown in Figure 10, where $f_{IF1}$and f_{IF2} are the desired IF outputs. In addition, the third–order IM (IM3) products f_{IM1} and f_{IM2} also appear at the output port. IM3 products are generated from f_1 and f_2 mixing with one another and then beating with the mixer's LO according to the expressions:

$$(2f_1 - f_2) - f_{LO} = f_{IM1}$$

$$(2f_2 - f_1) - f_{LO} = f_{IM2}$$

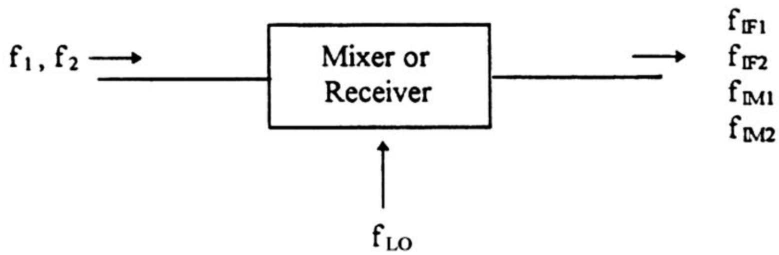


Figure 10: Signals generated from two RF signals by a nonlinear device

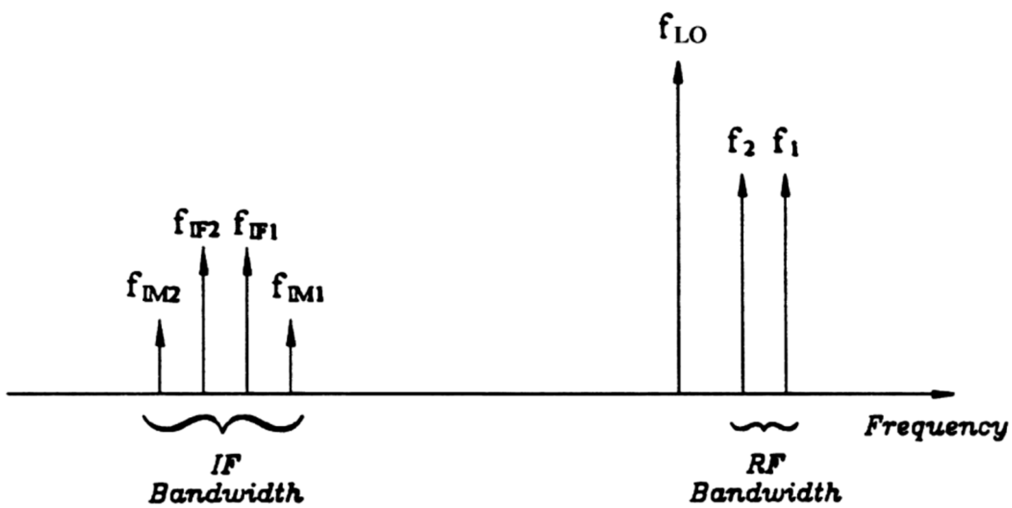


Figure 11: IM3 products in the IF bandwidth

where $f_{IM1}$and f_{IM2} are shown in Figure 11 with IF products for $f_{IF1}$and f_{IF2} generated by the mixer or receiver:

$$f_1 - f_{LO} = f_{IF1}$$

$$f_2 - f_{LO} = f_{IF2}$$

Note that the frequency separation in Figure 11 is

$$\Delta = f_1 - f_2 = f_{IM1} - f_{IF1} = f_{IF2} - f_{IM2} = f_{IF1} - f_{IF2}$$

These IM products are usually of primary interest because of their relatively large magnitude and because they are difficult to filter from the desired mixer outputs ($f_{IF1}$and f_{IF2}) if Δ is small.

Third-Order Intercept Point IP3

The intercept point, measured in dBm, is a figure of merit for intermodulation product suppression. A high intercept point indicates a high suppression of undesired IM products. The third-order intercept point (IP_3) is the theoretical point where the desired signal and the third-order distortion have equal magnitudes. The IP_3 is an important measure of the system's linearity.

In Figure 9, the output power of the first-order, or linear, product is proportional to the input power, and so the line describing this response has a slope of unity (before the onset of compression). The line describing the response of the third-order products has a slope of 3. (The second-order products would have a slope of 2, but since these products are generally not in the passband of the component, we have not plotted their response.)

In the linear region, for the IF signals, the output power is increased by 1 dB if the input power is increased by 1 dB. The IM3 products are increased by 3 dB for a 1–dB increase in P_{in}. The slope of the curve for the IM3 products is $3{:}1$.

Both the linear and third–order responses will exhibit compression at high input powers, so we show the extension of their idealized responses with dotted lines. Since these two lines have different slopes, they will intersect, typically at a point above the onset of compression, as shown in the figure. This hypothetical intersection point where the first–order and third–order powers would be equal is called the third–order intercept point, denoted as IP_3.

It may be specified as either an input power level (IIP_3), or an output power level (OIP_3). The relation between an intercept point referenced at the input versus the output is simply:

$$OIP_3 = G \times \ IIP_3$$

Or

$$OIP_3[\mathrm{dBm}] = IIP_3[\mathrm{dBm}] + G[\mathrm{dB}]$$

As with the 1 dB compression point, the reference for IP_3 is typically chosen to result in the largest value, so IP_3 is usually referenced at the output for amplifiers and at the input for mixers.

As depicted in Figure 9, IP_3 generally occurs at a higher power level than $P_{1\mathrm{dB}}$, the 1 dB compression point. Many practical components follow the approximate rule that IP_3 is 10–15 dB greater than $P_{1\mathrm{dB}}$, assuming these powers are referenced at the same point.

Intercept Point of a Cascaded System

For a cascaded system, the following procedure can be used to calculate the overall system intercept point:

1. Transfer all input intercept points to system input, subtracting gains and adding losses decibel for decibel.
2. Convert intercept points to powers (dBm to milliwatts). We have $(IIP_3)_1$, $(IIP_3)_2$, . . . $(IIP_3)_N$ for N cascaded components.
3. Assuming all input intercept points are independent and uncorrelated, add powers in ‘‘parallel’’ to obtain the input IIP_3 of the cascaded system:

$$\frac{1}{IIP_3} = \frac{1}{(IIP_3)_1} + \frac{1}{(IIP_3)_2} + \cdots + \frac{1}{(IIP_3)_N}$$

4. Convert IIP_3 from power (milliwatts) to dBm.

Note: The $IP_3 \rightarrow \infty$ for a filter (passive component).

Example 5: Third−order intermodulation

When two tons of $-10\ \mathrm{dBm}$ power level are applied to an amplifier, the level of the IM3 is $-50\ \mathrm{dBm}$. The amplifier has a gain of $10\ \mathrm{dB}$. Calculate the IM3 output power when the power level of the two−tone is $-20\ \mathrm{dBm}$. Also, indicate the IM3 power as decibels down from the wanted signal.

Solution

As shown in Figure 13, if the power level of the two−tone is $P_{in} = -20\ \mathrm{dBm}$, then the IM3 output power is:

$$IM3 = -50\ \mathrm{dBm} + 3 \times [-20\ \mathrm{dBm} - (-10\ \mathrm{dBm})]$$

$$IM3 = -50\ \text{dBm} - 30\ \text{dBm} = -80\ \text{dBm}$$

Then, wanted signal at $P_{in} = -20\ \text{dBm}$ has a power level at the output of the amplifier with gain of $10\ \text{dB}$:

$$P_{out} = P_{in} + G = -20\ \text{dBm} + 10\ \text{dB} = -10\ \text{dBm}$$

Difference between wanted signal and IM3 is

$$\Delta = -10\ \text{dBm} - (-80\ \text{dBm}) = 70\ \text{dB}$$

IM3 is $70\ \text{dB}$ down from the wanted signal.

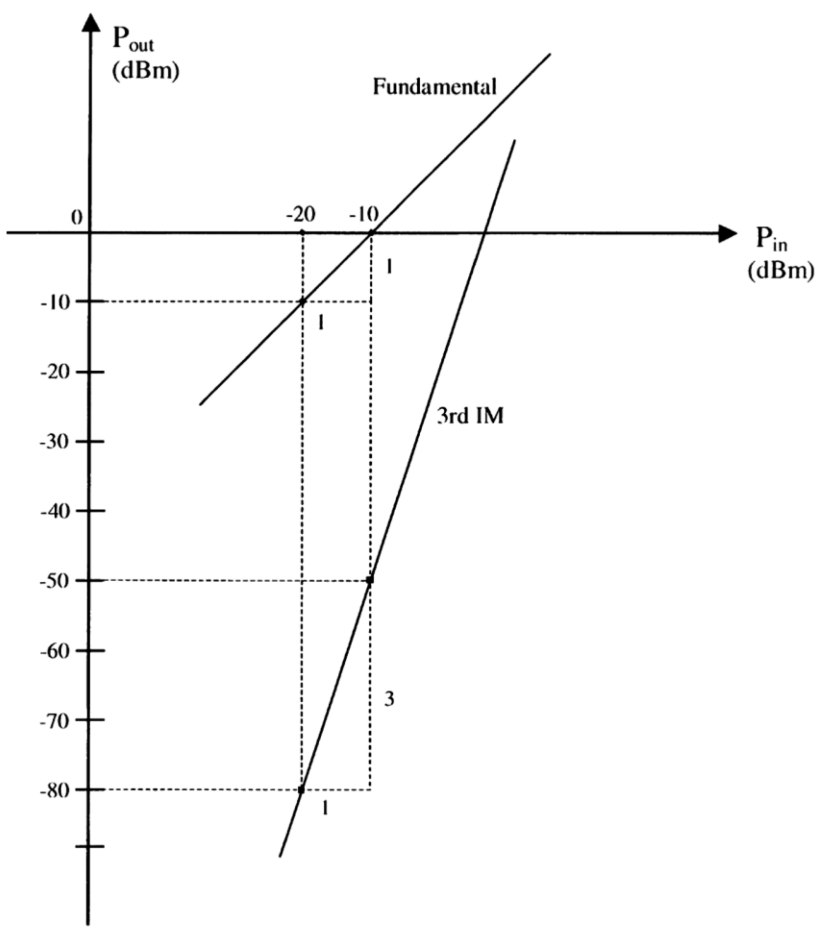


Figure 12: Third−order intermodulation

Example 6: Intercept Point of a Receiver System

A wireless receiver is shown in Figure 13. Calculate the overall IIP_3 in dBm.

Solution

Transfer all intercept points to system input; the results are shown in Figure 13. The overall IIP_3 is given by

$$\frac{1}{IIP_3} = \frac{1}{(IIP_3)_1} + \frac{1}{(IIP_3)_2} + \frac{1}{(IIP_3)_3} + \frac{1}{(IIP_3)_4} + \frac{1}{(IIP_3)_5}$$

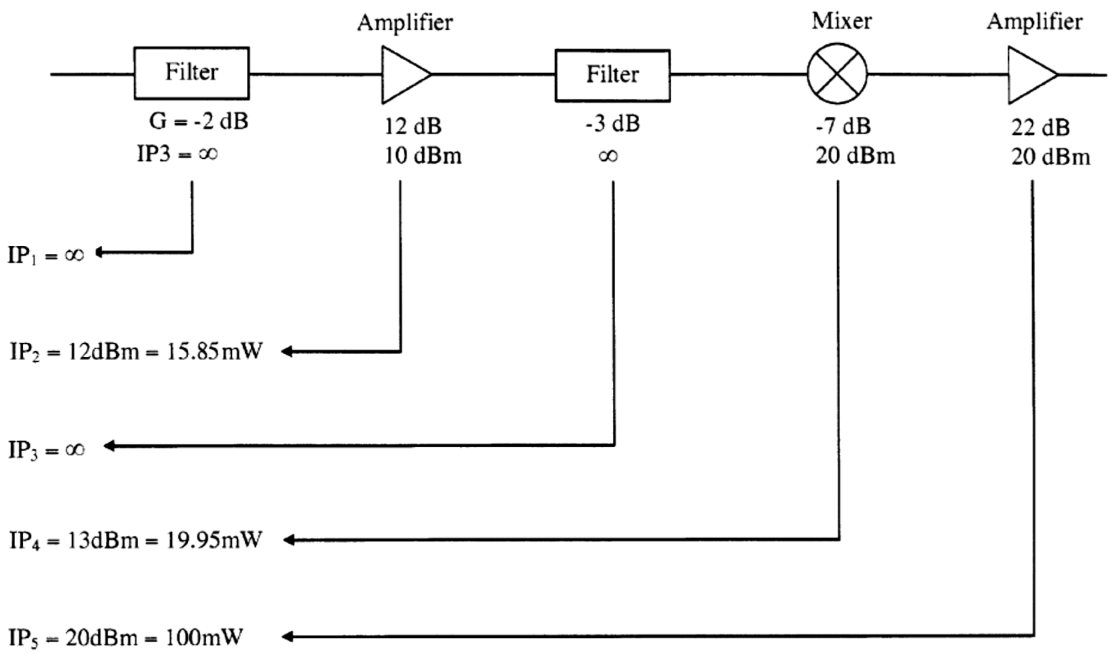


Figure 13: Receiver and its input intercept point

First and third stages are filters, then

$$(IIP_3)_1, (IIP_3)_3 \rightarrow \infty$$

Second stage is an amplifier with $IIP_3 = 10 \text{ dBm}$; transfer it to the system input gives:

$$(IIP_3)_2 = 10 \text{ dBm} - (-2 \text{ dB}) = 12 \text{ dBm} = 10^{1.2} = 15.85 \text{ mW}$$

Second stage is an amplifier with $IIP_3 = 10 \text{ dBm}$; transfer it to the system input gives:

$$(IIP_3)_2 = 10 \text{ dBm} - (-2 \text{ dB}) = 12 \text{ dBm} = 10^{1.2} = 15.85 \text{ mW}$$

Transfer the mixer $IIP_3 = 20 \text{ dBm}$ to the system input gives:

$$(IIP_3)_4 = 20 \text{ dBm} - (-3 \text{ dB}) - (12 \text{ dB}) - (-2 \text{ dB}) = 13 \text{ dBm} = 10^{1.3} = 20 \text{ mW}$$

Transfer the final amplifier with $IIP_3 = 20 \text{ dBm}$ to the system input gives:

$$(IIP_3)_5 = 20 \text{ dBm} - (-7 \text{ dB}) - (-3 \text{ dB}) - (12 \text{ dB}) - (-2 \text{ dB}) = 20 \text{ dBm}$$

$$(IIP_3)_5 = 10^2 = 100 \text{ mW}$$

Then, the overall IIP_3 is given by

$$\frac{1}{IIP_3} = \frac{1}{15.85} + \frac{1}{20} + \frac{1}{100} = \frac{1}{8.124} \rightarrow IIP_3 = 8.124 = 9.10 \text{ dBm}$$

Spurious Free Dynamic Range SFDR

The spurious free dynamic range of a nonlinear device is defined as the maximum output signal power for which the power of IM3 product is equal to the noise level of the component, divided by the output noise level. This situation is shown in Figure 9. Then the spurious free dynamic range can be expressed as

$$SFDR = \left(\frac{OIP_3}{N_o}\right)^{2/3}$$

In dB,

$$SFDR\ [\mathrm{dB}] = \frac{2}{3}(OIP_3\ [\mathrm{dBm}] - N_o\ [\mathrm{dBm}])$$

In a receiver it may be required to have a certain sensitivity, or minimum SNR, in order to achieve a specified performance level. This requires an increase in the input signal level, resulting in a corresponding decrease in dynamic range, since the spurious power level is still equal to the noise power. In this case, the spurious free dynamic range would be modified as:

$$SFDR\ [\mathrm{dB}] = \frac{2}{3}(IIP_3\ [\mathrm{dBm}] - S_{min}\ [\mathrm{dBm}])$$

$$SFDR\ [\mathrm{dB}] = \frac{2}{3}\left(OIP_3\ [\mathrm{dBm}] - (N_o\ [\mathrm{dBm}] + SNR_{min}\ [\mathrm{dB}])\right)$$

Example 7: Receiver spurious free dynamic range

A receiver has a noise figure of 7 dB, a 1 dB compression point of 25 dBm (referenced to output), a gain of 40 dB, and a third-order intercept point of 35 dBm (referenced to output). If the receiver is fed with an antenna having a noise temperature of $T_A = 150\ \mathrm{K}$, and the desired output SNR is 10 dB, find the linear and spurious free dynamic ranges. Assume a receiver bandwidth of 100 MHz.

Solution

The noise power at the receiver output can be calculated using noise temperatures as

$$N_o = GkB[T_A + (F-1)T_0] = 10^4(1.38 \times 10^{-23})(10^8)[150 + (4.01)(290)]$$

$$N_o = 1.8 \times 10^{-8}\ \mathrm{W} = 1.8 \times 10^{-5}\ \mathrm{mW} = -47.45\ \mathrm{dBm}$$

The linear dynamic range is, in dB:

$$LDR\ [\text{dB}] = OP_{1\text{dB}}\ [\text{dBm}] - N_o\ [\text{dBm}]$$

$$LDR = 25\ \text{dBm} + 47.45\ \text{dBm} = 72.45\ \text{dB}$$

The spurious free dynamic range is, in dB,

$$SFDR\ [\text{dB}] = \frac{2}{3}\big(OIP_3\ [\text{dBm}] - (N_o\ [\text{dBm}] + SNR_{min}\ [\text{dB}])\big)$$

$$SFDR = \frac{2}{3}\big(35\ \text{dBm} - (47.45\ \text{dBm} + 10\ \text{dB})\big) = 48.3\ \text{dB}$$

Observe that $SFDR \ll LDR$.

Main references:

David Pozar, “Microwave and RF Design of Wireless Systems”, John Wiley & Sons, 2001.

Kai Chang, “RF and Microwave Wireless Systems”, John Wiley & Sons, 2000.

Test

10 marks for each question, Successful test mark is $70/110$

Choose the correct answer.

1. The most important basic type of noise is

    a. Thermal noise
    b. Shot noise
    c. Flicker noise
    d. White noise

2. Thermal noise is also known as

    a. Johnson noise
    b. Solar noise
    c. Flicker noise
    d. Partition noise

3. Low frequency noise is

    a. Transit time noise
    b. Shot noise
    c. Flicker noise
    d. None of the above

4. For a noise power given by $P_N = kTB$:

    a. Systems with smaller bandwidth $\boldsymbol{B}$ collect less noise power
    b. Cooler systems generate more noise power
    c. This type of noise isn't thermal noise
    d. This type of noise isn't white noise

5. Noise power at the resistor is affected by the value of the resistor as

a. Directly proportional to the value of the resistor
b. Inversely proportional to the value of the resistor
c. Unaffected by the value of the resistor
d. Becomes half as the resistance value is doubled

6. A noise factor $F > 1$ means that

a. The system is noisy
b. The system is noiseless
c. The output signal power is less than the output noise power
d. The input signal power is less than the input noise power

7. Which of the following can cause frequency intermodulation products in a system?

a. Active components like diodes and transistors (amplifiers)
b. Two tones at the input of a nonlinear device
c. Spurious frequency components due to device nonlinearities,
d. All of the above

8. What is the spurious−free dynamic range of a system with IIP3 = +30 dBm and a sensitivity level of −90 dBm?

a. 80 dB
b. 90 dB
c. 30 dB
d. 120 dB

9. What is the first harmonic of 1 GHz?

a. 1 GHz
b. 2 GHz
c. 3 GHz
d. 4 GHz

10. What happens to the noise figure of a receiver when a 10 dB attenuator is added at the input?

a. Noise figure increases by 10 dB
b. Noise figure decreases by 10 dB
c. Noise figure decreases by 1 dB
d. Noise figure doesn't change

11. An RF system has a linear gain of $+10$ dB and an output 3rd−order intercept point (OIP3) of $+30$ dBm. What is the input 3rd−order intercept point (IIP3)?

a. $+40$ dBm
b. $+30$ dBm
c. $+20$ dBm
d. $+10$ dBm

Question number	Answer
1	a
2	a
3	c
4	a
5	c
6	a
7	d
8	a
9	a
10	a
11	c

Chapter Four

Microwave Amplifiers: Technologies and Specifications

Keywords

Operating power gain, Available power gain, Transducer power gain, Power Amplifier (PA), Gain Block Amplifier, Driver Amplifier, Low Noise Amplifier (LNA), Variable Gain Amplifies (VGA), Surface–mount technology (SMT) Surface–Mount Device (SMD), Monolithic Microwave Integrated Circuits (MMICs).

Abstract

In this chapter we aim at understanding the basic specifications of microwave amplifiers, which can be obtained from the datasheets produced by manufacturers. Therefore, students will be able to read datasheets, extract main and detailed specifications in terms of frequency and operating point, and to choose the suitable amplifier for the relevant application.

Learning Objectives

In this chapter, student will be able to:

- Read datasheets of various types of amplifiers needed in a wireless system,
- Extract main and detailed specifications of amplifiers from their datasheets in terms of frequency, operating point, and other parameters
- Choose the suitable amplifier for the relevant application to achieve the performance, current consumption, cost, and size as targeted.

Introduction

Microwave amplifiers, oscillators and mixers are basic and prevalent circuits in modern RF and microwave systems. Due to the dramatic improvements and innovations in solid–state technology that have occurred, most RF and microwave components today use transistor devices such as Si BJTs, GaAs or SiGe HBTs, Si MOSFETs, GaAs MESFETs, or GaAs or GaN HEMTs, and can be easily integrated in both hybrid and Monolithic Microwave Integrated Circuits (MMICs).

Transistor amplifiers can be used at frequencies in excess of 100 GHz in a wide range of applications requiring small size, low noise figure, broad bandwidth, and medium to high power capacity.

A solid–state oscillator uses an active nonlinear device, such as a diode or transistor, in conjunction with a passive circuit to convert DC to a sinusoidal steady–state RF signal. Basic transistor oscillator circuits can generally be used at low frequencies, often with crystal resonators to provide improved frequency stability and low noise performance. At higher frequencies, diodes or transistors biased to a negative resistance operating point can be used with cavity, transmission line, or dielectric resonators to produce fundamental frequency oscillations up to 100 GHz.

A mixer is a three–port device that uses a nonlinear active device, such as a diode or a transistor, to achieve frequency conversion. Modern RF and microwave systems typically use several mixers and filters to perform the functions of frequency up–conversion and down–conversion between baseband signal frequencies and RF carrier frequencies.

The receiver in a modern RF and microwave system is formed from amplifiers, oscillators and mixers, which are like building blocks, and it is apparent that the overall performance of the receiver depends on the characteristics of each individual building block. The objectives of the receiver system design are not only to make the designed receiver achieving the performance, current consumption, cost, and size as targeted, but also to define the specifications of each individual device in the receiver and thus to ensure the targeted receiver performance and other goals achieved.

Figure 1 shows a simplified block diagram of such a receiver, where the basic blocks are filters, amplifiers, mixers and oscillators.

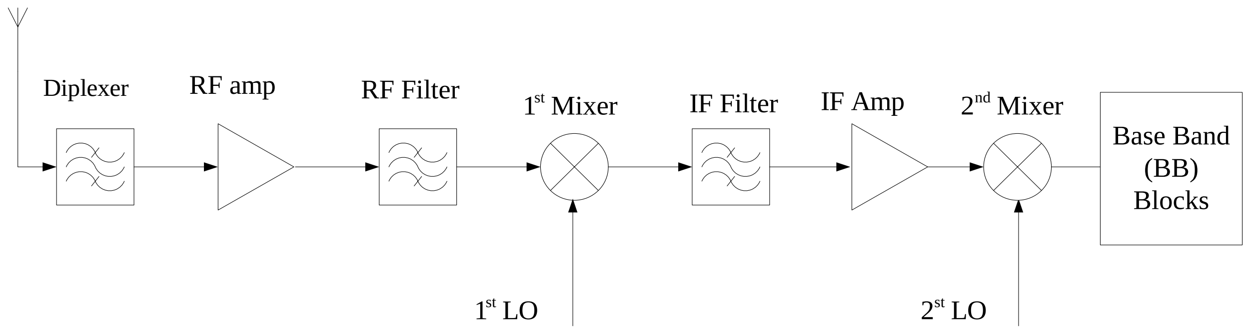


Figure 1: simplified block diagram of a receiver in a modern RF and microwave system.

The wireless systems use full−duplex transceivers, where the transmitter and the receiver simultaneously operate at offset frequencies. The configuration of a full duplex transceiver is usually more complicated as depicted in Figure 2. In this figure, various types of amplifiers, oscillators and mixers are present.

It is essential for a receiver system designer to understand the basic specifications of various function devices consisting of an RF receiver and to know their possible performance based on the present technology. These devices are the building blocks of the receiver, and proper selecting and specifying them are the major tasks of the receiver system design after the receiver architecture has been defined.

The main objectives of this chapter and the next chapter are to understand the basic specifications of amplifiers, oscillators and mixers, which can be obtained from the datasheets of these components. Therefore, students will be able to read datasheets and to choose the suitable amplifier, oscillator or mixer for the relevant application.

We will begin with some general definitions of two−port power gains that are useful for amplifiers.

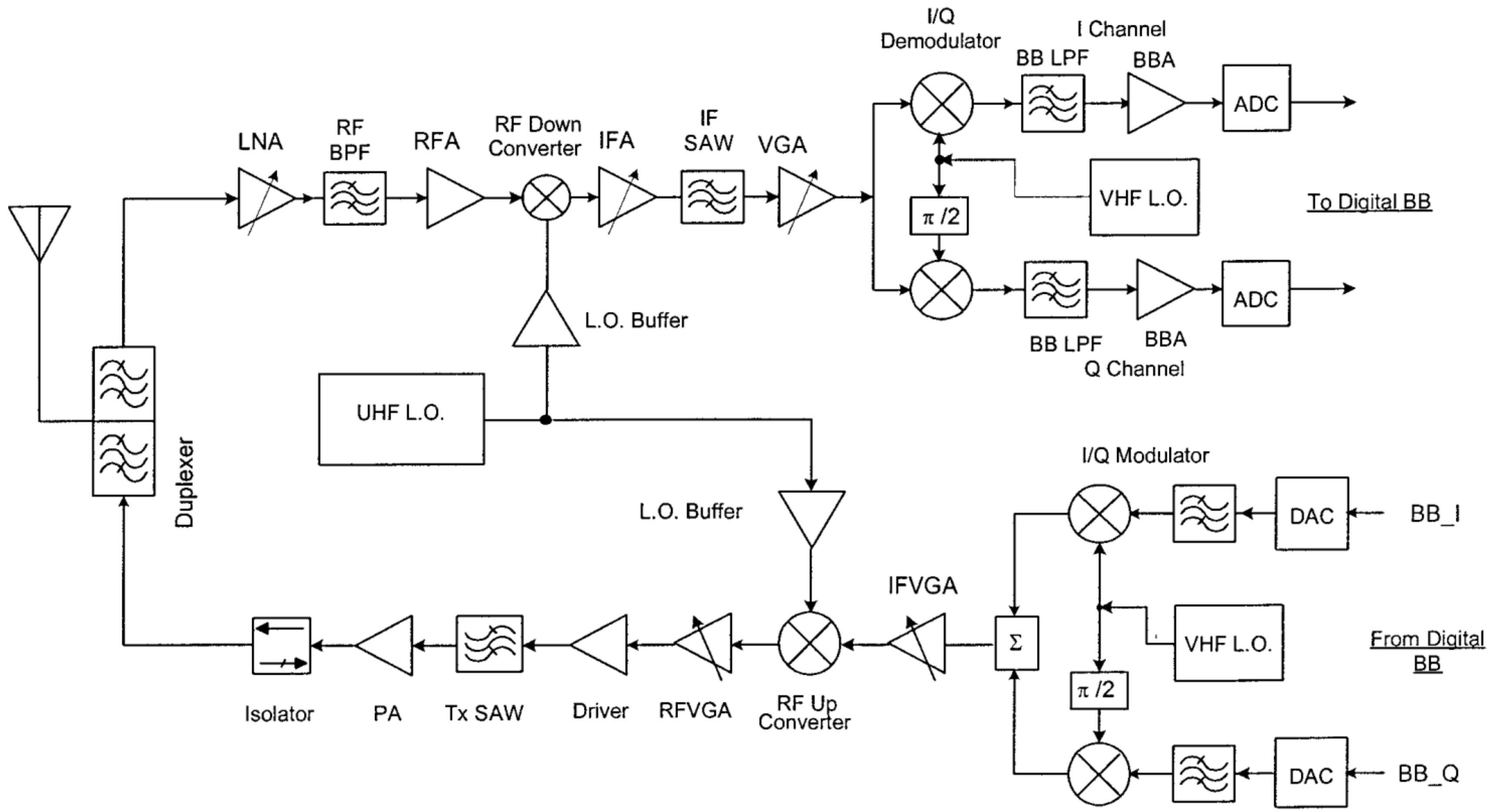


Figure 2. Block diagram of a superheterodyne full−duplxer transceiver

Definition of Two−Port Power Gains

Referring to a generic two−port network circuit as shown in figure 3, there are three commonly used definitions of power gains.

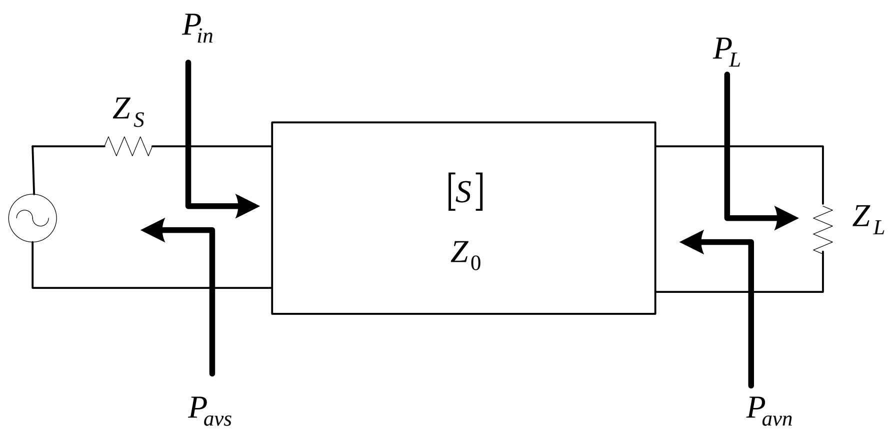


Figure 3: Definition of Two−Port Power Gains

Operating power gain

This is the ratio of the average power dissipated in a load, P_L, to the average power delivered to the network, P_{in}.

$$G_p = \frac{P_L}{P_{in}}$$

Available power gain

This is the ratio of the average power available from the two−port network, P_{avn}, to the average power available from the source, P_{avs}. This assumes conjugate matching of both the source ($Z_{in} = Z_S^*$) and the load ($Z_{out} = Z_L^*$).

$$G_A = \frac{P_{avn}}{P_{avs}}$$

Transducer power gain

This is the ratio of the average power delivered to the load, P_L, to the average power available from the source, P_{avs}.

$$G_T = \frac{P_L}{P_{avs}}$$

These three definitions of power gain are used for different types of amplifiers:

- Operating power gain is used for maximum linear output power amplifiers.
- Available power gain is used for low noise amplifiers (LNA).
- Transducer power gain is used for maximum linear gain amplifiers.

Scattering Parameters of Amplifiers

The S parameters of microwave amplifiers are usually available for the amplifier in chip and packaged form. Amplifiers in chip form are used when the best performance in gain, bandwidth, and noise is desired, but this form needs a special technology to implement it on printed circuit board (PCB). Packaged amplifiers are very popular because they come in sealed enclosures and are easy to work with. The parasitic elements introduced by the package produce a degradation in the amplifier AC performance.

Manufacturers usually measure and provide S parameters of microwave amplifiers as function of frequency at a given DC bias. Since the minimum noise figure, linear output power, and maximum gain require different DC bias settings, the manufacturers usually provide several sets of S parameters.

Manufacturers usually provide S parameters of microwave amplifiers in S2P text file, and it can be used directly by CAD software like AWR© Microwave Office and Advanced Design System (ADS) from Keysight Technologies. The header of the file is in the form:

MHZ S MA R 50

This means that the first column is the frequency and it is given in MHz (it can be in GHz), other columns are S parameters given in Magnitude (in numerical values and sometimes in dB) and Angle (in degrees). S parameters are given in the following order: $S_{11}, S_{21}, S_{12}, S_{22}$. Figure 4 shows an example of S parameters for MGA−621P8 LNA.

MGA-621P8
Data Sheet

MGA-621P8 Typical Scattering Parameters, Vdd = 4V, Idd = 65 mA

MGA-621P8 Typical Scattering Parameters, Vdd = 4V, Idd = 65 mA

freq GHz	S11		S21			S12		S22	
	Mag.	Ang.	dB	Mag.	Ang.	Mag.	Ang.	Mag.	Ang.
0.1	0.30	−54.42	25.77	19.43	157.88	0.03	19.85	0.13	-147.21
0.2	0.23	−75.20	24.58	16.95	148.21	0.04	22.58	0.13	178.18
0.3	0.20	−90.13	23.63	15.19	140.01	0.04	26.09	0.13	152.37
0.4	0.19	−101.37	22.79	13.79	132.53	0.04	29.85	0.12	133.25
0.5	0.18	−110.56	21.99	12.58	125.66	0.05	34.17	0.11	118.14
0.6	0.18	−118.55	21.20	11.48	119.41	0.05	36.86	0.10	103.00
0.7	0.18	−125.27	20.44	10.52	113.82	0.05	39.57	0.09	88.54
0.8	0.18	−131.26	19.71	9.68	108.74	0.06	41.79	0.08	74.16
0.9	0.18	−136.18	19.02	8.93	104.10	0.06	43.41	0.07	60.99
1.0	0.19	−142.93	18.36	8.28	99.87	0.07	44.47	0.07	47.62
1.1	0.19	−147.17	17.72	7.69	95.94	0.07	45.39	0.07	36.37
1.2	0.19	−150.80	17.13	7.18	92.25	0.08	46.01	0.06	22.69
1.3	0.19	−154.37	16.56	6.73	88.81	0.08	46.12	0.06	10.09
1.4	0.19	−157.18	16.02	6.32	85.52	0.09	45.95	0.06	−3.29
1.5	0.19	−159.86	15.50	5.95	82.41	0.09	45.81	0.06	−14.70
2.0	0.18	−170.21	13.24	4.59	68.66	0.12	43.37	0.08	−54.98
3.0	0.15	−173.07	9.90	3.12	46.26	0.17	34.49	0.13	−82.88
4.0	0.13	−160.54	7.50	2.37	27.07	0.21	23.39	0.17	−92.71
5.0	0.15	−161.16	5.66	1.92	8.67	0.26	10.75	0.17	−115.74
6.0	0.16	−168.32	4.15	1.61	−9.33	0.30	−3.13	0.18	−148.55
7.0	0.18	−169.02	2.81	1.38	−26.77	0.34	−17.35	0.23	−171.42
8.0	0.22	−172.11	1.54	1.19	−44.09	0.37	−32.48	0.27	−179.22
9.0	0.28	−179.13	0.40	1.05	−56.93	0.39	−44.39	0.34	172.55
10.0	0.36	167.24	−0.65	0.93	−72.87	0.41	−59.89	0.32	165.25
11.0	0.40	150.74	−1.67	0.83	−88.88	0.43	−76.41	0.36	143.82
12.0	0.42	117.93	−3.74	0.65	−106.42	0.38	−95.94	0.45	119.98
13.0	0.48	75.82	−7.44	0.43	−106.46	0.26	−99.26	0.65	96.63
14.0	0.36	46.19	−6.28	0.49	−88.91	0.31	−76.27	0.74	78.06
15.0	0.12	54.36	−3.61	0.66	−103.00	0.49	−90.86	0.65	60.03
16.0	0.14	105.64	−2.80	0.72	−126.00	0.59	−116.07	0.55	29.26
17.0	0.16	99.88	−3.19	0.69	−147.98	0.60	−140.52	0.59	−2.26
18.0	0.20	87.16	−4.00	0.63	−165.87	0.58	−161.06	0.65	−14.54
19.0	0.40	52.83	−4.69	0.58	−171.24	0.55	−169.34	0.53	−10.12
20.0	0.60	41.84	−5.39	0.54	176.07	0.52	175.04	0.57	10.07

Figure 4: An example of S parameters for MGA−621P8 LNA.

Remember that:

S_{11} and S_{22} account for matching of the amplifier at the input and the output respectively. For good matching $|S_{11}|$ and $|S_{22}|$ must be less than 0.2, this means that $RL \geq 14\ \mathrm{dB}$ or $VSWR \leq 1.5$.

S_{21} accounts for the gain of the amplifier, $|S_{21}|$ must be as greater as possible than 1.

S_{12} accounts for reverse isolation of the amplifier, and it must be $|S_{12}| \approx 0$.

Categories of Amplifiers

When searching for an amplifier, manufacturers classify them in main categories similar to their functions in RF and microwave systems. In this section we will take examples from two manufacturers: **Analog Devices** (www.analog.com) and **Broadcom Limited** (www.broadcom.com).

Analog Devices (www.analog.com) offers fixed–gain and variable–gain RF amplifiers can be used in applications covering DC to RF, microwave and millimeter wave bands. ADI's industry–leading designs make optimal use of many different MMIC technologies to cover a wide range of bandwidths and power levels. RF amplifiers are fully specified over frequency, temperature and supply voltage with offerings in both bare die and high performance surface–mount packages.

Categories: Driver Amplifiers, Fully Differential Amplifiers, Gain Blocks, Low Noise Amplifier, Power Amplifiers, RF Amplifiers Bias Controllers, Wideband Distributed Amplifiers, Variable Gain Amplifiers (analog and digital control VGAs).

Example from Analog Devices website:

To search for an amplifier from **Analog Devices**, Figure 5 shows the webpage products where all categories are listed. For RF and microwave applications, RF amplifiers contains several subcategories depending on the function of the amplifier as shown.

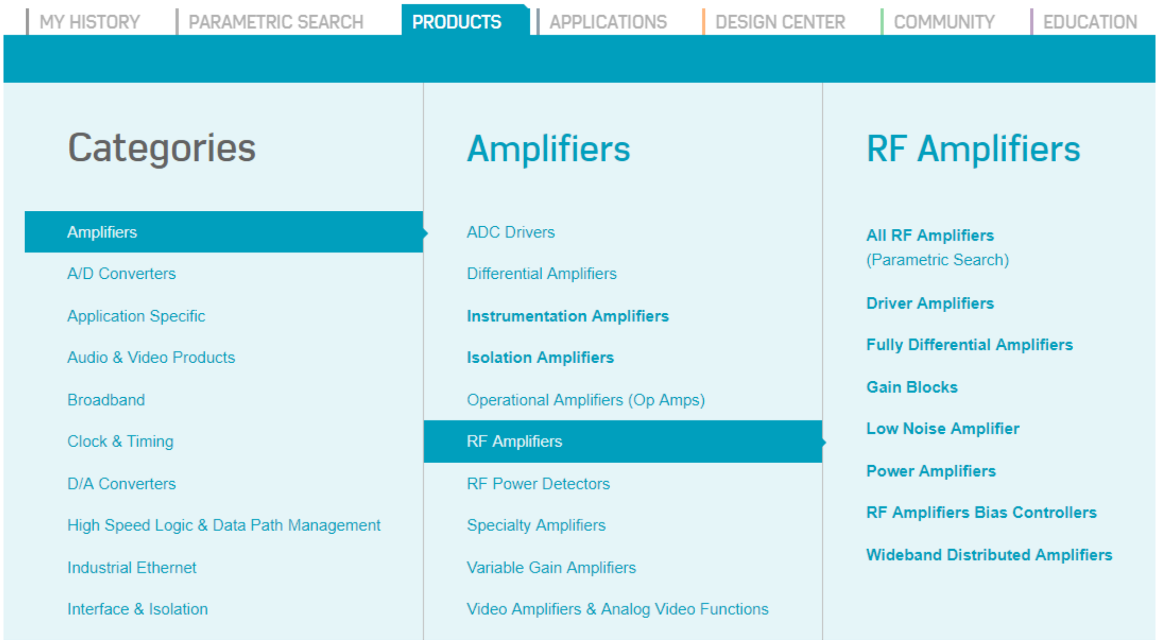


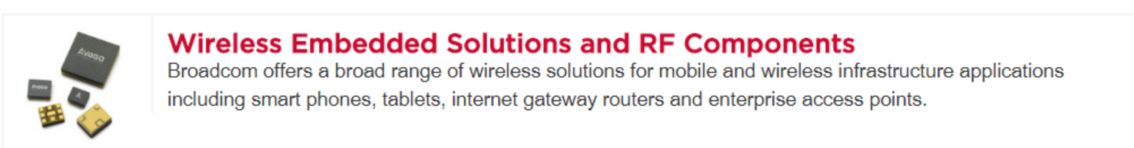
Figure 5: RF amplifier categories from Analog Devices.

Broadcom Limited (AVAGO Technologies) (www.broadcom.com) offers a broad selection of RF and microwave amplifiers including Power Amplifiers (PAs), Gain Blocks/Drivers, Low Noise Amplifiers (LNAs) and variable gain amplifiers (VGAs) for cellular applications and diverse markets such as wireless infrastructure, WiMAX, WLAN, test & measurement and other wireless microwave applications.

Example from Broadcom Limited website:

To search for an amplifier from **Broadcom Limited**, Figure 6 shows the webpage products where all categories are listed. For RF and microwave applications, RF amplifiers contains several subcategories depending on the function of the amplifier as shown.

From PRODUCTS we choose



Then we choose

The different categories of amplifiers are presented as in Figure 6.

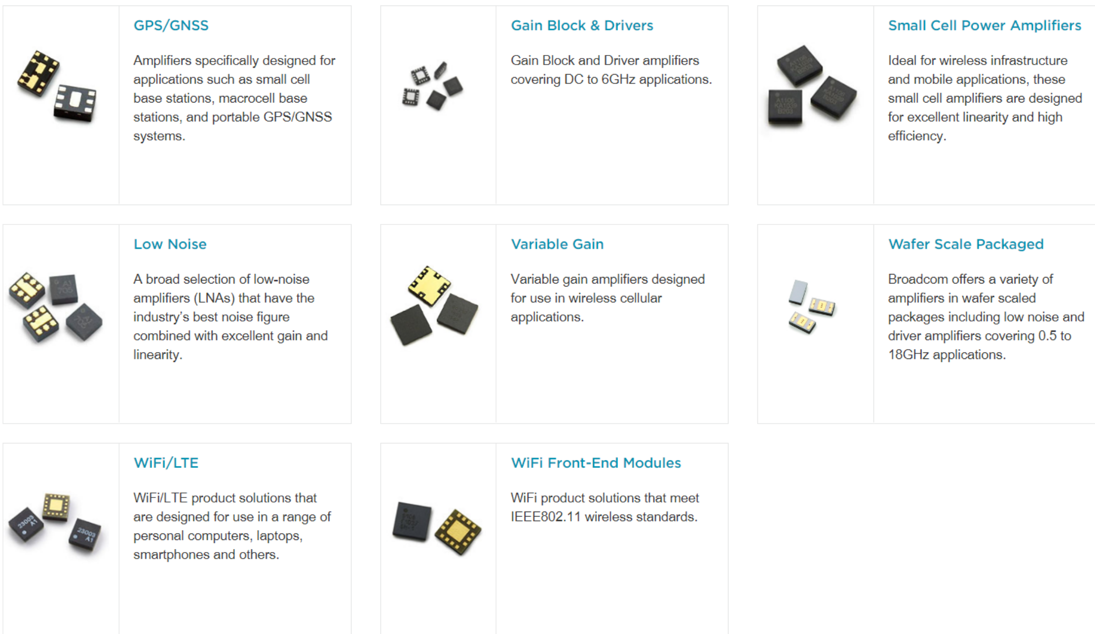


Figure 6: RF amplifier categories from Broadcom Limited.

Definition of amplifier categories

The low-noise amplifier (LNA) is the first amplifier used in the RF front-end of receiver systems after the antenna. It plays an important role in achieving good reception sensitivity by amplifying a very low-power signal without significantly degrading its signal-to-noise ratio. LNA reduces the effect of noise from subsequent stages of the receive chain by its gain that can be stepped-controlled to cope with the receiver dynamic range.

The gain-block amplifier may be utilized as second or third stage in the receive chain in the RF front-end receiver to further amplify the received RF signal.

The variable gain amplifier may be utilized as a gain-block amplifier with variable gain in the RF front-end receiver chain or in the IF/baseband chain.

The driver amplifier may be utilized in the transmit chain or in the receive chain. It is utilized in the final stage of the transmitter to further amplify the RF signal to a power level that is enough to drive the power amplifier (PA).

The power amplifier boosts the desired RF signal to a power level that is high enough to make the transmission power at antenna port being still greater than the minimum requirement.

Amplifier specifications

It is essential to understand the basic specifications of various amplifiers and to know their possible performance based on the present technology, and proper selecting and specifying them are the major tasks in practice.

The main specifications of an amplifier are as follows:

- Operating frequency band (MHz or GHz),
- Nominal gain (dB),
- Noise figure (dB),
- $P_{1\mathrm{dB}}$, (dBm),
- IP_3, (dBm),
- Input and output matching or return loss (dB).

We will explore two examples of low-noise amplifiers from the two manufacturers **Broadcom Limited** and **Analog Devices**.

Low-noise amplifier from Broadcom Limited

When we choose the category **Low Noise**, a table containing all low noise amplifiers appears and may be downloaded in Excel sheet form (see attached file "**Low-Noise-AVAGO.xlsx**"). Let's choose MGA-622P8 LNA and download datasheet (see attached datasheet "**MGA-622P8_LNA.pdf**"). This datasheet offers all information and specifications of the amplifier that help us to choose the more suitable one for our application. We will present these information and specifications and focus on the more important ones.

Description

Avago Technologies' MGA−622P8 is an economical, easy−to−use GaAs MMIC Low Noise Amplifier (LNA). The LNA has low noise and high linearity achieved through the use of Avago Technologies' proprietary $0.25\mu m$ GaAs Enhancement−mode pHEMT process.

We read here that this MMIC LNA is low price and can be implemented using conventional PCB and available technology. The technology GaAs E−pHEMT is usually used for low noise and high linearity (high IP_3) components.

It is housed in the miniature 2.0 mm × 2.0 mm × 0.75 mm 8−pin Dual−Flat−Non−Lead (DFN) package.

This information about the package of the amplifier is important for the implementation.

Applications

- Low noise amplifier for small cell base station application.
- Other low noise application.

Electrical Specifications

RF performance at T_A = 25°C, Z_0 = 50 Ω, Vdd = 4V, measured on demo board in Figure 6 with component list in Table 1 for 1.9 GHz.

Symbol	Parameter and Test Condition	Frequency (GHz)	Unit	Min	Typ	Max
Vdd	Device operating voltage		V	—	4.0	—
Idd	Device current		mA	50	61	75
Gain	Gain	1.9	dB	17.5	18.6	20.0
NF	Noise Figure	1.9	dB	—	0.56	0.9
OIP3[a]	Output Third order intercept point	1.9	dBm	32.0	35.0	—
OP1dB	Output Power at 1dB Gain Compression	1.9	dBm	—	20.4	—
IRL	Input Return Loss, 50 Ω source	1.9	dB	—	20.0	—
ORL	Output Return Loss, 50 Ω load	1.9	dB	—	12.0	—

a. 2-tone OIP3 test condition: FRF1, FRF2 = 1MHz separation with input power = −10dBm per tone.

This table contains of the LNA and the conditions of measurement. Figure 7 gives the PCB layout and the schematic, where the LNA MMIC and the external components are shown. Table 1 gives the values and types of these external components that are surface−mount device (SMD)[1]. Murata is famous for manufacturing SMD (or chip) capacitors and inductors.

From the electrical specifications, we can read the values of key specifications of MGA−622P8 LNA at 1.9 GHz as follows:

- Operating frequency band: $[1.5, 2.7]$ GHz.
- Nominal gain: $G_p = 18.6$ dB.
- Noise figure: $NF = 0.56$ dB.
- $OP_{1\text{dB}} = 20.4$ dBm,
- $OIP_3 = 35$ dBm,
- Input return loss: $IRL = 20$ dB.
- Output return loss: $ORL = 12$ dB.

From these electrical specifications, we conclude that this LNA has a very low NF, and high linearity because $OIP_3 - OP_{1\text{dB}} \approx 15$ dB.

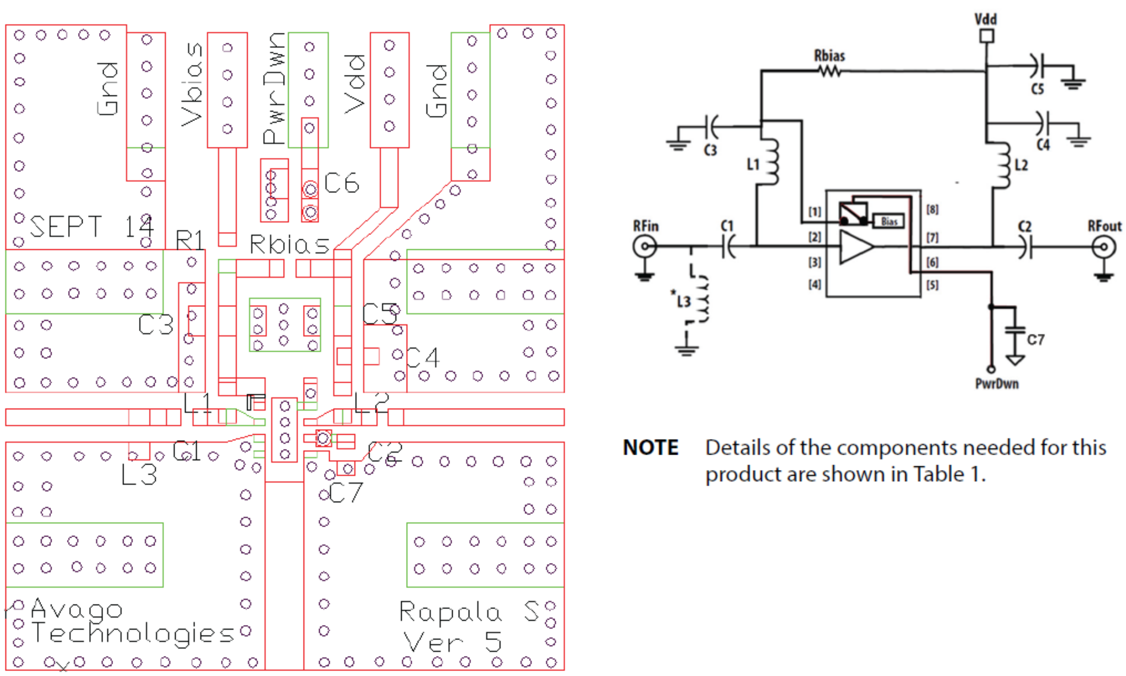


Figure 7: Demo board layout and schematic diagrams

[1] Surface-mount technology (SMT) is a method for producing electronic circuits in which the components are mounted or placed directly onto the surface of printed circuit boards (PCBs). An electronic device so made is called a surface-mount device (SMD).

Table 1 Component List for 1.5 GHz to 2.7 GHz Matching (Refer to the Application Note for Other Application Frequencies)

Part	Size	Value	Detail Part Number
C1, C2, C4	0402	100 pF (Murata)	GRM1555C1H101JA01D
C3, C5	0603	4.7 μF (Murata)	GRM188R60J475KE19D
C6	0402	Not used	
C7	0402	10 pF	GRM1555C1H100JA01D
L1	0402	5.6 nH (Murata)	LQP15MN5N6G00
L2	0402	6.8 nH (Murata)	LQP15MN6N8G00
L3	0402	Not used	
Rbias	0402	2.49 kΩ (Kamaya)	RMC1/16SK2491FTH

More detailed results are given in the datasheet in graphical form as function of frequency. This helps to examine the amplifier specifications at frequencies of interests.

S parameters are also measured over a wide range of frequencies at $V_{dd} = 4\text{ V}$, $I_{dd} = 60\text{ mA}$.

Low−noise amplifier from Analog Devices

We choose a LNA from **Analog Devices** from the same category and package as previous one to compare the key specifications. The chosen LNA is HMC374 (see attached datasheet "**hmc374_LNA.pdf**").

Here, electrical specifications are gives for three ranges of frequencies as shown below. Detailed results are also given in the datasheet in graphical form as function of frequency.

Electrical Specifications, T_A = +25° C, Vdd= +5V

Parameter	Min.	Typ.	Max.	Min.	Typ.	Max.	Min.	Typ.	Max.	Units
Frequency Range		0.3 - 1.0			1.0 - 2.0			2.0 - 3.0		GHz
Gain	12	15		10	13		6	9		dB
Gain Variation Over Temperature		0.01	0.02		0.01	0.02		0.01	0.02	dB/°C
Noise Figure		1.5	1.9		1.6	2.0		1.8	2.2	dB
Input Return Loss		5			8			13		dB
Output Return Loss		7			9			9		dB
Output 1 dB Compression (P1dB)		22			22			22		dBm
Saturated Output Power (Psat)		23			23			23		dBm
Output Third Order Intercept (IP3)		37			37			37		dBm
Supply Current (Idd) (Vdd = +5V)		90			90			90		mA

For the frequency range $[1,2]$ GHz, the key specifications are:

- Operating frequency band: $[1, 2]$ GHz.
- Nominal gain: $G_p = 13$ dB.
- Noise figure: $NF = 1.6$ dB.
- $OP_{1\mathrm{dB}} = 22$ dBm,
- $OIP_3 = 37$ dBm,
- Input return loss: $IRL = 8$ dB.
- Output return loss: $ORL = 9$ dB.

Note that the two LNAs have similar linearity, but other specifications of HMC374 are less than MGA−622P8. The most important is that the NF of HMC374 is higher than that of MGA−622P8. This is due to that the technology E−pHEMT is preferable for low noise applications.

Note: Better input and output matching could be obtained if input and output external network matching are added, like in the measurement circuit of MGA−622P8.

S parameters are given in s2p file named "**HMC374 deembedded.s2p**" (attached file). As mentioned previously, this file may be used directly by the simulation software.

Exercises

1. Search in the internet for manufacturers of microwave amplifiers other than Broadcom Limited and Analog Devices.
2. Choose different types of amplifiers from different manufacturers, download their datasheets and S parameters in form of s2p file, extract their main specifications, and compare them in terms of noise performance, linearity, gain, dynamic range and matching.

Test

10 marks for each question, Successful test mark is $60/100$

Choose the correct answer.

1. The gain of an amplifier decreases with frequency in general
   a. True
   b. False
2. For $|S_{11}| \leq 0.2$, the matching of the amplifier is
   a. Good
   b. A matching network should be added
   c. Not acceptable
   d. No information about matching
3. For $|S_{22}| > 0.2$, the matching of the amplifier is
   a. Good
   b. An input matching network should be added
   c. An output matching network should be added
   d. No information about matching
4. For $|S_{21}| \leq 0 \text{ dB}$, the amplifier is
   a. Useful for low−noise applications
   b. Used as an attenuator
   c. Useful for low power applications
   d. Linear
5. A high gain amplifier means that
   a. The output power is high
   b. The input signal level is highly increased
   c. The amplifier is very noisy
   d. The amplifier has a large dynamic range
6. An amplifier has a high OIP3 means that
   a. The output power of the amplifier is high
   b. The amplifier has high linearity

   c. The amplifier has high distortion
   d. The output noise of the amplifier is high

7. An amplifier has a high OP1dB means that
   a. The output power of the amplifier is high
   b. The amplifier has high linearity
   c. The amplifier has a large spurious−free dynamic range
   d. The gain of the amplifier is high

8. For $NF \leq 2\ \mathrm{dB}$, the amplifier is
   a. Useful for low−noise applications
   b. Used as a final stage amplifier in a receiver
   c. Useful for high power applications
   d. No information about noise

9. The gain−block amplifier may be utilized as second or third stage amplifier to further amplify the transmitted RF signal.
   a. True
   b. False

10. A LNA with high gain reduces more the effect of noise from subsequent stages of the receive chain.
    a. True
    b. False

Question number	Answer	Feedback
1	a	Revise scattering parameters of amplifiers
2	a	Revise scattering parameters of amplifiers
3	c	Revise scattering parameters of amplifiers
4	b	Revise scattering parameters of amplifiers
5	b	Revise definition of power gain
6	b	Revise definition of third order intercept point
7	a	Revise definition of 1dB compression point
8	a	Revise definition of noise figure
9	b	Revise definition of amplifier categories
10	a	Revise definition of amplifier categories

Chapter Five:
Microwave Oscillators and Mixers: Technologies and Specifications

Keywords

Voltage Controlled Oscillator (VCO), Microwave Frequency Synthesizer, Direct Digital Synthesis (DDS), Phase–Locked Loops (PLL), Temperature–Compensated Crystal Oscillator (TCXO), Oven–Controlled Crystal Oscillator (OCXO), Local Oscillator (LO), N Counter, Dual–Modulus Prescaler, Loop Filter, Phase Noise, Mixer, Frequency Conversion, Up–Conversion, Down–Conversion, Intermediate Frequency (IF), Image Frequency, Conversion Loss, Single–Ended Mixer, Balanced Mixer, Double–balanced mixer.

Abstract

In this chapter we aim at understanding the basic specifications of microwave oscillators and mixers, which can be obtained from the datasheets produced by manufacturers. Therefore, students will be able to read datasheets of different microwave oscillator and mixer technologies, extract main and detailed specifications in terms of frequency and power level, and to choose the suitable oscillator and mixer for the relevant application.

Learning Objectives

In this chapter, student will be able to:

- Read datasheets of various technologies of microwave oscillators and mixers needed in a wireless system,
- Extract main and detailed specifications of microwave oscillators and mixers from their datasheets in terms of frequency, power level, and other parameters,
- Choose the suitable microwave oscillator and mixer for the relevant application to achieve the performance, current consumption, cost, and size as targeted.

Introduction

In chapter 4 we studied basic specifications of amplifiers, one of three key components in wireless systems. We continue in this chapter our study about two other key components in wireless systems: oscillators and mixers, trying to achieve the same objectives of understanding their basic specifications, which can be obtained from the datasheets of these components. Therefore, students will be able to read datasheets and to choose the suitable oscillator or mixer for the relevant application.

The oscillator is used in a receiver as a local oscillator (LO) as shown in Figure 1 of chapter 4. By varying the frequency of the local oscillator, the mixer is used for frequency conversion (down conversion) from the received RF frequency to the intermediate frequency (IF). Therefore, the wanted information signal is obtained as we will see later.

In the transmitter, the local oscillator is used for carrier generation, and the mixer is used for frequency conversion (up conversion) from the intermediate frequency (IF), which contains the information signal, to the transmitted RF frequency (see Figure 2 - chapter 4).

In recent years, personal communications in high Megahertz and low Gigahertz frequency ranges are booming. Behind this achievements was the technological progress in MMIC integrated circuitry on one hand and application of frequency synthesis on the other hand.

Oscillators and frequency synthesizers are key components in RF and microwave transmitters and receivers, providing precisely controlled sources for frequency conversion and carrier generation.

In many wireless applications (e.g., cellular mobile communications) it is necessary to vary the frequency of the local oscillator. Simple transistor oscillators (such as Voltage Controlled Oscillator VCO) usually lack the frequency stability and low-noise performance required for modern wireless systems. Frequency synthesis methods can then be used to precisely derive higher frequencies from a crystal controlled source, and allow the generation of a closely-spaced local oscillator frequencies required for multichannel wireless transceivers. Phase-Locked Loops (PLL) are often used for this purpose.

Microwave Frequency Synthesizers

Frequency synthesizers provide a large number of precisely controlled frequencies derived from a stable oscillator. The stable reference source is usually a crystal-controlled oscillator, which may be housed in a temperature controlled environment for even great stability, such as Temperature-compensated crystal oscillator (TCXO) and oven-controlled crystal oscillator (OCXO).

Modern microwave frequency synthesizers are available in a form of MMIC[1] integrated circuitry, and are found in virtually all modern radios, cellular phones, and wireless data equipment.

There are two methods that can be used for frequency synthesis:

- Phase-Locked Loop (PLL) is used to derive multiples of a reference frequency.
- Direct Digital Synthesis (DDS) uses look-up table for the sine function and a digital-to-analog converter (DAC) to construct a sine wave of arbitrary frequency.

Direct digital synthesis

Direct digital synthesis (DDS) is a method of producing an analog waveform—usually a sine wave—by generating a time-varying signal in digital form and then performing a digital-to-analog conversion as shown in Figure 1. Because operations within a DDS device are primarily digital, it can offer:

- Fast switching between output frequencies,
- Fine frequency resolution,
- Operation over a broad spectrum of frequencies.

Note: With advances in design and process technology, today's DDS devices are very compact and draw little power.

---

[1] Monolithic Microwave Integrated Circuits (MMICs)

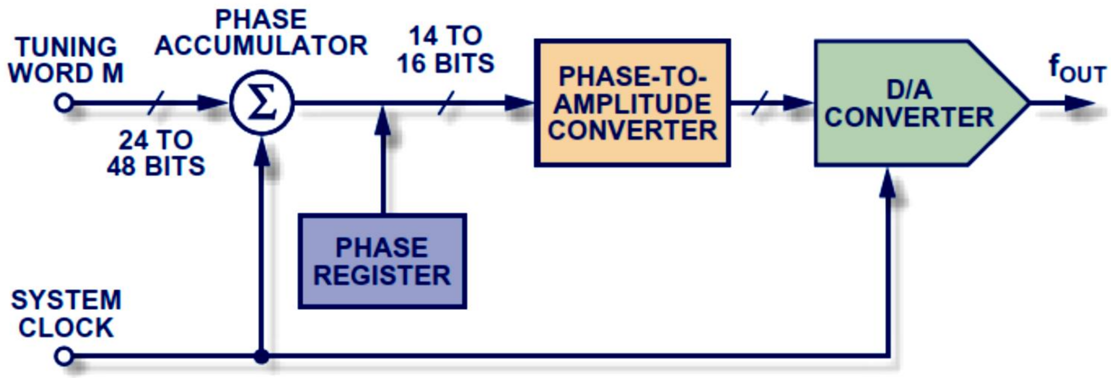


Figure 1: Block diagram of a DDS

A DDS produces a sine wave at a given frequency. The frequency depends on two variables, the reference-clock frequency and the binary number programmed into the frequency register (tuning word) as shown in Figure 2.

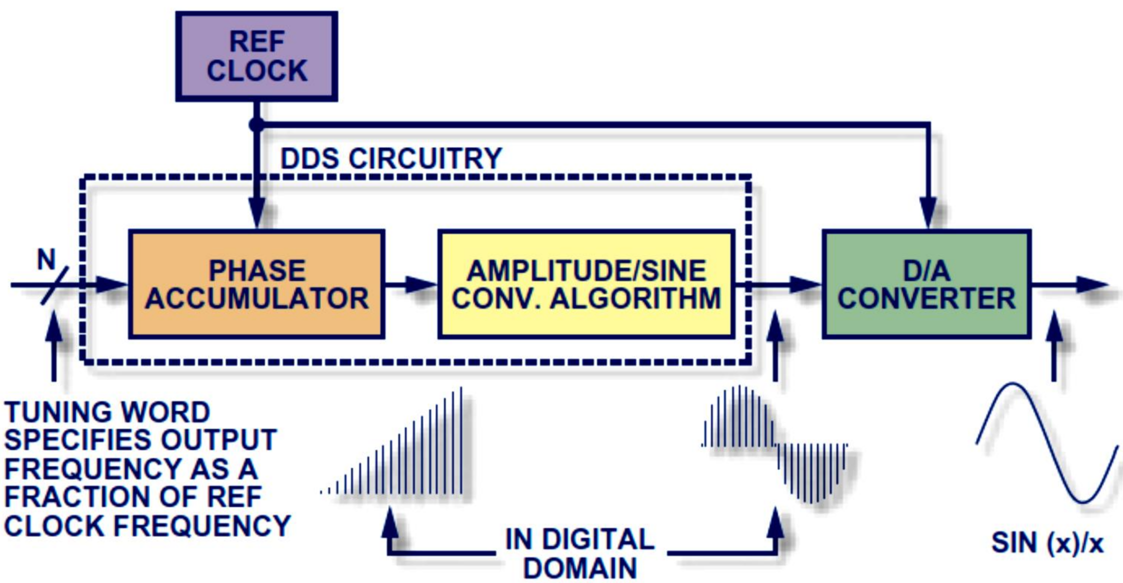


Figure 2: Signal flow through the DDS architecture

Applications of DDS

Applications currently using DDS-based waveform generation fall into two principal categories:

- Designers of communications systems requiring agile (i.e., immediately responding) frequency sources with excellent phase noise,
- Low spurious performance often choose DDS for its combination of spectral performance and frequency-tuning resolution.

Such applications include using a DDS for modulation, as a reference for a PLL to enhance overall frequency tunability, as a local oscillator (LO), or even for direct RF transmission.

Today's cost- competitive, high - performance, functionally integrated DDS ICs are becoming common in both communication systems and sensor applications. The advantages that make them attractive to design engineers include:

- Digitally controlled micro-hertz frequency-tuning and sub-degree phase-tuning capability,
- Extremely fast hopping speed in tuning output frequency (or phase),
- The digital architecture of DDS eliminates the need for the manual tuning and tweaking related to component aging and temperature drift in analog synthesizer solutions,
- The digital control interface of the DDS architecture facilitates an environment where systems can be remotely controlled and optimized with high resolution under processor control.

Note: for more information and questions about DDS, a useful document from Analog Devices entitled "All about Direct Digital Synthesis" is available as "ddspage.pdf".

Example from Analog Devices

To search for a DDS from Analog Devices,

Product Categories

RF & Microwave

Direct Digital Synthesis

We choose for example AD9914:

Features and Benefits

- 3.5 GSPS internal clock speed
- Integrated 12–bit DAC
- Frequency tuning resolution to 190 pHz $(190 \times 10^{-12}\ \mathrm{Hz})$
- 16–bit phase tuning resolution
- 12–bit amplitude scaling
- Programmable modulus
- Automatic linear and nonlinear frequency sweeping capability
- 32–bit parallel datapath interface
- 8 frequency/phase offset profiles
- Phase noise: -128 dBc/Hz (1 kHz offset at 1396 MHz)
- Wideband SFDR $< -50\ \mathrm{dBc}$
- Serial or parallel input/output control
- 1.8 V / 3.3 V power supplies

See data sheet "AD9914.pdf" for additional features.

Phase–Locked Loop Frequency Synthesizer

A phase–locked loop (PLL) is a frequency synthesis method, where a phase/frequency detector (PFD) compares a fed back frequency with a divided–down version of the reference frequency (Figure 3). The PFD's output current pulses are filtered and integrated to generate a voltage. This voltage drives an external voltage–controlled oscillator (VCO) to increase or decrease the output frequency so as to drive the PFD's average output towards zero.

The frequency synthesizer is obtained by programming the value of N counter, so it allows the generation of a variety of output frequencies as multiples of a single reference frequency (Figure 3). One main application is in generating local oscillator (LO) signals for the up- and down-conversion of RF signals as shown in Figure 2 of chapter 4.

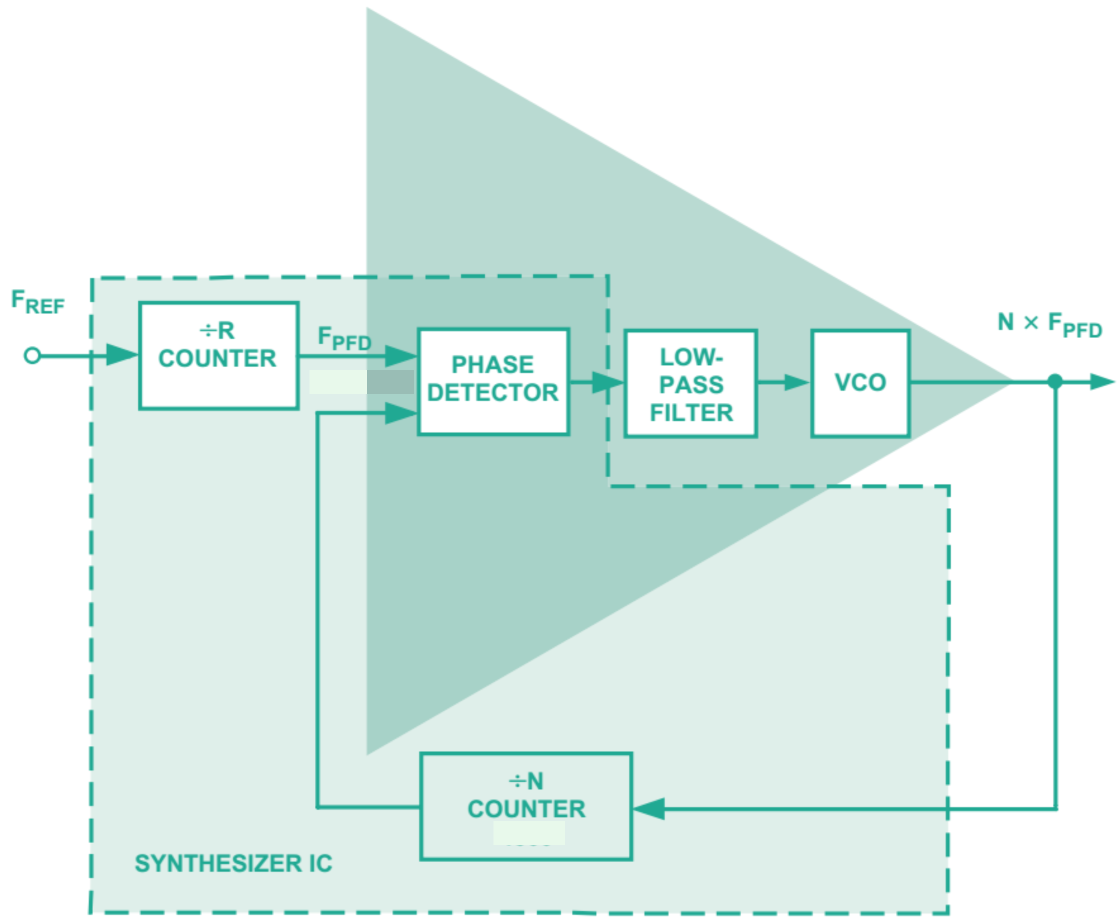


Figure 3. Block diagram of a PLL synthesizer.

Note: the term PLL technically refers to the entire system shown in Figure 3; however, sometimes it is meant to refer to the entire system except for the crystal oscillator, low-pass filter, and VCO. This is because these components are difficult to integrate on a PLL synthesizer IC. Recently, a PLL synthesizer IC with VCO is commercially available.

Fundamental components of PLL

The basic blocks of the PLL are shown in Figure the Error Detector (composed of a phase frequency detector PFD and a charge pump CP), Loop Filter (low-pass filter LPF), VCO, and a Feedback Divider, N. Negative feedback forces the error signal, $e(s)$, at the output of PFD to approach zero at which point the feedback divider output and the reference frequency are in phase and frequency lock, and $F_O = N \times F_{REF}$.

Referring to Figure 4, a system for using a PLL to generate higher frequencies than the input, the VCO oscillates at a frequency of F_O. A portion of this signal is fed back to the error detector, via a frequency divider with a ratio $1/N$. This divided down frequency is fed to one input of the error detector. The other input is a fixed reference frequency F_{REF}.

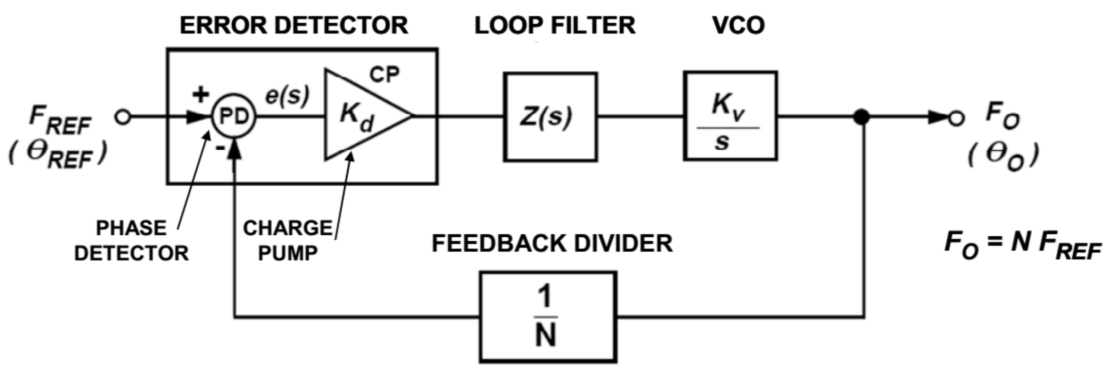


Figure 4. Basic blocks of a PLL.

The error detector compares the signals at both inputs. When the two signal inputs are equal in phase and frequency, $\boldsymbol{F_{REF} = F_O/N}$, the error will be constant and the loop is said to be in a "locked" condition.

The Modern Phase Frequency Detector with Charge Pump and its Advantages

The classical phase detector is a device that converts the differences in the two phases from N and R counters into an output voltage. Depending on the technology, this output voltage can either be applied directly to the loop filter, or converted to a current by the charge pump.

The modern PFD does a much better job dealing with a large error in frequency. It is typically accompanied with a charge pump. The PFD converts the phase error presented to it into a voltage, which in turn is converted by the charge pump into a correction current. These two devices are typically integrated together on the same chip and work together. The term of charge pump is only used to refer to the device that converts the error voltage, $e(s)$, to a correction current.

The charge pump PLL offers several advantages over the voltage phase detector and has all but replaced it. Using the PFD, the PLL is able to lock to any frequency; regardless of how far off it initially is in frequency and does not have a steady state phase error.

Note: The only case where the op−amp is really necessary is when the VCO tuning voltage needs to be higher than the charge pump can supply. In this case, an active filter is necessary.

N counter vs. P prescaler

The PLL starts with a stable crystal reference frequency (the frequency of a crystal oscillator). In Figure 5(A), the R counter divides this frequency to a lower one, F_1, which is called the comparison frequency. This is one of the inputs to the PFD.

N counter divides the VCO frequency and phase by N. If the output frequency of the VCO is low enough (on the order of 200 MHz or less), it can be implemented with a programmable digital counter fabricated with a low frequency process, such as CMOS.

It is desirable to implement as much of the N counter in CMOS as possible, for lower cost and current consumption.

However, if the VCO frequency is much higher than 200 MHz, then prescalers are often used to divide down the VCO frequency to something that can be handled with a lower frequency process, as shown in Figure 5(B).

Prescalers often divide by some power of two, since this makes them easier to implement. The most common implementations of prescalers are single modulus, dual modulus, and quadruple modulus. Of these, the dual modulus prescaler is most commonly used.

The N counter is the programmable element that sets the relationship between the input and output frequencies in the PLL. In addition to a straightforward N counter, it has evolved to include a prescaler, which can have a dual modulus when very high−frequency outputs are required. For example, let's assume that a 900 MHz output is required with 10 kHz spacing. A 10 MHz reference frequency might be used, with the R divider set at 1000. Then, the N−value in the feedback would need to be of the order of 90,000. This would mean at least a 17−bit counter capable of dealing with an input frequency of 900 MHz. To handle this range, it makes sense to precede the programmable counter with a fixed counter element to bring the very high input frequency down to a range at which standard CMOS will operate. This counter, called a prescaler, is shown in Figure 5(B).

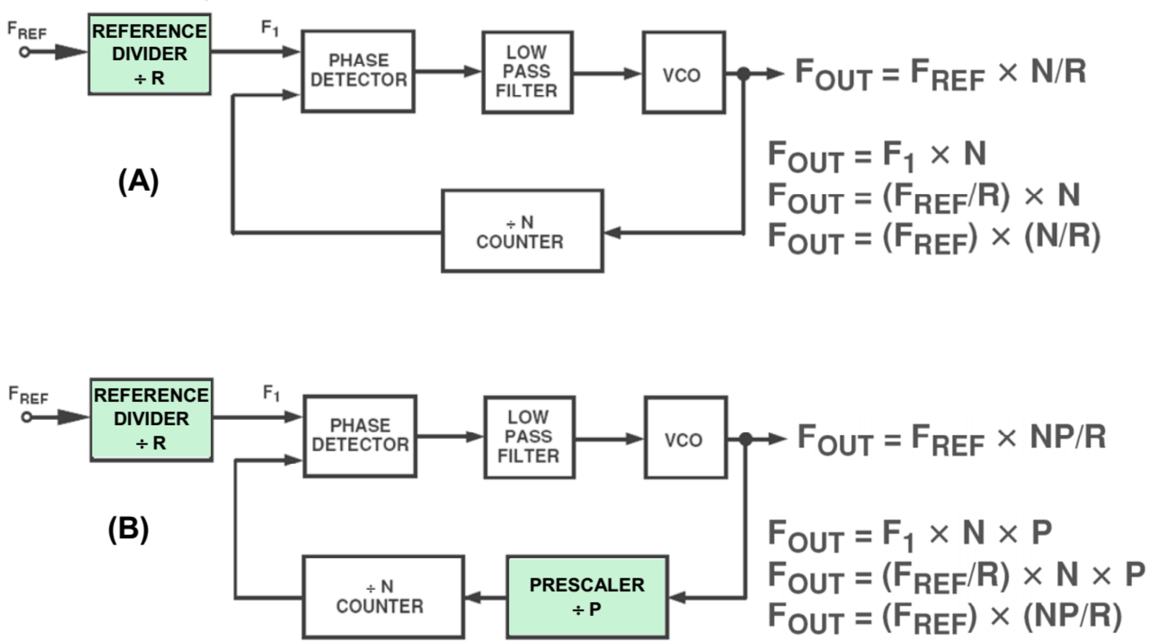


Figure 5: Adding an input reference divider and a prescaler to the basic PLL

However, note that using a fixed P prescaler as shown reduces the system resolution to $F_1 \times P$. This issue can be addressed by using a dual−modulus prescaler which has the advantages of P prescaler, but without loss of resolution. A dual−modulus prescaler is a counter whose division ratio can be switched from one value to another by an external control signal. By using the dual−modulus prescaler with an A and B counter, one can still maintain output resolution of F_1 with $N = BP + A$, as shown in Figure 6.

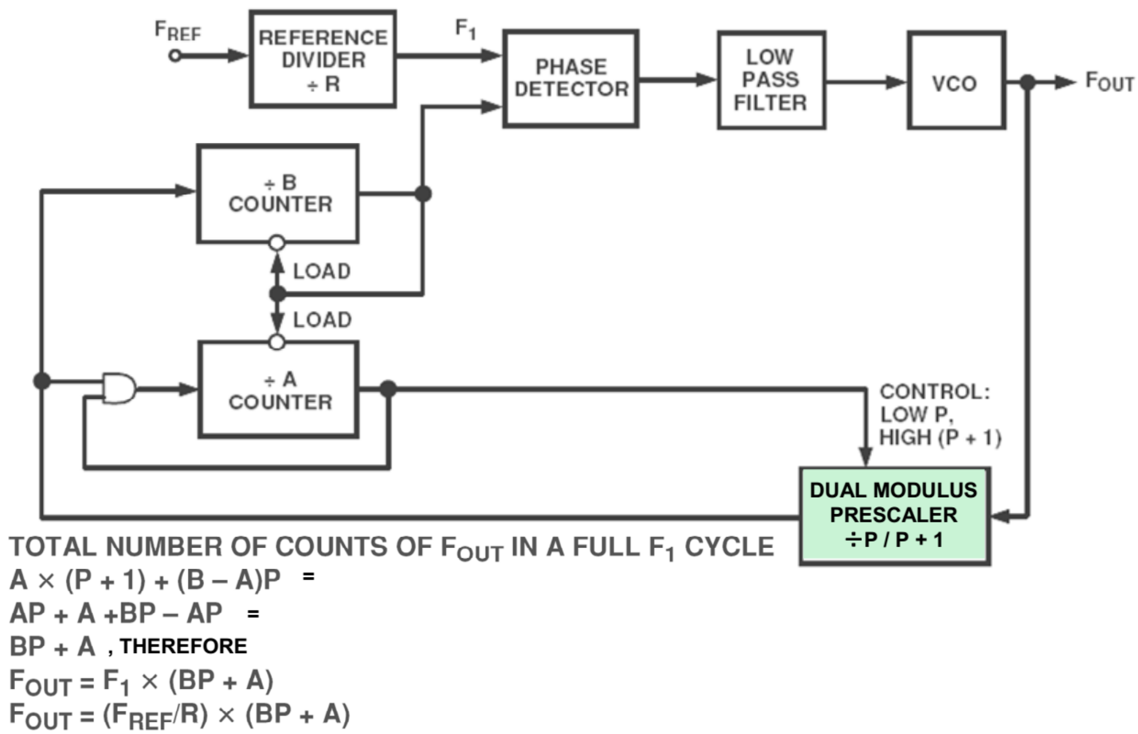


Figure 6: Adding a Dual Modulus Prescaler to the PLL

Determining the Loop Filter Topology and Order

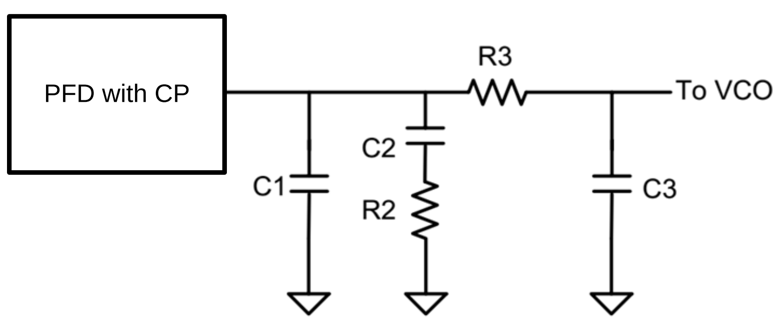


Figure 7: Third Order Passive Loop Filter

A third order passive loop filter is show in Figure 7. Passive loop filters are usually recommended above active loop filters, because adding active devices adds phase noise, complexity, and cost. However, there are cases where an active filter is necessary. The most common case arises when the maximum PLL charge pump voltage is lower than the VCO tuning voltage requirements. If higher tuning voltages are supplied to a VCO, then either the tuning range can be expanded or the phase noise reduced. In terms of filter order, the most basic is the second order filter. Additional RC low pass filtering stages can be added to reduce the reference spurs.

Choosing the Phase Margin and Loop Bandwidth

The phase margin relates to the stability of a system. This parameter is typically chosen between 40 and 55 degrees. Simulations show that a phase margin of about 48 degrees yields the optimal lock time. For minimum RMS phase error designs, 50 degrees is a good starting point for phase margin.

The loop bandwidth (the cut–off frequency of the low–pass filter) is the most critical parameter of the loop filter. The choice of loop bandwidth typically involves a trade–off between noise/spur levels and lock time. A smaller loop bandwidth will improve the noise/spur levels at the expense of lock time. A larger loop bandwidth yields the opposite effect. So it is important to find the loop bandwidth that provides a good balance between noise/spur levels and lock time. In cases where lock time and noise/spurs are not a major consideration, it often makes sense to choose the loop bandwidth at the frequency where the PLL noise equals the VCO noise for an optimal RMS phase error design. Another consideration is the loop filter capacitor sizes. If the loop bandwidth is too narrow, these may become unrealistically large. If it is too wide, these may become too small and be swamped out by parasitic capacitances and the input capacitance of the VCO.

Simplifying PLL/Loop Filter Design Using ADIsimPLL™

The ADIsimPLL™ software is a complete PLL design package which can be downloaded from the **Analog Devices'** website. The software has a user–friendly graphical interface, and a complete comprehensive tutorial for first–time users.

Traditionally, PLL Synthesizer design relied on published application notes to assist in the design of the PLL loop filter. It was necessary to build prototype circuits to determine key performance parameters such as lock time, phase noise, and reference spurious levels. Optimization was accomplished by "tweaking" component values on the bench and repeating lengthy measurements.

ADIsimPLL™ both streamlines and improves the traditional design process. Starting with the “new PLL wizard,” a designer constructs a PLL by specifying the frequency requirements of the PLL, selecting an integer−N or fractional−N implementation, and then choosing from a library of PLL chips, library or custom VCO, and a loop filter from a range of topologies. The program designs a loop filter and displays key parameters including phase noise, reference spurs, lock time, lock detect performance, and others.

The basic design process using ADIsimPLL™ can be summarized as follows:

1. Choose reference frequency, output frequency range, and channel spacing
2. Select PLL chip from **Analog Devices** list
3. Select VCO from **Analog Devices** or choose a PLL chip with VCO
4. Select loop filter configuration
5. Select loop filter bandwidth and phase margin
6. Run simulation
7. Evaluate time and frequency domain results
8. Optimize

ADIsimPLL™ works for integer−N or fractional−N PLLs, but does not simulate fractional−N spurs. Phase noise prediction for fractional−N devices assumes the device is operating in the "lowest phase noise" mode.

Advantages of Fractional−N over integer−N PLL

- The resolution at the fractional−N PLL output is reduced to small fractions of the PFD frequency, while the resolution at the output of an integer−N PLL is limited to steps of the PFD frequency, F_1. Thus, it is possible to generate output frequencies with resolutions of 100s of Hz, while maintaining a high PFD frequency. As a result the N−value is significantly less than for integer−N.
- Since noise at the charge pump is multiplied up to the output at a rate of $20\ logN$, significant improvements in phase noise are possible.

- Also offering a significant advantage is the lock−time improvement made possible by fractional−N. The PFD frequency set to 20 MHz and loop bandwidth of 150 kHz will allow the synthesizer jump 30 MHz in $< 30\ \mu s$. With the super−fast lock times of fractional−N, synthesizers must have lock time specs that allow the 2 "ping−pong" PLLs to be replaced with a single fractional−N PLL block.

If fractional−N offers all these advantages, why are integer−N PLLs still so popular?

Spurious levels! A fractional−N divide by 19.1 consists of the N−divider dividing by nineteen 90% of the time, and by twenty 10% of the time.

$$19.1 = 19 \times 0.90 + 20 \times 0.10$$

The average division is correct, but the instantaneous division is incorrect. Because of this, the PFD and charge pump are constantly trying to correct for instantaneous phase errors. The heavy digital activity of the sigma−delta modulator, which provides the averaging function, creates spurious components at the output. The digital noise, combined with inaccuracies in matching the hard−working charge pump, results in spurious levels greater than those allowable by most communications standards.

Practical PLL Synthesizer circuits

In this type of application, the PLL is used to generate a set of discrete frequencies.

FM radio

In FM radio, the valid stations range from 88 to 108 MHz, and are spaced 0.1 MHz apart. The PLL generates a frequency that is 10.7 MHz less than the desired channel, since the received signal is mixed with the PLL signal to always generate a fixed IF (Intermediate Frequency) of 10.7 MHz. Therefore, the PLL generates frequencies ranging from 77.3 MHz to 97.3 MHz. The channel spacing would be equal to the comparison frequency, which would be 100 kHz.

A fixed crystal frequency of 10 MHz can be divided by an R value of 100 to yield a comparison frequency of 100 kHz. Then the N value ranging from 773 to 973

is programmed into the PLL. If the user is listening to a station at 99.3 MHz and decides to change the channel to 103.4 MHz, then the R value remains at 100, but the N value changes from 886 to 927. The performance of the radio will be impacted by the spectral purity of the PLL signal produced and also the time it takes for the PLL to switch frequencies.

GSM Receiver

Figure 8 shows a typical application5m of dual PLL frequency synthesizer in a superheterodyne receiver for GSM 900 MHz (the RF frequency range is $890-915$ MHz). The output frequency of the 1^{st} synthesizer is tuned over the range ($650-675$ MHz) and used as 1^{st} local oscillator to down convert the RF frequency range to the 1^{st} IF (240 MHz).

$$f_{RF} - f_{LO} = IF = 240\ \mathrm{MHz}$$

The 2^{nd} synthesizer is used to generate a fixed frequency (229.3 MHz). This frequency is mixed with the 1^{st} IF to produce the fixed 2^{nd} IF.

$$240\ \mathrm{MHz} - 229.3\ \mathrm{MHz} = 10.7\ \mathrm{MHz}$$

In this example, an ADF4350 synthesizer from Analog Devices (AD) can be used with an internal VCO and an external filter and TCXO as reference oscillator. An input reference (R) counter reduces the reference input frequency (13 MHz in this example) to PFD frequency ($f_{PFD} = f_{REF}/R$); and a feedback (N) counter reduces the output frequency for comparison with the scaled reference frequency at the PFD. At equilibrium, the two frequencies are equal, and the output frequency is $N \times\ f_{PFD}$. The feedback counter is a dual−modulus prescaler type, with A and B counters ($N\ =\ BP\ +\ A$, where P is the prescale value).

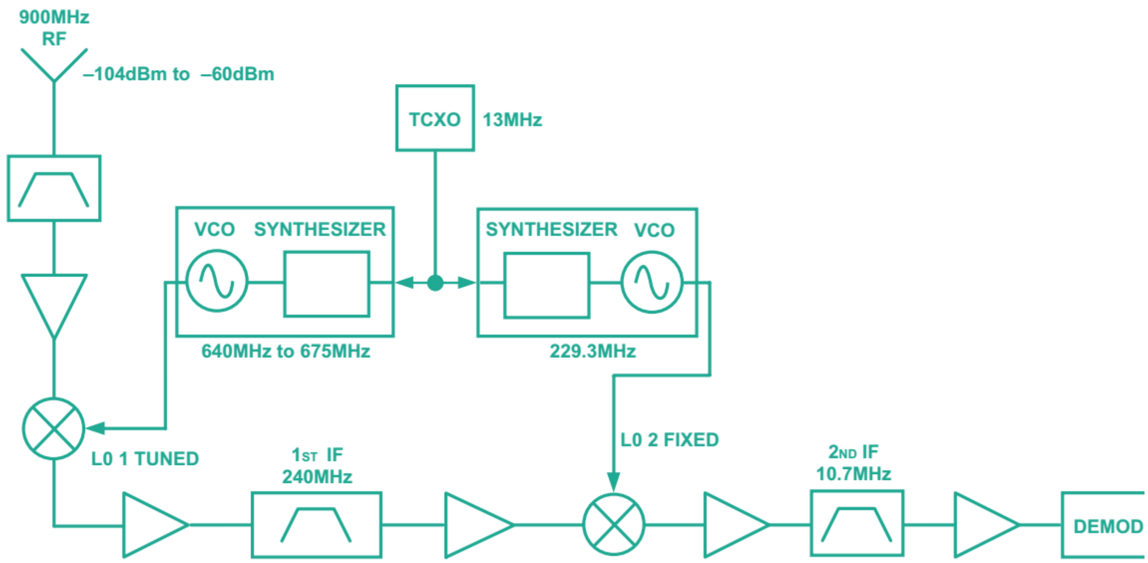


Figure 8. Dual PLL synthesizer used to convert down from GSM RF to baseband.

PLL Synthesizer Applications

Some of its earlier applications included keeping power generators in phase and synchronizing to the sync pulse in a TV Set, recovering a clock from asynchronous data, and demodulating an FM modulated signal.

Base station and handset LOs are common application, but synthesizers are also found in low frequency clock generators, wireless LANs, radar systems, and collision–avoidance systems.

Key performance parameters to be considered in selecting a PLL synthesizer

The major ones are: phase noise, reference spurs, and lock time.

Phase Noise

A frequency synthesizer is a type of oscillator, and in any oscillator design, frequency stability is of critical importance. We are interested in both long–term and short–term stability.

Long−term frequency stability is concerned with how the output frequency varies over a long period of time (hours, days, or months). It is usually specified as the ratio, $\Delta f/f$ for a given period of time, expressed as a percentage or in dB.

Short−term stability, on the other hand, is concerned with variations that occur over a period of seconds or less. These variations can be random or periodic. A spectrum analyzer can be used to examine the short−term stability of a signal. Figure 9 shows a typical spectrum, with random and discrete frequency components causing a broad skirt and spurious peaks.

Oscillator phase noise refers to the short−term random fluctuation in the frequency (or phase) of an oscillator signal. Oscillator phase noise $\mathfrak{L}(f_m)$ is defined as the ratio of single−sideband noise power in 1−Hz BW @ f_m offset from carrier, to the carrier signal power.

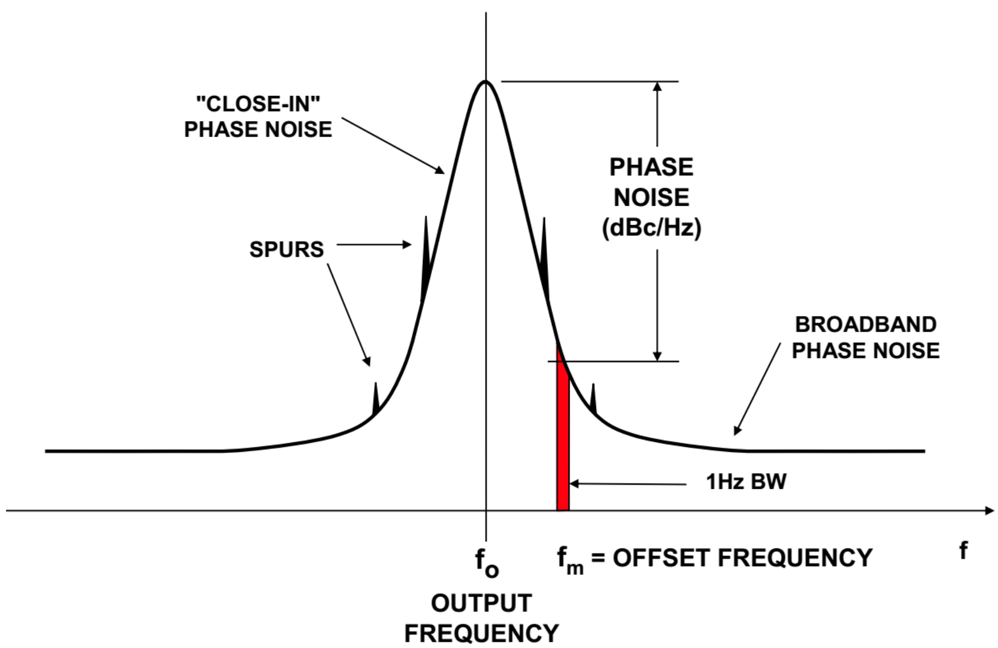


Figure 9: Phase noise and spurs

An ideal oscillator at a frequency f_0 is a sine wave with a spectrum consisting of a Dirac pulse at f_0, as shown in Figure below. Real oscillator output signal has a spectrum with random and discrete frequency components causing a broad skirt and spurious peaks (Figure 10).

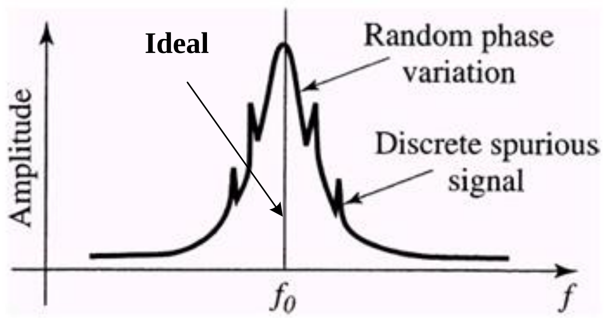


Figure 10: Real oscillator output signal spectrum

Oscillator phase noise may degrade the selectivity of the receiver by causing down−conversion of signals located nearby the desired signal frequency (reciprocal mixing), the adjacent channels spacing, and the Bit Error Rate (BER) for digital modulation methods, because the probability of error depends on the variance of the noise power.

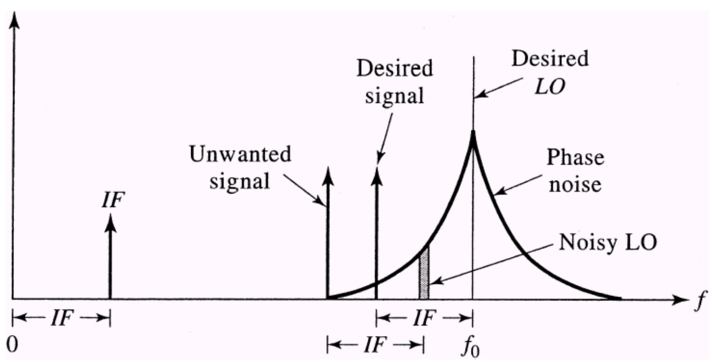


Figure 11: Reciprocal mixing

The discrete spurious components (spurs) could be caused by known clock frequencies in the signal source, power line interference, and mixer products.

The broadening caused by random noise fluctuation is due to phase noise. It can be the result of thermal noise, shot noise, and/or flicker noise in active and passive devices. The phase noise spectrum of an oscillator shows the noise power in a 1 Hz bandwidth as a function of frequency. Phase noise is defined as the ratio of the noise in a 1 Hz bandwidth at a specified frequency offset, f_m, to the oscillator signal amplitude at frequency f_0. The phase noise is measured in dBc/Hz at a frequency offset, f_m, with the frequency axis on a log scale.

Reference Spurs

These are spurious peaks at discrete offset frequencies generated by the internal counters and charge pump operation at the PFD frequency. These spurs will be increased by mismatched up and down currents from the charge pump, charge-pump leakage, and inadequate decoupling of supplies. The spurious tones will get mixed down on top of the wanted signal and decrease receiver sensitivity.

Lock Time

The lock time of a PLL is the time it takes to jump from one specified frequency to another specified frequency within a given frequency tolerance. The jump size is normally determined by the maximum jump the PLL will have to accomplish when operating in its allocated frequency band. The step-size for GSM-900 is 45 MHz and for GSM-1800 is 95 MHz. The required frequency tolerances are 90 Hz and 180 Hz, respectively. The PLL must complete the required frequency step in less than 1.5 time slots, where each time slot is 577 µs.

ADF4350 synthesizer

The ADF4350 allows implementation of fractional-N or integer-N phase-locked loop (PLL) frequency synthesizers if used with an external loop filter and external reference frequency.

The ADF4350 has an integrated voltage controlled oscillator (VCO) with a fundamental output frequency ranging from 2200 MHz to 4400 MHz. In addition, divide-by-1/2/4/8 or 16 circuits allow the user to generate RF output frequencies as low as 137.5 MHz. For applications that require isolation, the RF output stage can be muted.

Main features of the ADF4350 synthesizer are: (for further details see the attached datasheet “ADF4350.pdf”)

- Output frequency range: 137.5 MHz to 4400 MHz
- Fractional-N synthesizer and integer-N synthesizer
- Low phase noise VCO
- Power supply: 3.0 V to 3.6 V

- Programmable dual−modulus prescaler of $4/5$ or $8/9$
- Programmable output power level

Parameter	B Version Min	Typ	Max	Unit	Conditions/Comments
RF OUTPUT CHARACTERISTICS					
Maximum VCO Output Frequency			4400	MHz	
Minimum VCO Output Frequency	2200			MHz	Fundamental VCO mode
Minimum VCO Output Frequency Using Dividers	137.5			MHz	2200 MHz fundamental output and divide by 16 selected
VCO Sensitivity		33		MHz/V	
Frequency Pushing (Open-Loop)		1		MHz/V	
Frequency Pulling (Open-Loop)		90		kHz	Into 2.00 VSWR load
Harmonic Content (Second)		−19		dBc	Fundamental VCO output
Harmonic Content (Third)		−13		dBc	Fundamental VCO output
Harmonic Content (Second)		−20		dBc	Divided VCO output
Harmonic Content (Third)		−10		dBc	Divided VCO output
Minimum RF Output Power [5]		−4		dBm	Programmable in 3 dB steps
Maximum RF Output Power [5]		5		dBm	
Output Power Variation		±1		dB	
Minimum VCO Tuning Voltage		0.5		V	
Maximum VCO Tuning Voltage		2.5		V	
NOISE CHARACTERISTICS					
VCO Phase-Noise Performance [6]		−89		dBc/Hz	10 kHz offset from 2.2 GHz carrier
		−114		dBc/Hz	100 kHz offset from 2.2 GHz carrier
		−134		dBc/Hz	1 MHz offset from 2.2 GHz carrier
		−148		dBc/Hz	5 MHz offset from 2.2 GHz carrier
		−86		dBc/Hz	10 kHz offset from 3.3 GHz carrier
		−111		dBc/Hz	100 kHz offset from 3.3 GHz carrier
		−134		dBc/Hz	1 MHz offset from 3.3 GHz carrier
		−145		dBc/Hz	5 MHz offset from 3.3 GHz carrier
		−83		dBc/Hz	10 kHz offset from 4.4 GHz carrier
		−110		dBc/Hz	100 kHz offset from 4.4 GHz carrier
		−132		dBc/Hz	1 MHz offset from 4.4 GHz carrier
		−145		dBc/Hz	5 MHz offset from 4.4 GHz carrier
Normalized Phase Noise Floor (PN_{SYNTH}) [7]		−220		dBc/Hz	PLL Loop BW = 500 kHz
Normalized 1/f Noise (PN_{1_f}) [8]		−111		dBc/Hz	10 kHz offset; normalized to 1 GHz
In-Band Phase Noise [9]		−97		dBc/Hz	3 kHz offset from 2113.5 MHz carrier
Integrated RMS Jitter [10]		0.5		ps	
Spurious Signals Due to PFD Frequency		−70		dBc	
Level of Signal With RF Mute Enabled		−40		dBm	

Oscillators

Phase-Locked Loops need two types of oscillators: a crystal oscillator as a reference frequency, and a Voltage Controlled Oscillator to generate the output frequency.

Crystal oscillators

Quartz crystals are useful for good frequency stability, especially at frequencies below a few 100 MHz. Crystal-controlled oscillators therefore find extensive use as stable frequency source in wireless system.

Further stability can be obtained by controlling the temperature of the quartz crystal. Temperature-compensated crystal oscillator (TCXO) and oven-controlled crystal oscillator (OCXO) are examples.

Frequency accuracies of ten Parts Per Million (PPM) are not uncommon for crystal oscillators (Crystal unit). The main cause of frequency accuracy in these oscillators is drift over temperature. The TCXO (Temperature Compensated Crystal Oscillator) has a temperature sensor and compensation (temperature compensation circuit) to correct the crystal frequency over a wide temperature range. This improves the frequency accuracy by a factor of ten. The OCXO (Oven Controlled Crystal Oscillator) improves the performance by approximately another factor of ten by having an oven heat the crystal to a constant temperature.

Specifications of crystal oscillators

Let's take an example from Murata (www.murata.com). Murata is a global leader in the design, manufacture and supply of advanced electronic materials, leading edge electronic components, and multi-functional, high-density modules. Murata innovations can be found in a wide range of applications from mobile phones to home appliances, and automotive applications to energy management systems and healthcare devices.

Products of interests from Murata:

- Capacitors
- Inductors
- Noise Suppression Products / EMI Suppression Filters / ESD Protection Devices
- Resistors
- Thermistors
- Sensors
- Timing Devices (Ceramic Resonator / Crystal Unit / Oscillator)

Murata provides a selection guide for crystal oscillators (see the attached “p79e.pdf” file). Crystal oscillators of interests are in page 35. The main specifications are:

Part Number	Type	Frequency (MHz)	Frequency Tolerance (ppm max.) [at 25°C±3°C]	Frequency Shift by Temperature (ppm max.) [Standard Condition: +25°C]	Frequency Aging (ppm max./Year)	Current Consumption (mA max.)	Frequency Controlled Range (ppm)

- Frequency of oscillation: from 10 up to 52 MHz.
- Frequency tolerance of the frequency of oscillation (in PPM): For example, a frequency of 10 MHz, has a frequency tolerance of $\pm 1.0 \text{ PPM}$ at $25°\text{C} \pm 3°\text{C}$. This means that the frequency of oscillation is: $10 \text{ MHz} \pm 10 \text{ Hz}$.
- Frequency shift by temperature (short–term stability): $\pm 0.5 \text{ PPM}$ over a temperature range from $-30°\text{C}$ to $85°\text{C}$.
- Frequency aging (long–term stability): $\pm 1.0 \text{ PPM/Year}$, it is the change in frequency with time due to internal changes in the oscillator. It is the frequency change with time while factors external to the oscillator (environment, power supply, etc.) are kept constant. Generally aging, rather than shift or drift, is what one measures during oscillator evaluation.
- Frequency–controlled range for voltage–controlled crystal oscillators (VC–TCXO): from $\pm 5 \text{ PPM}$ to $\pm 15 \text{ PPM}$.

Product ID	
XT	VC-TCXO
XN	TCXO

Voltage Controlled Oscillator (VCO)

A voltage–controlled capacitor, called varactor, is used to tune the frequency of the VCO. The varactor is a diode whose junction capacitance may be controlled by changing the DC reverse bias applied to the diode. Then the resonant frequency of the oscillator is an LC circuit, and the resulting configuration of the transistor and the LC resonant circuit is called a Voltage Controlled Oscillator (VCO).This VCO is used in PLL to precisely control the VCO frequency.

VCO Performance Parameters

Frequency Range

The frequency range of the VCO is set by the capacitors in the circuit and how much the varactor diode capacitance can change. A wider frequency range is always desirable, but this comes at the cost of phase noise. Because the frequency accuracy of the VCO is typically process and temperature dependent, the guaranteed frequency range of the datasheet is typically much less than the actual frequency range the VCO is capable of tuning over.

VCO Gain

The gain of the VCO is expressed in MHz/V and is how much the output frequency changes for a change in the input voltage. This gain typically amplifies noise voltages, and therefore it is desirable to keep this low for noise purposes, but this makes the tuning range narrower. A simplified way of viewing the VCO gain is to treat this as a constant. If this assumption is grossly wrong, then the tuning range can be broken up into several regions. Under this assumption, it can be calculated from the extreme frequencies and extreme tuning voltages.

$$K_{VCO} \approx \frac{f_{max} - f_{min}}{V_{Tune,max} - V_{Tune,min}}$$

Pushing and Power Supply Noise Rejection

Pushing refers to how much a change in voltage at the power supply pins of the VCO impact the output frequency.

Pulling

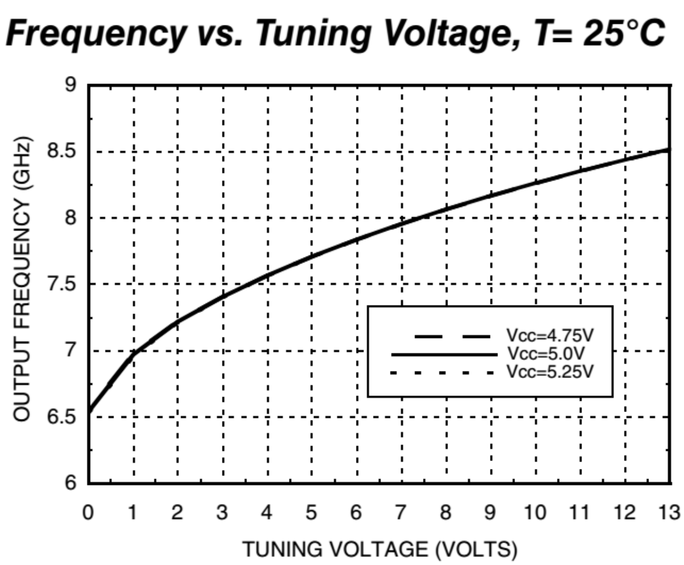
Pulling refers to how much the VCO frequency will shift when a load is placed on the output.

Harmonics

VCOs generate harmonics, which occur at a multiple of the out frequency. In general, these are considered undesirable. The desired frequency (the fundamental) is the first harmonic.

Frequency Tuning Characteristic

Frequency versus tuning voltage performance for a given VCO. This is usually graphed as frequency vs. voltage (figure 12).



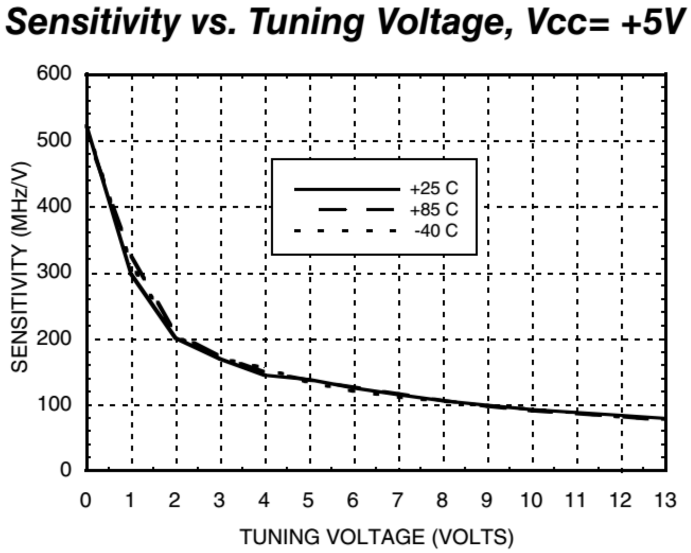


Figure 12(a): Frequency Tuning Characteristic and sensitivity

of the VCO "hmc508"

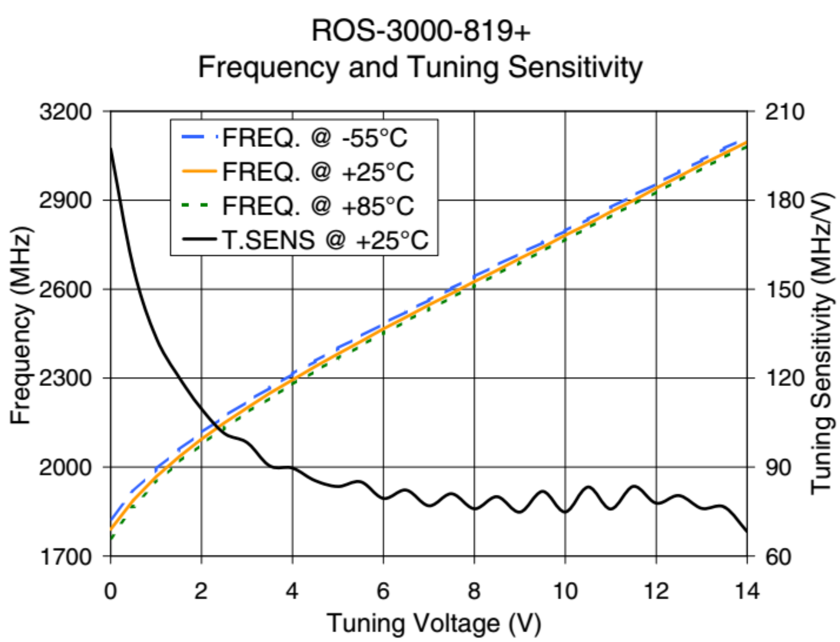


Figure 12(b): Frequency Tuning Characteristic and sensitivity

of the VCO "ROS3000"

Tuning linearity

Tuning sensitivity as a function of tuning voltage is a measure of tuning linearity. For any given application, specify the minimum and maximum of the tuning sensitivity (Figure 12).

- Tuning Sensitivity:

This is the slope of the tuning characteristic and is expressed as frequency change per unit voltage change $\Delta f/\Delta V$ (MHz/V).

- Tuning linearity

Tuning linearity is the deviation of the frequency versus tuning voltage characteristic from a best-fit straight line. It is sometimes called tuning nonlinearity.

- Output Power:

The fundamental sinusoidal frequency output of the VCO measured into a 50-ohm load, expressed in dBm.

Exercises

1. Extract the specifications and key performance parameters of the VCO "hmc508" from Analog Devices (datasheet attached "hmc508.pdf").
2. Extract the specifications and key performance parameters of the VCO "ROS3000" from Mini-Circuits (datasheet attached "ROS-3000-819+.pdf").

Mixers

Frequency conversion is a critical function in all wireless systems: it is needed in a transmitter for up conversion (to obtain higher frequencies), and in a receiver for down conversion (to obtain lower frequencies). A key component used to accomplish this is the mixer.

In chapter 4, Figure 1 shows a simplified block diagram of a receiver, where the mixer is used for down conversion. A mixer is a nonlinear three-port device to achieve frequency conversion as shown in Figure 13. This figure shows that an ideal mixer produces an output consisting of the sum and difference frequencies of its two input signals. Operation of practical RF and microwave mixers is usually based on the nonlinearity provided by either a diode or a transistor. Since a nonlinear component generates a wide variety of harmonics and other products of input frequencies, then filtering must be used to select the desired frequency components. Modern microwave systems typically use several mixers and filters to perform the functions of frequency up-conversion and down-conversion (Chapter 4, Figure 1) between baseband signal frequencies and RF carrier frequencies.

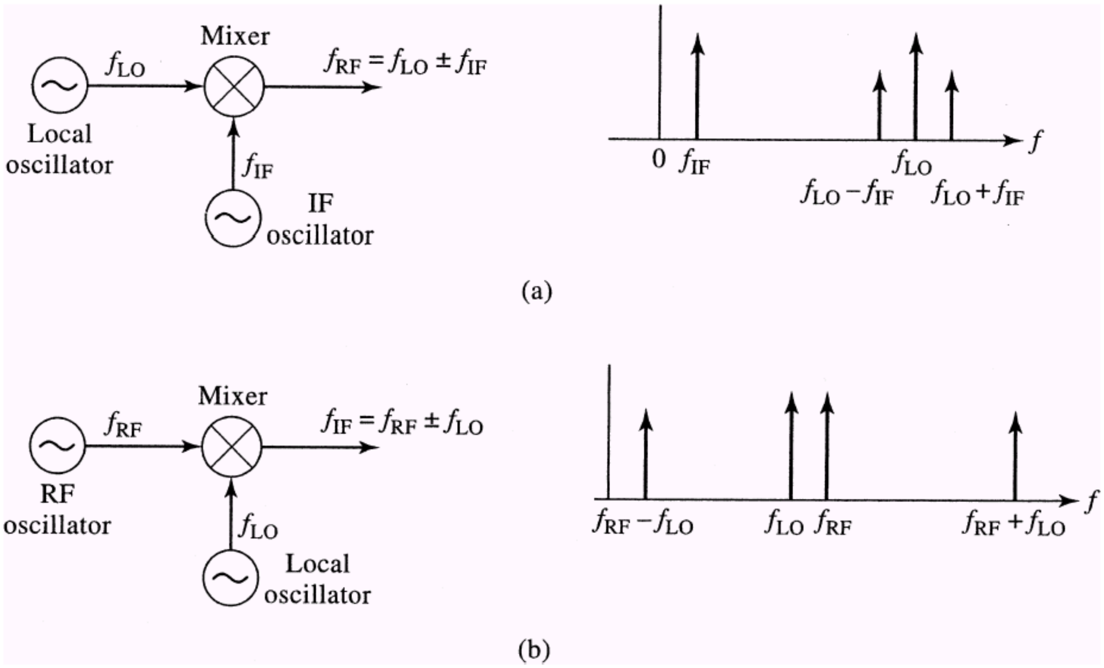


Figure 13: Frequency conversion using a mixer. (a) Up–conversion. (b) Down–conversion.

Mixer Terminology

Up–conversion

Figure 13(a) illustrates the operation of frequency up–conversion, as occurs in a transmitter. A local oscillator (LO) signal at the relatively high frequency f_{LO} is connected to the LO input port of the mixer. A lower frequency baseband or intermediate frequency (IF) signal is applied to the other mixer input. This signal typically contains the information or data to be transmitted.

The output of the ideal mixer is given by the product of the LO and IF signals. The RF output is seen to consist of the sum and difference of the input signal frequencies:

$$f_{RF} = f_{LO} \pm f_{IF}$$

The sum and difference frequencies are called the sidebands of the carrier frequency f_{LO}, with $f_{LO} + f_{IF}$ being the upper sideband (USB), and $f_{LO} - f_{IF}$ being the lower sideband (LSB). A double–sideband (DSB) signal contains both upper and lower sidebands, while a single–sideband (SSB) signal can be produced by filtering or by using a single–sideband mixer.

Down–conversion

Figure 13(b) illustrates the operation of frequency down–conversion, as occurs in a receiver. In this case, the received RF signal is applied to the RF port of the mixer, along with the LO signal connected to the LO input port of the mixer. Thus the mixer output consists of the sum and difference of the input signal frequencies:

$$f_{IF} = f_{LO} \pm f_{RF}$$

In practice, the RF and LO frequencies are relatively close together, so the sum frequency is approximately twice the RF frequency, while the difference is much smaller than f_{RF}. The desired IF output in a receiver is the difference frequency, which is easily selected by low–pass filtering:

$$f_{IF} = f_{RF} - f_{LO}$$

In a receiver, when the LO frequency is below the RF, it is called low–side injection and the mixer a low–side down–converter; when the LO is above the RF, it is called high–side injection, and the mixer a high–side down–converter.

Image frequency

In a receiver the RF input signal at frequency f_{RF} is typically delivered from the antenna, which may receive RF signals over a relatively wide band of frequencies. For a receiver with an LO frequency f_{LO} and IF frequency f_{IF}, the RF input frequency that will be down–converted to the IF frequency is

$$f_{RF} = f_{LO} + f_{IF}$$

Now consider the RF input frequency given by

$$f_{IM} = f_{LO} - f_{IF}$$

The down conversion of f_{IM} by the mixer gives a negative intermediate frequency $-f_{IF}$. Mathematically, this frequency is identical to f_{IF} because the Fourier spectrum of any real signal is symmetric about zero frequency, and thus contains negative as well as positive frequencies. The RF frequency f_{IM} is called the image frequency. The image frequency is important in receiver design because a received RF signal at the image frequency $f_{IM} = f_{LO} - f_{IF}$ is indistinguishable at the IF stage from the desired RF signal of frequency $f_{RF} = f_{LO} + f_{IF}$.

The choice of which RF frequency $f_{RF} = f_{LO} \pm f_{IF}$ is the desired and which is the image is arbitrary, depending on whether the LO frequency is above or below the desired RF frequency. Another way of viewing this difference is to note that f_{IF} may be negative. Observe that the desired and image frequencies are separated by $2f_{IF}$.

This means that the mixer will develop an intermediate frequency output when the input signal frequency is greater or less than the local oscillator frequency by an amount equal to the intermediate frequency. That is, there are two input frequencies— namely, $|f_{LO} \pm f_{IF}|$, that will result in f_{IF} at the mixer output. This introduces the possibility of simultaneous reception of two signals differing in frequency by $2f_{IF}$.

Conversion loss

For an ideal mixer, each of the outputs is only half the amplitude (one−quarter the power) of the individual inputs; thus, there is a loss of 6 dB in this ideal linear mixer. Conversion loss is a measure of the efficiency of the mixer in providing frequency down−conversion from the RF input signal to the IF output signal. Since only one sideband is utilized in most applications, the specifications given in datasheets are for a single−sideband output. If both sidebands are utilized, the conversion loss is 3 dB lower than in the single−sideband case.

An important figure of merit for a mixer is therefore the conversion loss, which is defined as the ratio of available RF input power to the available IF output power, expressed as a positive number in dB.

Conversion Gain

An active mixer can have an internal amplifier in one or more of the three signal paths. When the amplifier is in the RF or IF path, it generally provides IF output power that is greater than the RF input power. Therefore, conversion gain is specified instead of conversion loss; it is equal to the ratio of the IF single−sideband output power to the RF input power, expressed as a positive number in dB.

Isolation

Isolation is a measure of the circuit balance within the mixer. When the isolation is high, the amount of "leakage" or "feed through" between the mixer ports will be very small. Generally, the double−balanced mixers provide the best isolation.

The LO−to−RF isolation is the amount the LO drive power attenuated when it is measured at the RF port, the IF port being terminated with 50 ohms. The LO−to−IF isolation is the amount the LO drive power is attenuated when it is measured at the IF port, the RF port being terminated with 50 ohms. Normally, only the LO isolations are specified, not RF isolation. This is because the RF signal power is much lower than the LO drive level; therefore, RF leakage is usually not a limiting performance factor.

Example 1

Consider $f_{LO} = 10\,\mathrm{MHz}$ and $f_{IF} = 1\,\mathrm{MHz}$; the wanted response is at the IF frequency, $f_{IF} = 1\,\mathrm{MHz}$ for $f_{RF} = 11\,\mathrm{MHz}$. However, the mixer produces the same IF in response to the image frequency, $f_{IM} = 9\,\mathrm{MHz}$ as shown in Figure 14.

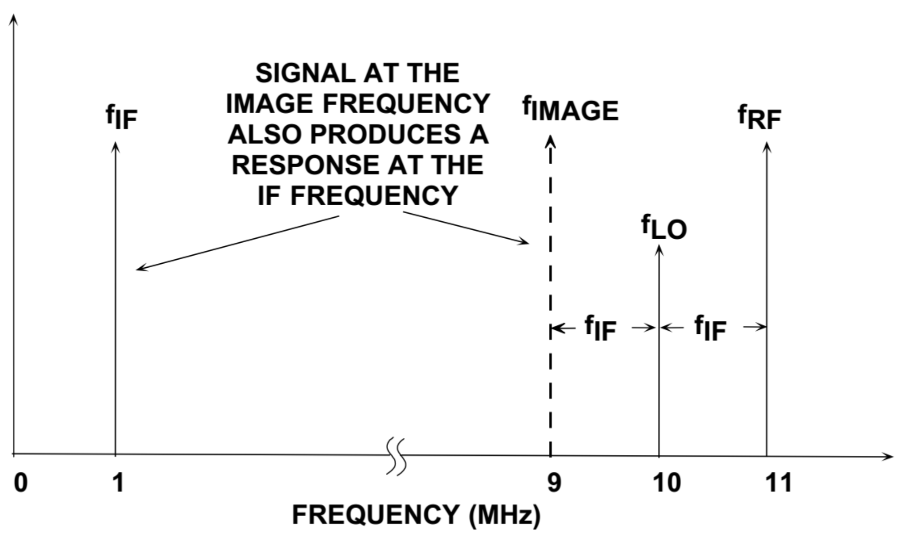
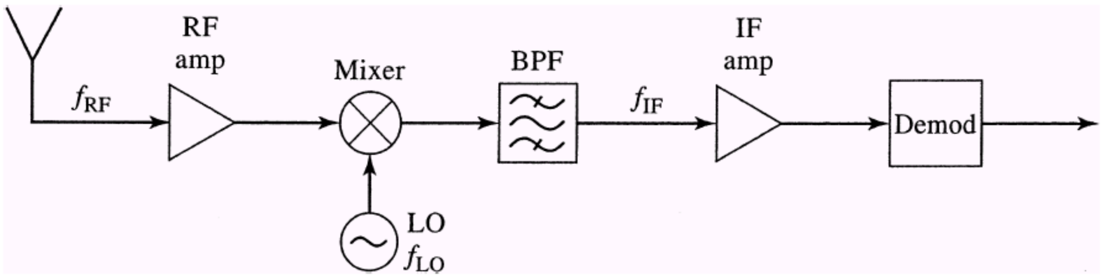


Figure 14: Image frequency.

Example 2

A GSM cellular telephone system uses the superheterodyne receive shown below to receive the range of frequency [935, 960] MHz.



What is the IF frequency for the upper LO frequency tuning range [1010, 1035] MHz of the local oscillator? Determine the range of image frequency. Does the image frequency fall within the receive passband?

Solution

The IF frequency is

$$f_{LO} - f_{RF} = 75 \text{ MHz}$$

The range of the RF image frequency range is

$$f_{IM} = f_{RF} + 2 \times f_{IF} \in [935 + 2 \times 75, 960 + 2 \times 75] \text{ MHz}$$

$$\mathrm{f_{IM}} \in [1085, 1110] \text{ MHz}$$

which is well outside the receive passband.

Mixer Configurations

Single–Ended Mixers

A basic diode mixer circuit is shown in Figure 15 and the circuit for a single–ended FET mixer is shown in Figure16. The RF and LO inputs are combined in a diplexer, which superimposes the two input voltages to drive the diode. The duplexing function can be implemented using a directional coupler or hybrid junction to provide signal combining as well as isolation between the two inputs.

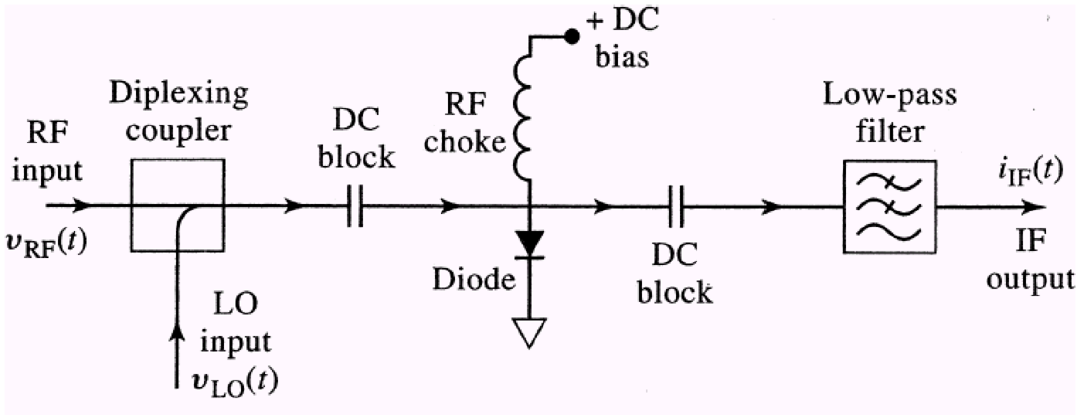


Figure 15: Single–ended diode mixer circuit.

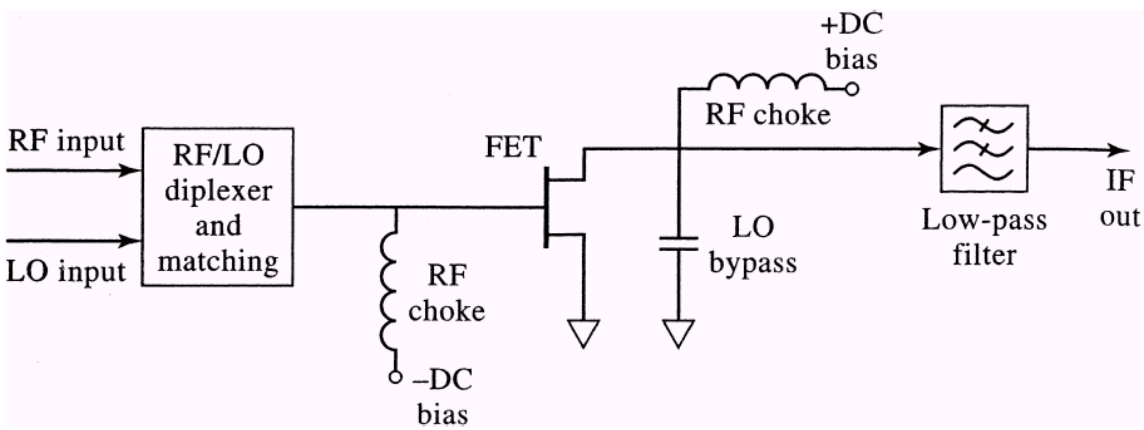


Figure 16: single–ended FET mixer circuit.

Balanced Mixers

RF input matching and RF–LO isolation can be improved using a balanced mixer, which consists of two single–ended mixers combined with a hybrid junction as shown in Figure 17.

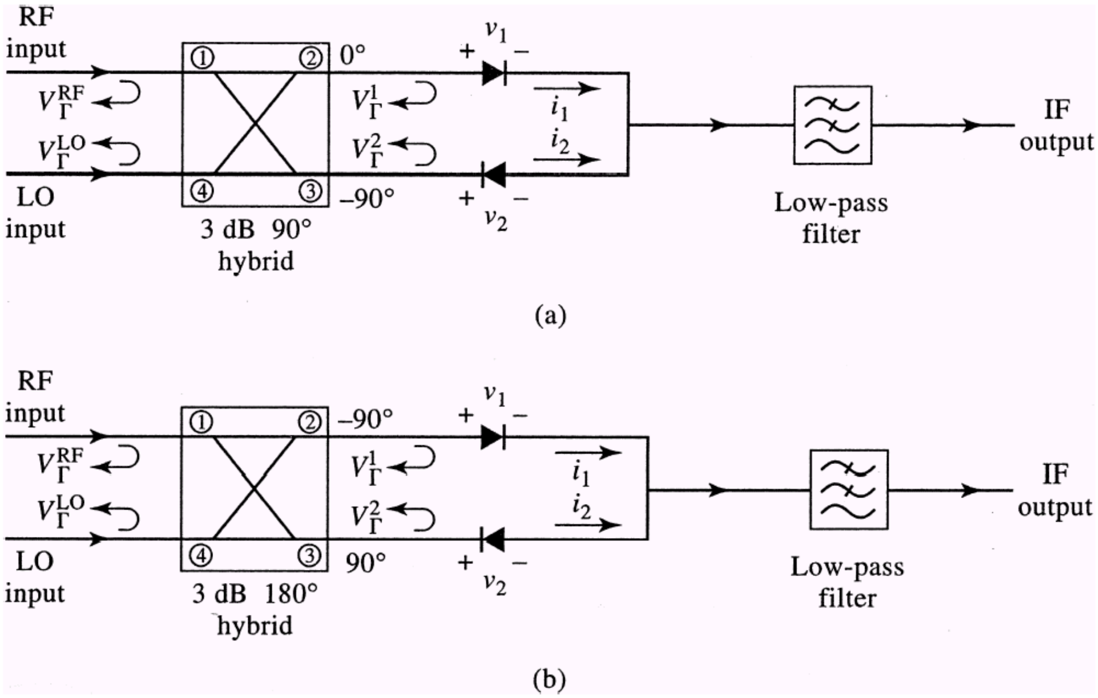


Figure 17: Balanced mixer circuits. (a) Using a 90° hybrid. (b) Using a 180° hybrid.

Double−balanced mixer

The double−balanced mixer shown in Figure 18 provides good isolation between all three ports, as well as rejection of all even harmonics of the RF and LO signals. This leads to very good conversion loss, but less than ideal input matching at the RF port. The double−balanced mixer also provides a higher third−order intercept point than either a single−ended mixer or a balanced mixer.

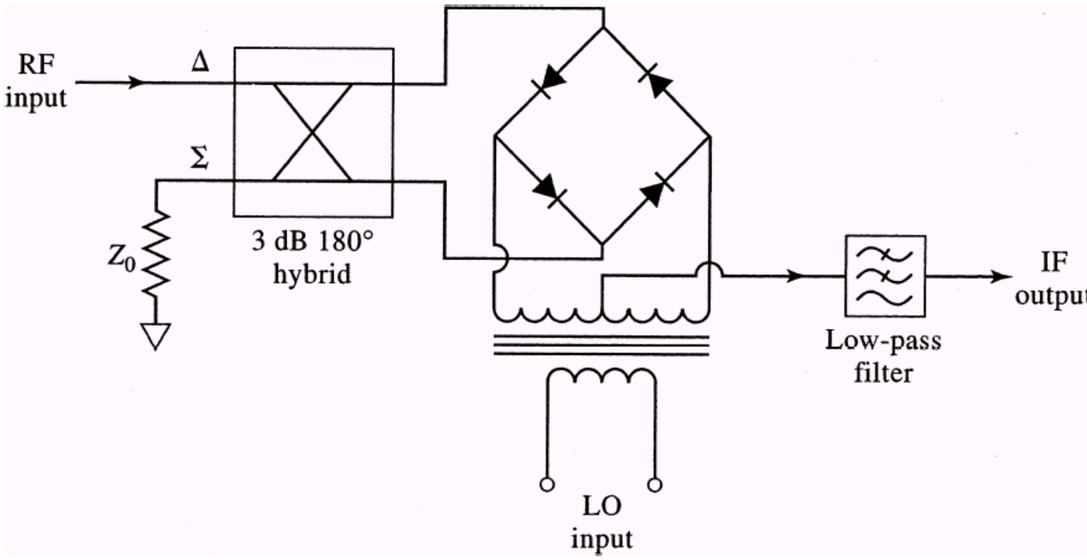


Figure 18: Double balanced mixer circuits using a 180° hybrid.

Down−converter Specifications

The specification of a down−converter has the following main requirements:

1. Operating frequency band (MHz),
2. Conversion loss or gain (dB),
3. Noise figure (dB),
4. IIP_3, (dBm),
5. Isolation between RF/IF and LO ports (dB),
6. Nominal LO power (dBm),
7. Input and output impedance (Ω),
8. Input return loss (dB).

Categories of Mixers

When searching for a mixer, manufacturers classify them in main categories. For example, Analog Devices (www.analog.com) has the following categories:

I/Q and Image Reject Mixers	RF Mixers with Integrated LO
I/Q Downconverters/Receivers	Single, Double & Triple Balanced Mixers
I/Q Upconverters/Transmitters	Sub-Harmonic Mixers

I/Q and Image Reject Mixers

In-phase quadrature (I/Q) mixers and image reject mixers (IRM) offer a robust frequency mixing solution for applications that require high image rejection or sideband suppression. These mixers can be used as an up-converter or down-converter. I/Q or IRM mixers use two double-balanced mixers with a $90°$ hybrid. The inherent architecture of I/Q mixers eliminates the resulting product due to image frequency during down-conversion and allows the selection of desired sideband during up-conversion.

I/Q Down-converters/Receivers

In-phase quadrature (I/Q) down-converters integrate an image reject mixer, a low noise amplifier, a $90°$ hybrid, and an LO buffer amplifier in one package. These devices are used in receiver applications to down-convert radio frequency (RF) signals to the desired intermediate frequency (IF) while ensuring low noise and high image rejection performance.

I/Q Up-converters/Transmitters

In-phase quadrature (I/Q) up-converters integrate an I/Q mixer, RF amplifier, $90°$ hybrid, and an LO buffer in one package. These devices are used in transmitter applications to up-convert intermediate frequency (IF) to a desired radio frequency (RF) while reducing the need for filtering unwanted sideband.

RF Mixers with Integrated LO

Analog Devices RF mixers/multipliers include both active and passive, as well as broadband and narrowband, products. These devices promise high performance and flexibility to help designers achieve their target specifications, reduce design time, and optimize costs. With leading specifications including speed, dynamic range, noise figure under blocking, SFDR, and ESD performance, ADI RF mixers and multipliers help to raise the bar within RF systems.

Single, Double & Triple Balanced Mixers

Single, double or triple balanced mixers are the fundamental mixer architecture ranges, which means that all ports in the mixer utilize a single-ended to differential balun structure to reduce undesired signals. Analog Devices' fundamental mixer portfolio's frequency coverage from sub MHz to 90 GHz can offer a discrete mixer solution in either packaged or die form to meet the current market's end application requirements. Options include local oscillator (LO) buffer amplifiers or intermediate frequency (IF) amplifiers to further meet customer-specific design objectives. Table 1 presents some mixers of this category.

Table 1. Passive Mixers

RF Frequency (MHz)	Single Mixer	Single Mixer and IF Amplifier	Dual Mixer and IF Amplifier
500 to 1700	ADL5367	ADL5357	ADL5358
1200 to 2500	ADL5365	ADL5355	ADL5356
2300 to 2900	ADL5363	ADL5353	ADL5354

The specifications of the mixer HMC785LP4E are given in the attached datasheet ("hmc785.pdf"). We can easily read the main specifications from the following table:

Electrical Specifications,
T_A = +25° C, LO = 0 dBm, Vcc = Vcc1, 2, 3 = +5V, G_Bias = +2.5V *

Parameter	Min.	Typ.	Max.	Units
Frequency Range, RF		1.7 - 2.2		GHz
Frequency Range, LO		1.5 - 2.2		GHz
Frequency Range, IF		50 - 300		MHz
Conversion Loss		8	10	dB
Noise Figure (SSB)		8		dB
IP3 (Input)		36		dBm
1 dB Compression (Input)		26		dBm
LO to RF Isolation	18	30		dB
LO to IF Isolation	18	25		dB
RF to IF Isolation	25	39		dB
LO Drive Input Level (Typical)		-6 to +6		dBm
Supply Current Icc total		160	180	mA

Exercise 1

The comparison of the balanced mixer ADL5365 and the double balanced mixer ADL5812 shows that the later has higher LO to RF isolation. Search for the datasheets of these two mixers, read and compare their LO to RF isolation values.

Sub−Harmonic Mixers

Sub−harmonic mixers are composed of a passive mixer with a sub−harmonically pumped $(\times 2)$ local oscillator. These mixers allow designers to use low LO frequency, which eases the need to generate a high frequency LO signal. For high frequency application design, Analog Devices sub−harmonic mixers offer a simpler alternative to traditional mixers with no LO multiplier.

Exercise 2

Search for the datasheet of the mixer LTC5585. What is the name of the manufacturer? What are the main specifications of this mixer. What is its main application? What about the price of this mixer compared to the two mixers of exercise 1?

Test

10 marks for each question, Successful test mark is 60/100

Choose the correct answer.

1. Phase–Locked Loop is used to derive multiples of a reference frequency
   a. True
   b. False
2. Direct digital synthesis is a method used for generating digital signals
   a. True
   b. False
3. What is the name given to a mixer configuration that cancels out received signal frequencies that are equidistant from the oscillator frequency as the desired signal, but on the opposite side of the oscillator?
   a. Sideband reject
   b. Mirror reject
   c. Image reject
   d. Opposite reject
4. In a phase–locked loop oscillator circuit, what does the phase detector do?
   a. Detects phase
   b. Compares relative phase between the reference input and oscillator output and introduces a shift as necessary to maintain phase coherence
   c. Injects random phase shifts at the input to stabilize the output frequency
   d. Phase detectors are not used in PLL circuits
5. For a fundamental frequency of 100 MHz, what is the first harmonic frequency?
   a. 50 MHz
   b. 100 MHz
   c. 200 MHz
   d. 1000 MHz

6. What is the spectrum of an ideal oscillator at a frequency f_0 ?

   a. A comb of Dirac pulses at the fundamental $\boldsymbol{f_0}$ and all harmonics

   b. A Dirac pulse at $\boldsymbol{f_0}$

   c. A sine wave

   d. An intermodulation of all harmonics

7. When is harmonic mixing typically used?

   a. When harmonics cannot be filtered out.

   b. When multiple frequencies are needed.

   c. When the LO signal needs to be multiplied.

   d. Never.

8. What advantage does a Fractional−N synthesizer have over Integer−N?

   a. Fractional−N is less expensive than Integer−N.

   b. Fractional−N consumes less PCB area than Integer−N.

   c. Fractional−N is much simpler than Integer−N.

   d. Fractional−N permits output frequency steps that are small that the input reference source.

9. What advantage does P prescaler has over N counter in a synthesizer?

   a. P prescaler is less expensive than N counter.

   b. P prescaler consumes less current than N counter.

   c. P prescaler is programmable.

   d. $\boldsymbol{P}$ prescaler operates at higher frequencies than $\boldsymbol{N}$ counter.

10. In an Integer−N synthesizer, a 10 MHz reference frequency is used and 200 kHz output frequency spacing is required. What is the value of the R divider?

    a. $\mathbf{50}$

    b. 500

    c. 200

    d. 1000

Question number	Answer	Feedback
1	a	Revise Microwave Frequency Synthesizers
2	b	Revise Direct digital synthesis
3	c	Revise Mixer categories
4	b	Revise Fundamental components of PLL
5	b	Revise Phase–Locked Loop Frequency Synthesizer
6	b	None
7	c	None
8	d	None
9	d	None
10	a	None

Chapter Six:

RF and Microwave system architectures and performance

Keywords

Superheterodyne receiver, direct conversion receiver, low IF receiver, direct RF sampling receiver, receiver dynamic range, automatic gain control (AGC) system, frequency division duplex (FDD), time division duplex (TDD), minimum discernible signal (MDS) receiver sensitivity receiver selectivity, spurious response rejection.

Abstract

In this chapter we aim at reviewing different architectures of RF and microwave systems, their performance and technical issues. Superheterodyne, direct conversion, low IF and direct RF sampling architectures are presented. The receiver dynamic range and its automatic gain control (AGC) system are also presented. The objectives are not only to make the receiver achieving the performance, current consumption, cost, and size as targeted, but also to define the specifications of each individual device in the receiver and thus to ensure the targeted receiver performance and other goals achieved.

Learning Objectives

In this chapter, student will be able to:

- Review basic concepts of a communication system, communication channels, duplexing methods, and multiple access schemes.
- Define different receiver system considerations.
- Recognize superheterodyne, direct conversion, low IF and direct RF sampling receiver architectures.
- Recognize receiver dynamic range and components of receiver AGC system.
- Understand performance tradeoffs between receiver architectures.

Introduction

In the previous course "Antennas and Wave Propagation" we have studied the basic propagation models, the link budget of a point-to-point link, and propagation impairments and performance enhancement techniques in wireless systems.

In this chapter we focus on the different architectures of RF and microwave communication systems and their performance.

Brief history of RF and microwave systems

By unifying the works in electromagnetism of Lorentz, Faraday, Ampere, and Gauss, Maxwell presented in 1864 Maxwell's equations, and predicted the propagation of electromagnetic waves in free space at the speed of light. 20 years later, Hertz validated the electromagnetic wave (wireless) propagation in the laboratory. Almost two decades later, Marconi commercialized the use of electromagnetic wave propagation for wireless communications and allowed the transfer of information from one continent to another without a physical connection. The telegraph became the means of fast communications.

In the early 1900s, most wireless transmission occurred at very long wavelengths. Wireless communications using telegraphs, broadcasting, telephones, and point-to-point radio links were available before World War II. The widespread use of these communication methods was accelerated during and after the war.

After 1980, cordless phones and cellular phones became popular. Nowadays, personal communication systems (PCSs) operating at higher frequencies with wider bandwidths are emerging with various services such as voice, video, messaging, email, data, and on-line services. The direct link between satellites and PCSs can provide these services to everyone, at anytime, anywhere in the world, even in the most remote regions of the globe.

Figure 1 summarizes this brief history of RF and microwave systems. In the 1990s, the direct broadcast satellite (DBS) systems have replaced cable television, and the global positioning systems (GPS), RF identification (RFID) systems, and remote sensing and surveillance systems have found many commercial applications.

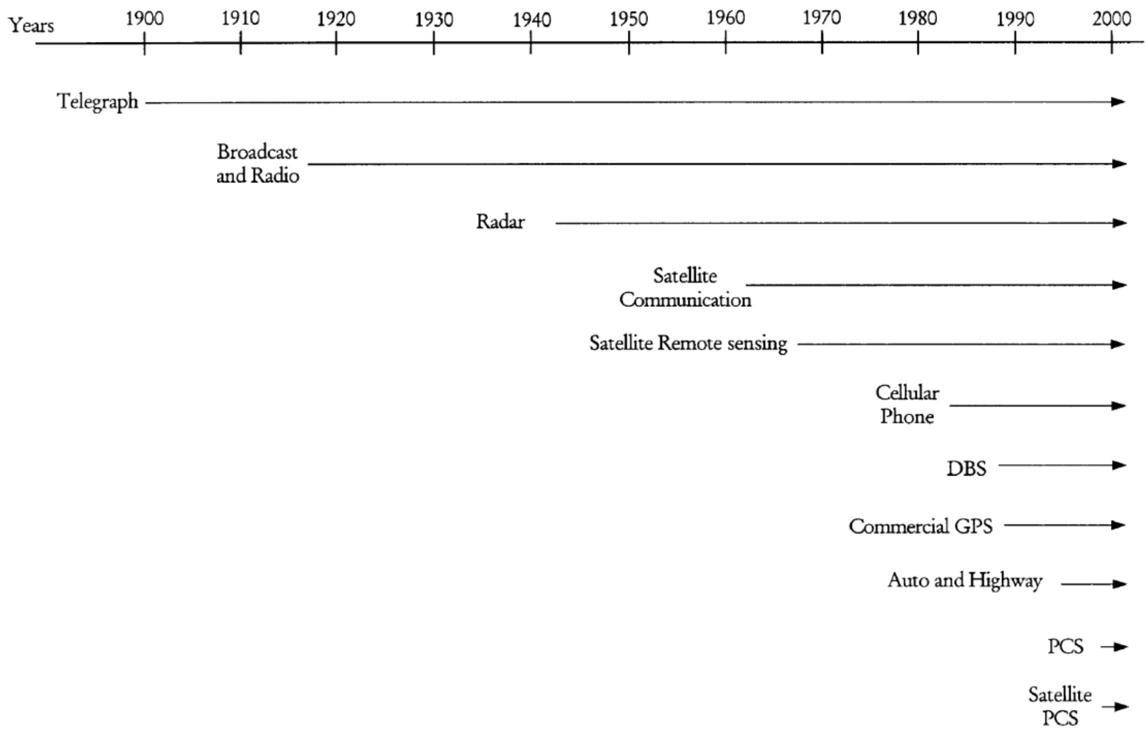


Figure 1. Brief history of RF and microwave systems

Basic Communication System

Figure 2 highlights the basic structure of a communication system. The task of the transmitter is to convert the message signal produced by a source of information into a form suitable for transmission over the channel. The channel transports the message signal and delivers it to the receiver. In wireless system, free-space is the communication channel, where the signal is distorted due to channel imperfections, noise and interfering signals (originating from other sources), with the result that the received signal is a corrupted version of the transmitted signal. The task of the receiver is to produce an estimate of the original message signal for the user of information. We say an “estimate” here because of the unavoidable deviation, however small, of the receiver output compared to the transmitter input, the deviation being attributed to channel imperfections, noise, and interference.

The transmitter and the receiver consist of two main sections: Baseband and RF. Baseband section deals with low–frequency signals, and RF section deals with high frequency signals. In this course, we are interested in the RF section.

In general, an RF receiver is defined from the port connected to a receiver antenna to an analog–to–digital converter (ADC), and an RF transmitter is defined from a digital–to–analog converter (DAC) to the port connected to a transmitter antenna (see Figure 2 of chapter 4). The RF receiver and transmitter are usually formed not only by RF circuitry and devices but also by intermediate frequency (IF) circuitry and devices. The ADC and DCA are often used as a boundary between the RF transceiver and its digital counterpart.

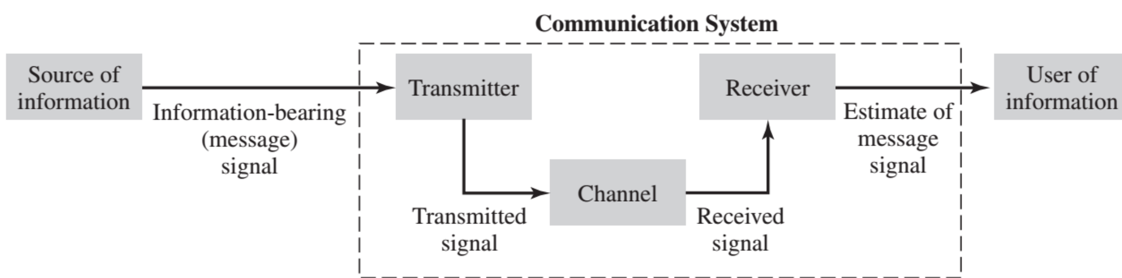


Figure 2. Basic block diagram of a communication system

A transceiver is a device combined of a transmitter and a receiver in one unit. The main blocks of an RF transceiver can be functionally classified into the following categories: filters, amplifiers, mixers, oscillators, synthesizers, modulator/demodulators, ADC/DAC, coupler/divider/combiner/attenuators, etc. An RF transceiver will use most of these function blocks but may not be all of them. The characterization and specification of most of these function blocks have been addressed in previous chapters of this course, or in previous courses.

Types of Communication Channels

A communication channel may be simplex, half duplex, or full duplex. In simplex channel, signals are transmitted in only one direction; from one device as a transmitter to the other as a receiver. In half–duplex channel, both devices may transmit, but only one at a time. In full–duplex channel (also called duplex), both devices may transmit and receive simultaneously. In the latter case, the channel is carrying signals in both directions at the same time. Table 1 summarizes these types of communication channels.

Type of Channel	Properties	Applications
Simplex	One-way only	FM radio, television
Half duplex	Two-way, only one at a time	Police radio
Full duplex	Two-way, both at the same time	Mobile systems

Table 1. Types of communication channels

Duplexing Methods

In a mobile communication system, for example, multiple users (mobile stations) may be assigned different channels to communicate with the base station according to various techniques, known as multiple access schemes. To enable simultaneous two–way communication between mobile station and base station, in this full–duplex system, two duplexing methods are possible:

1. Frequency Division Duplex

   In frequency division duplex (FDD), the total bandwidth allocated to the system is divided into two sub–bands: uplink (UL) from mobile station to base station, and downlink (DL) from base station to mobile station. FDD duplexing method illustrated in Figure 3 has the following features:

- Each user is allocated paired channels (UL/DL) separated by a duplexing frequency distance to keep interference between transmission and reception to a minimum.
- FDD requires either 2 antennas or a diplexer[1] to enable device to use both frequency channels with single antenna.

[1] A diplexer is a component used to separate two frequency bands. It is commonly used in a transceiver to separate the transmitting and receiving frequency bands.

- Because it requires less power to transmit a lower frequency over a given distance, uplink frequencies in mobile systems are always the lower band of frequencies – this saves valuable battery power of the mobile stations.

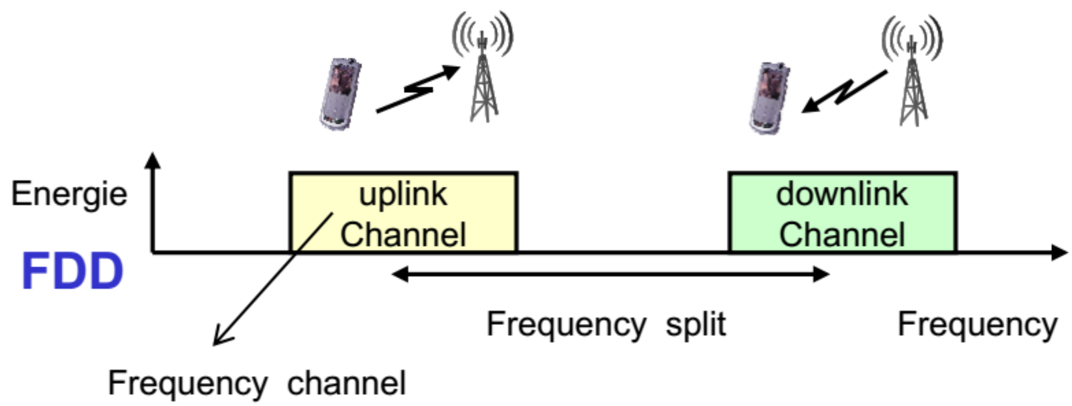


Figure 3. Principle of FDD duplexing method

2. Time Division Duplex

In time division duplex (TDD), the total bandwidth allocated to the system is shared, and different time slots are assigned for uplink and downlink, as shown in Figure 4. In a mobile communication system, different mobile stations can use the same frequency without interference using this scheme, where each connection is allotted its own up- and downlink pair. TDD duplexing method has the following features:

- TDD uses time instead of frequency to provide uplink and downlink.
- Duplex channel consists of an uplink and a downlink time slot.
- If time separation between uplink and downlink time slots is small then transmission and reception of data appears simultaneous.

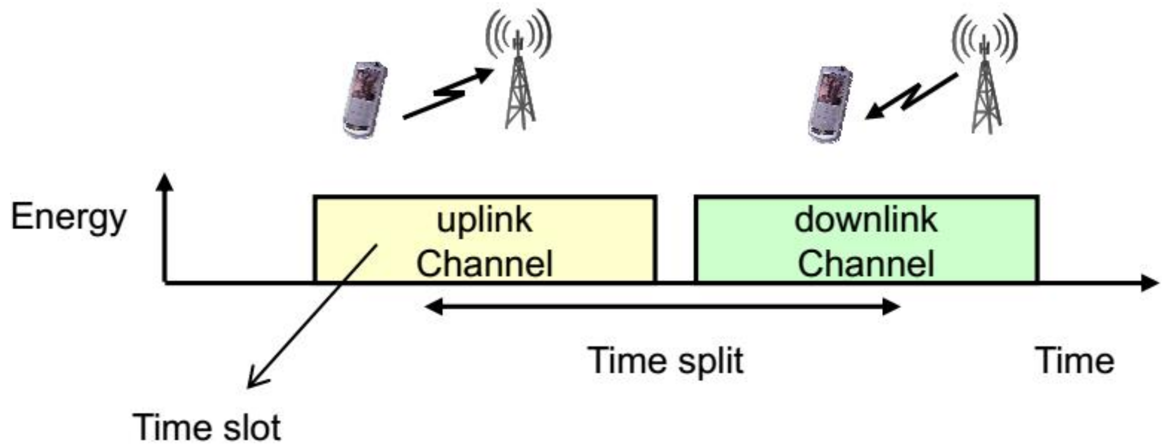


Figure 4. Principle of TDD duplexing method

Multiple Access Schemes

3. Frequency Division Multiple Access FDMA

In FDMA, the total bandwidth allocated to the system is divided into two sub-bands (UL/DL) and each of these is further divided into channels of fixed bandwidth, which are then assigned to different users. While a user is assigned a given channel, no one else is allowed to transmit in that channel. In a mobile communication system, each mobile station is allocated a pair of channels, separated by the duplex spacing, one in the uplink sub-band, for transmitting to the base station, the other in the downlink sub-band, for reception from the base station. Figure 5 shows how frequency sub-bands and paired channels are allocated in GSM-900 mobile system.

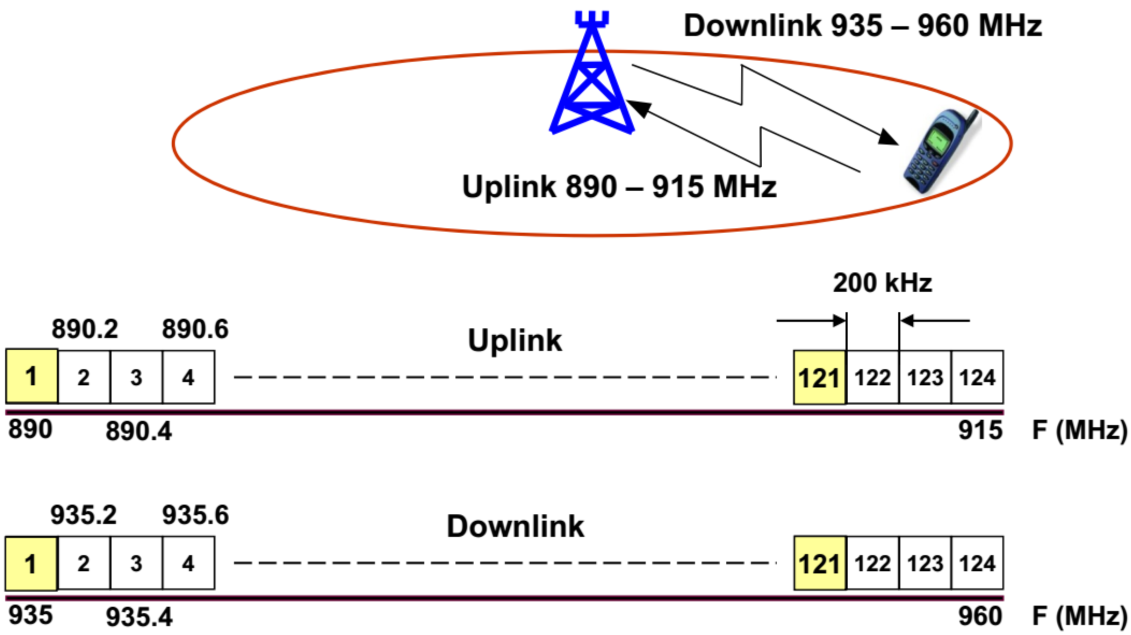


Figure 5. GSM−900 Frequency Allocation

4. Time Division Multiple Access TDMA

In TDMA, time is divided into intervals of regular length called frames, and then each frame is subdivided into time slots. Each user is assigned a slot number, and can transmit over the entire bandwidth during its slot within each frame.

This scheme is illustrated in Figure 6 for GSM−900 mobile system. It combines TDMA and FDD, requiring two frequencies to provide duplex operation, just as in FDMA. Time is divided into frames and each frame is further divided into 8 slots. Mobiles are allocated a pair of time slots, one at the uplink frequency and the other at the downlink frequency, chosen so that they do not coincide in time.

TDMA can also be used with TDD, by allocating half the slots to the uplink and half to the downlink, avoiding the need for frequency switching between transmission and reception.

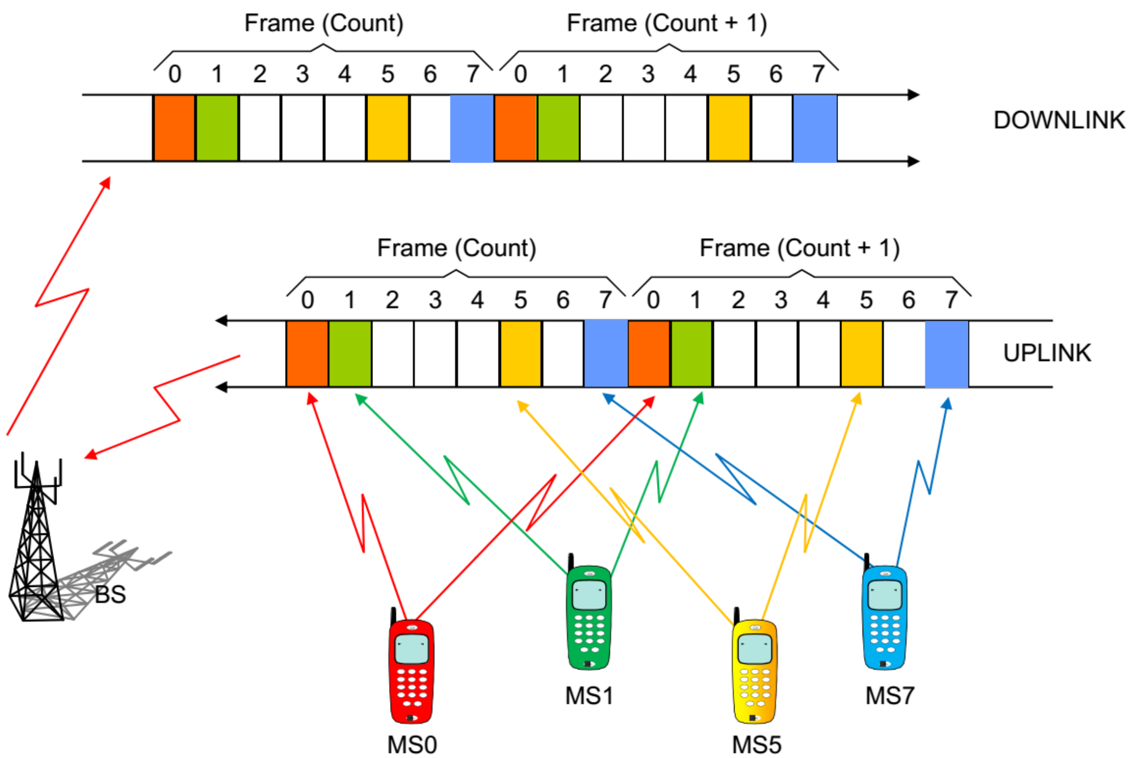


Figure 6: TDMA/FDMA for GSM-900 mobile system.

5. Code Division Multiple Access CDMA

Codes with orthogonality characteristics can be applied to the transmission to enable the use of code division multiplexing (CDM). Code division multiple access use exactly these codes to separate different users in code space and to enable access to a shared medium without interference.

In CDM, all channels use the same frequency band at the same time. Separation of users by codes, and guard spaces corresponds to the distance between codes (orthogonal codes). CDM offers good protection against interference, but needs high complexity of the receivers, and precise synchronization between sender and receiver. It was initially used in military applications. In mobile communications, CDMA is the designated multiplexing technique for UMTS/IMT-2000.

Conclusion

Advanced multiple access schemes were developed for advanced mobile communication systems. For Long-Term Evolution (LTE) system, orthogonal frequency-division multiplexing (OFDM) is used. It is a method of encoding digital data on multiple carrier frequencies, where a large number of closely spaced orthogonal sub-carrier signals are used to carry data on several parallel data streams or channels.

RF and Microwave system architectures

When designing and developing an RF transceiver for a wireless (RF and Microwave) system, we shall first determine what kind of architecture will be employed based on considerations of performance, cost, power consumption, and robust implementation. In this paragraph we are going to present the architectures of the RF and Microwave systems, their performance and technical issues.

We will focus on receiver architecture, performance and technical issues, because the receiver is usually the most critical component of a wireless system, having the overall purpose of reliably recovering the desired signal from a wide spectrum of transmitting sources, interference, and noise. The receiver picks up the modulated carrier signal from its antenna, then the task of its architecture is to process the incoming signal into useful information, adding minimal distortion. The performance of the receiver depends on the system design, circuit design, and working environment. The acceptable level of distortion or noise varies with the application. Noise and interference, which are unwanted signals that appear at the output of the receiver, set a lower limit on the usable signal level at the output. For the output signal to be useful, the signal power must be larger than the noise power by an amount specified by the required minimum signal-to-noise ratio. The minimum signal-to-noise ratio depends on the application, for example, 30 dB for a telephone line, 40 dB for a TV system, and 60 dB for a good music system.

We mentioned in the introduction of chapter 4 that a receiver in a modern RF and microwave system is formed from amplifiers, oscillators and mixers, which are like building blocks (Figure 1 of chapter 4), and it is apparent that the overall performance of the receiver depends on the characteristics of each individual building block. The objectives of the receiver system design are not only to make the designed receiver achieving the performance, current consumption, cost, and size as targeted, but also to define the specifications of each individual device in the receiver and thus to ensure the targeted receiver performance and other goals achieved.

History

Increased requirements for low power, small form factor, low cost and reduced cost in mobile communications has driven academia and industry to revive the practice of the direct conversion receiver since 1980. Long abandoned in favor of the mature superheterodyne receiver, direct conversion has emerged over the last few decades, or so thanks to improved semiconductor process technologies and astute design techniques.

Very much like its well established superheterodyne receiver counterpart, first introduced in 1918 by Armstrong, the origins of the direct conversion receiver (DCR) date back to the first half of last century when a single down−conversion receiver was first described by F.M. Colebrook in 1924, and the term homodyne was applied. To overcome some issues of direct conversion architecture, a modified architecture referred to as low IF architecture is then created. Some wireless communication receivers especially based on the CMOS technology started to employ this architecture to cope with the flicker noise and the DC offset problems of the direct conversion.

It is important to say that over the last few decades, the drive of the wireless market and enabling monolithic integration technology have triggered research activities on direct conversion receivers, which integrated with the remaining analog and digital sections of the transceiver, have the potential to reach the "one−chip radio" goal. Besides, it favors multi−mode, multi−standard applications and thereby constitutes another step towards software radio.

Receiver Considerations

The receiver system considerations are:

Minimum Discernible Signal (MDS) is the input that gives an output signal–to–noise ratio of 0 dB (where the signal power is equal to the noise power).

$$MDS[\text{watts}] = kT_0BF$$
$$MDS[\text{dBm}] = -174\ [\text{dBm/Hz}] + 10\log B + NF$$

Sensitivity S_{min}: Receiver sensitivity is the smallest signal we can put in to get a specified minimum SNR out of the receiver. The sensitivity of a receiver is directly related to the overall noise figure of the receiver. Then the noise generated inside the receiver is what limits the sensitivity.

$$S_{min}[\text{dBm}] = -174\ [\text{dBm/Hz}] + 10\log B + NF + SNR_{min}$$

For example, the required SNR_{min} at the output of the RF section of the CDMA mobile receiver is -1 dBm, and the receiver bandwidth is 1.25 MHz. Assuming that the overall noise figure NF is 10 dB, the receiver sensitivity is thus

$$S_{min} = -174\ + 10\log(1.25 \times 10^6) + 10 - 1 = -104.1\ \text{dBm}$$

$$S_{min} = 3.9 \times 10^{-11}\ \text{mW} = 0.04\ \text{pW}$$

In fact, it is more convenient for the RF receiver system design to employ the receiver noise figure NF, than using the sensitivity. The sensitivity for the GSM 900 small mobile receiver is defined as -102 dBm, and the SNR_{min} is approximately 8 dB. If the receiver bandwidth is 250 kHz, the maximum acceptable noise figure for such receiver sensitivity is

$$NF = 174 + S_{min} - 10\log B - SNR_{min} = 174 - 102 - 54 - 8 = 10\ \text{dB}$$

Selectivity: Receiver selectivity is the ability to reject unwanted signals on adjacent channel frequencies. The channel filtering (IF filter) characteristics and the phase noise of the local oscillator dominate the receiver selectivity. Selectivity can be obtained by using a narrow bandpass filter at the RF stage of the receiver, but the bandwidth and cutoff requirements for such a filter are usually impractical to realize at RF frequencies. It is more effective to achieve selectivity by down–converting a relatively wide RF bandwidth around the desired signal, and using a sharp cutoff bandpass filter at the IF stage to select only the desired frequency band. In addition, many wireless systems use a number of narrow but closely spaced channels, which must be selected using a tuned local oscillator, while the IF passband is fixed.

Spurious Response Rejection: The receiver ability to reject undesirable channel responses is important in reducing interference. This can be accomplished by properly choosing the IF by frequency planning process, and using various filters, especially for superheterodyne receiver.

The frequency band allocation for a wireless system will have substantial influence to the frequency planning of a superheterodyne transceiver used in this system. The frequency planning is mainly to search and select intermediate frequencies, which should minimize the spurious response problem of the superheterodyne transceiver, and should provide the possibility of developing a transceiver with excellent performance. However, this is a very long and monotonous task even if the transmitter and receiver each has only one IF block as presented in Fig. 3.1. It is more difficult to develop a frequency plan for a full-duplex transceiver than for a half duplex one. In a full-duplex transceiver, the receiver and the transmitter operate simultaneously, and the signals in both sides and their harmonics and mixing products need be taken into account in the frequency planning. In this case, it becomes really essential to avoid that any low-order transmitter spurious appears in the receiver band.

A superheterodyne transceiver may contain the following fundamental signals: an UHF LO (local oscillator) signal, a reference oscillator signal, two or multiple VHF LO signals, two or multiple IF signals, a weak RF reception signal and a powerful transmission signal. The real problem is that these signals may produce a tremendous amount of mixing products and harmonics because most of the devices compromising the transceiver have certain nonlinear characteristics. In the frequency planning, we must analyze these potential mixing products and harmonics-i.e., locate their position and determine their strength. It is not too hard to locate spurs of different order, but it is really difficult to determine the strength of the mixing products and harmonics without proper nonlinear models of corresponding devices. Fortunately, we are able to roughly estimate spurious levels based on the orders of the mixing products and harmonics and based on the order being an even or odd number as well.

Intermodulation Rejection: The receiver has the tendency to generate its own on-channel interference from one or more RF signals. These interference signals are called intermodulation (IM) products. The linearity of a receiver, especially the third-order distortion, is the main factor to determine the intermodulation distortion (IMD) performance.

Frequency Stability: The stability of the local oscillator is important for low phase noise. Stabilized sources using dielectric resonators, phase-locked techniques, or synthesizers are commonly used (as seen in chapter 5, paragraph 3).

Radiation Emission: The local oscillator signal could leak through the mixer to the antenna and radiate into free space. This radiation causes interference and needs to be less than a certain level specified by the Federal Communications Commission (FCC). The allowable emission level within the receiver band for wireless mobile stations is in the range of -60 to -80 dBm.

Receiver architectures

A receiver can have different architectures, such as superheterodyne, direct conversion, or low IF. No matter which architecture is employed, the receiver should possess a well-defined function and performance. This means that the receivers operating in a wireless system may be different in their architectures, but they must have many common characteristics to achieve the unique performance specified in the wireless system standards.

In full-duplex systems, the receiver and the transmitter are operating simultaneously, but they operate in different frequency bands. It is apparent that in the full-duplex system, the transmission of the transmitter is a strong interference source to the receiver since the transmission power at the receiver antenna port can be 120 dB or even 130 dB stronger than the desired reception signal, but the transmission frequency is only tens of MHz away from that of the reception signal. The corresponding receiver must be able to work properly under the constant attack of the strong transmission interference. **This makes the design of the receiver operating in a full-duplex system much harder than the one working in the half-duplex system**.

Superheterodyne receiver

This is the most popular architecture used in RF communication transceivers. It is based on the heterodyne process of mixing an incoming RF signal with an offset frequency local oscillator (LO) in a nonlinear device (mixer) to generate an intermediate frequency (IF) signal. The nonlinear device executing the heterodyne process is called a frequency mixer or frequency converter. Figure 7 shows a block diagram of single-conversion superheterodyne receiver. The IF frequency is generally selected to be between the RF frequency and baseband. A midrange IF allows the use of sharper cutoff filters for improved selectivity, and higher IF gain through the use of an IF amplifier. Tuning is conveniently accomplished by varying the frequency of the local oscillator so that the IF frequency remains constant. The superheterodyne receiver is used in the majority of broadcast radios and televisions, radar systems, cellular telephone systems, and data communications systems.

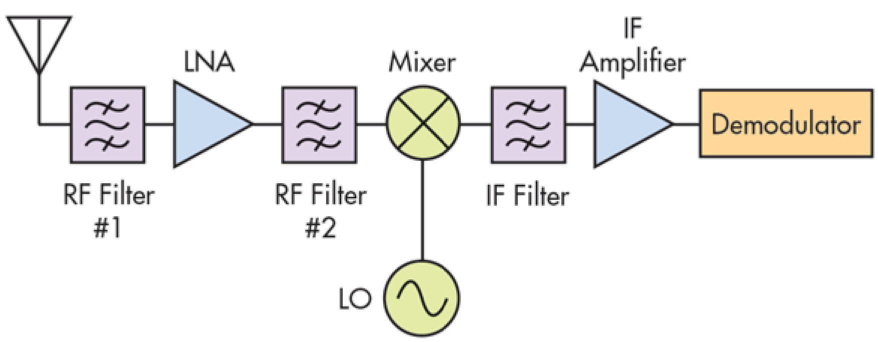


Figure 7: Simplified block diagram of single-conversion superheterodyne receiver.

Architecture description

- A received RF signal first enters a bandpass filter. This filter (RF Filter #1), often called a pre–select filter, rejects out–of–band signals.
- Next to the pre–select filter, a low–noise amplifier (LNA) performs the task of boosting the signal amplitude. This LNA is an extremely important component, as the overall noise figure of a superheterodyne receiver is highly dependent on the noise figure of the LNA.
- Another bandpass filter (RF Filter #2), known as an image–reject filter, follows the LNA. The purpose of this filter is to reject the unwanted image frequency band.
- A mixer then converts the RF signal to a lower–frequency IF signal. Both the RF signal and an LO signal enter the mixer, thereby generating the IF signal that appears at the mixer's output. The frequency of this IF signal is equal to the difference of the RF input signal's frequency and the LO signal's frequency.
- Following frequency down–conversion, bandpass filtering (IF Filter) is implemented in the IF stage to remove any unwanted signals.
- Next to the IF filter, an IF amplifier provides a significant amount of gain to the IF signal. Multiple amplifiers may be employed as well.
- The amplified IF signal is then demodulated, allowing the information to be processed.

Dual–conversion superheterodyne receiver

At microwave and millimeter wave frequencies it is often necessary to use two stages of down conversion to avoid problems due to LO stability. Such a dual–conversion superheterodyne receiver employs two local oscillators, two mixers, and two IF frequencies to achieve down–conversion to baseband, as shown in Figure 8.

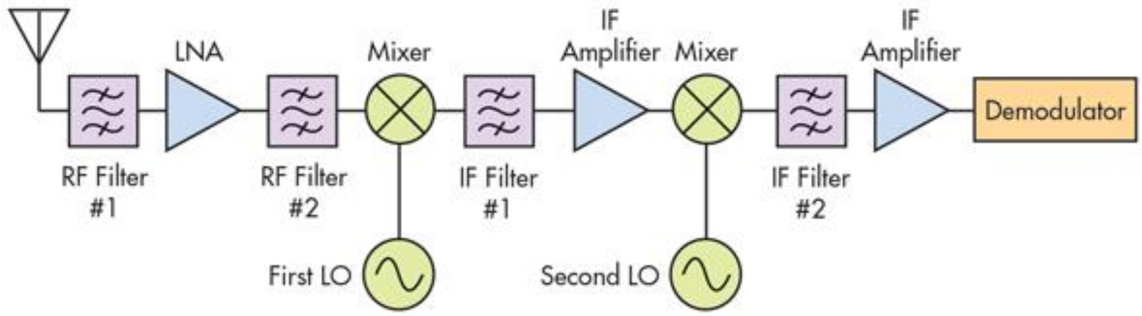


Figure 8: Simplified block diagram of dual–conversion superheterodyne receiver.

Image problem

It is apparent that the same IF can be generated by an incoming signal with a frequency either above or below the LO frequency (see chapter 5 – paragraph 4). Between these two frequencies, the one corresponding to the undesired signal is referred to as an image frequency, and the signal with this frequency is called as an image. The frequency difference between the desired signal and its image is twice the IF.

To suppress these unwanted image frequencies, a superheterodyne receiver requires filtering prior to the mixer. However, converting a high–frequency input signal to a very–low–frequency signal in just one conversion stage causes a problem: The image band will be very close in frequency to the actual RF input band. Thus, filtering these image frequencies becomes a challenge because of their close proximity to the actual RF input frequencies.

The dual–conversion superheterodyne receiver can remedy this problem. Employing two frequency–conversion stages makes it possible to space the image frequency band further away from the actual input frequency band. The increased separation between the input frequency band and the image frequency band is due to the higher first−IF frequency. Converting the RF input frequency to this first−IF frequency allows the image frequency band to be further away when compared with the approach of converting the input frequency to the lower final output frequency directly. Consequently, the task of filtering the image frequencies becomes much easier.

Image Rejection − Selectivity Trade−off

As stated, dual−conversion superheterodyne receivers have the benefit of a higher first−IF frequency, which allows them to achieve good image rejection. Furthermore, such receivers can also achieve excellent selectivity thanks to their lower second−IF frequency—greater selectivity is generally easier to obtain at lower frequencies. Thus, dual−conversion superheterodyne receivers can simultaneously achieve superior image rejection and selectivity.

Channel Filtering

The channel filtering of a superheterodyne receiver is performed in the IF blocks by means of passive filters with high selectivity. The reception channel tuning is often carried out through programming an RF synthesizer, and thus the frequency of each IF block can remain fixed. At fixed intermediate frequencies, it is relatively easier to obtain high and stable gains. The power consumption for achieving high gain at IF is significantly lower than that if the same gain is developed at RF. This is due to the fact that the channel filters effectively suppress strong unwanted signals or interferers before they are substantially amplified, and therefore a high dynamic range of IF amplifiers is not demanded. The high channel selectivity also helps to achieve higher receiver sensitivity since adequate gain prior to the channel filtering can be set for obtaining the best sensitivity but still not to saturate the later stage amplifiers.

Spurious Response Problem

The frequency band allocation for a wireless system will have substantial influence to the frequency planning of a superheterodyne transceiver used in this system. The frequency planning is mainly to search and select intermediate frequencies, which should minimize the spurious response problem of the superheterodyne transceiver, and should provide the possibility of developing a transceiver with excellent performance. However, this is a very long and monotonous task even for single–conversion superheterodyne receivers.

In a full–duplex transceiver, the receiver and the transmitter operate simultaneously, and the signals in both sides and their harmonics and mixing products need to be taken into account in the frequency planning. In this case, it becomes really essential to avoid that any low–order transmitter spurious appears in the receiver band.

A dual–conversion superheterodyne transceiver may contain the following fundamental signals (Figure 2, chapter 4):

- UHF LO signal,
- Reference oscillator signal,
- Two or multiple VHF LO signals,
- Two or multiple IF signals,
- Weak RF reception signal,
- Powerful transmission signal.

The real problem is that these signals may produce a tremendous amount of mixing products and harmonics because most of the devices compromising the transceiver have certain nonlinear characteristics. In the frequency planning, we must analyze these potential mixing products and harmonics–i.e., locate their position and determine their strength.

Direct−Conversion Receiver

A direct−conversion receiver (DCR) is another type of receiver architecture. Figure 9 shows a simplified block diagram of the principle of direct−conversion. Direct−conversion receivers convert an RF signal to a 0−Hz signal (baseband −BB− signal) in one stage. It is also known as a zero−IF receiver, because an RF signal is directly down−converted to a baseband signal without any intermediate frequency stages. The direct conversion receiver is also referred to as a homodyne when the LO is phase−locked with the carrier of the received signal $(f_{LO} = f_{RF})$. Direct−conversion receivers are generally considered low−cost solutions, mostly because they require few components. They need no IF, and thus the expensive IF passive filter (SAW filter) can be eliminated, and then the cost and size of the overall receiver are reduced.

In addition, direct−conversion receiver has several qualities making it very suitable for integrated−circuit (IC) designs, as well as multi−band, multi−standard operation. The configuration of the direct−conversion receiver looks simpler than that of the superheterodyne receiver, but its implementation is much more difficult since there are a number of technical challenges (as we will see later) in the direct conversion receiver, and system−on−chip is a suitable solution.

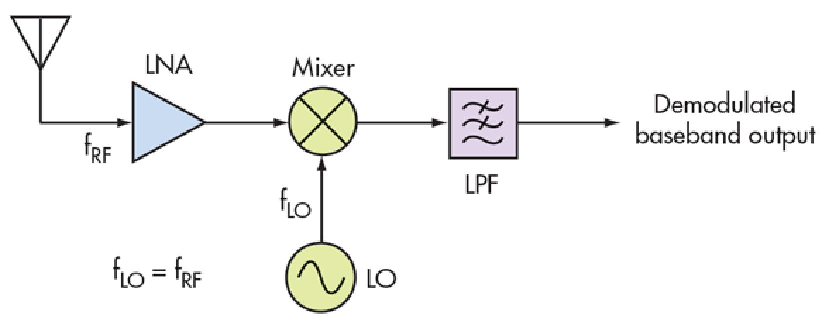


Figure 9: Block diagram of a direct−conversion receiver.

A more detailed block diagram of an RF direct conversion receiver is shown in Figure 10. In a full−duplex transceiver, a diplexer is used with the antenna to separate receiver and transmitter frequency bands. The function of the receiver bandpass filter (BPF) in the diplexer is to suppress the leakage power of the transmission and other out of receiver band interference.

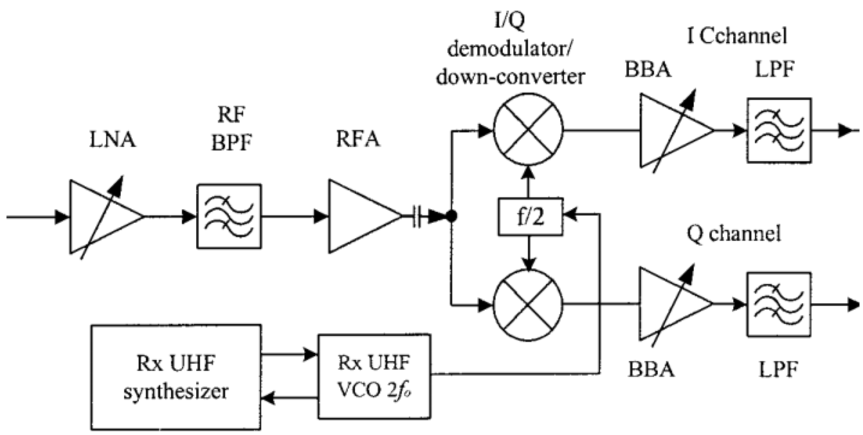


Figure 10: Detailed block diagram of an RF direct conversion receiver.

Architecture description

- The problem of the image has been eliminated, since the IF is zero and the image to the desired channel is the channel itself.
- The received signal after preselecting of the diplexer BPF[2] is amplified by an LNA, and it is further filtered by an RF filter and then amplified by an RFA.
- The filtered RF signal is then directly down−converted into I and Q[3] channel BB signals by an I/Q down−converter also called as quadrature demodulator.
- The UHF VCO in the receiver is running at twice of the receiver operating frequency. The VCO signal is applied to a frequency−by−two−divider. The two outputs of the divider with a 90° phase shift each other are used to drive the receiver I and Q down−converter, respectively.
- The I/Q signals can then be synchronously amplified and processed following the demodulation stage.

[2] The diplexer is not shown on the figure.
[3] I for in phase or the same phase, and Q for quadrature phase or phase difference of 90°.

- There exists a low–pass channel filter in each of I and Q channels. Unlike the superheterodyne receiver, the channel selectivity now mainly depends on the stop–band rejection of the low–pass filter without any passive band–pass filter assistance.
- The amplified and filtered BB analog signals in the I and Q channels are converted into digital signals by analog–to–digital converters (ADC), and the digital signals then pass through digital filters to further suppress nearby interferers and enhance the channel selectivity.

Technical issues

DC Offsets

In direct conversion, as the RF signal is directly converted to baseband in the receive chain, without any filtering other than RF band–selection, various phenomena contribute to the creation of DC signals, which directly appear as interfering signals in the band of interest, as shown in Figure 11.

Figure 11 (a): The LO may be conducted or radiated through an unintended path to the mixer's RF input port, thus effectively mixing with itself, producing an unwanted DC component at the mixer output. Worse still, this LO leakage may reach the LNA input, producing an even stronger result. This effect presents a high barrier against the integration of LO, mixer and LNA on a single silicon substrate, where numerous mechanisms can contribute to poor isolation. These include substrate coupling, ground bounce, bond wire radiation, and capacitive and magnetic coupling.

Figure 11 (b): A strong in–band interference signal, once amplified by the LNA, may find a path to the LO–input port of the mixer, thus once again producing self–mixing.

Figure 11 (c): Some amount of LO power will be conducted through the mixer and LNA (due to their non–ideal reverse isolation) to the antenna. The radiated power, appearing as an interferer to other receivers in the corresponding band, may violate emissions standards[4] of the given system. It is important to note that since $f_{LO} = f_{RF}$, the front–end filters do nothing to suppress this LO emission. Additionally, the radiated LO signal can then be reflected by buildings or moving objects and re–captured by the antenna. This effect, however, is not of significant importance compared to the LO self–mixing and blocking signal self–mixing.

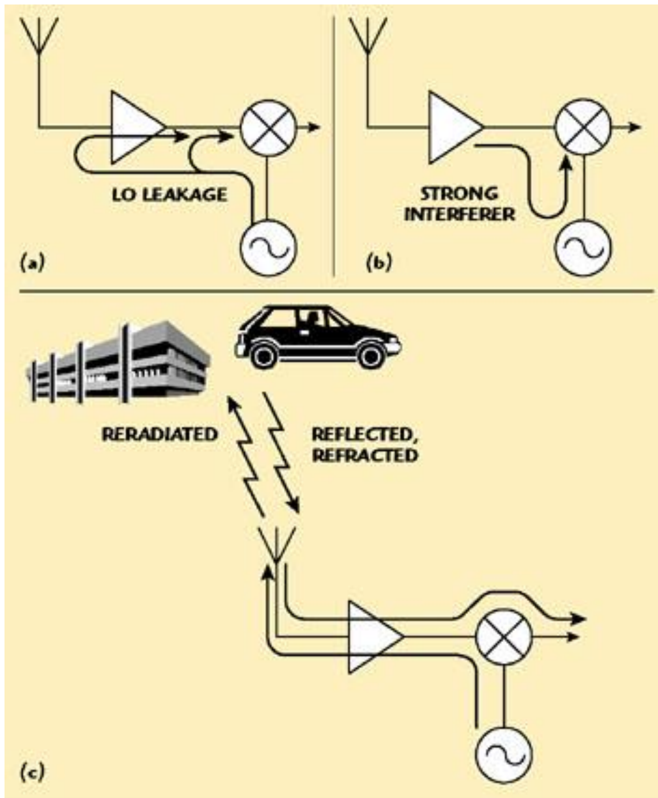


Figure 11: Causes of DC offset in DCR

Nonlinearities

The leakage of LO or RF signals to the opposite mixer port is not the only way in which unwanted DC can be produced. Any stage that exhibits even–order nonlinearity will also generate a DC output. This is the problem of second–order nonlinearity for the DCR.

Frequency–closed strong interferers and/or an interferer with an amplitude modulation (AM) can be turned into low–frequency in–channel interference

[4] The allowable emission level within the receiver band for mobile stations is in the range of -60 to -80 dBm.

products including the DC component by the second–order distortion. The in–channel bandwidth interferers may propagate in the BB block and then possibly deteriorate the DCR performance or even block the receiver.

More importantly, however, large blocking signals also cause DC in the direct conversion receiver, whether on a spurious frequency or not. The DC is produced at the mixer output and amplified by the BB stages. It is due primarily to second–order nonlinearity of the mixer, characterized by the second–order intercept point (IP2) and second–order intermodulation (IM2). It can be alleviated by extremely well–balanced circuit design.

In most systems, the third–order intermodulation is of importance as it usually falls in–band, in the vicinity of the signals of interest, and is characterized by the third–order intercept point (IP3). In direct conversion, second–order nonlinearity becomes critical, as it produces baseband signals, which now appear as interfering signals in the down–converted desired signal. IM2 is measured by the IP2. IP2 is defined in the same manner as IP3 (review chapter 3: Noise and Nonlinear Distortion in RF and Microwave systems).

Flicker noise

Low frequency noise becomes a great concern in a DCR, as significant gain is allocated to BB stages after the mixer. Weak signal levels of a few millivolts in baseband are still very vulnerable to noise. This requires stronger RF stage gain to alleviate the poor noise figure of baseband blocks, but of course this must be traded against the linearity problems, just described, that accompany higher RF gain.

Flicker noise, or $1/f$ noise, is the major baseband noise contributor. The flicker noise gets higher when the frequency goes lower. For a DCR, the gain for the desired signal before it is converted to a BB signal is only around 25 dB. Flicker noise contributed by the converter, BB amplifiers, and BB filter of the DCR may have a visibly deteriorative effect on the desired signal.

In practice, flicker noise becomes an issue for MOS devices more than bipolar. The flicker noise is much lower for the integrated circuits in terms of SiGe and BiCMOS technologies. The DCR based on these semiconductor technologies makes it possible to achieve high receiver sensitivity.

I/Q Mismatches

Due to the high frequency of the LO, it is not possible to implement the I/Q demodulator digitally. An analog I/Q demodulator exhibits gain and phase imbalances between the I and Q branches, as well as the introduction of DC offsets. Such imperfections distort the recovered constellation. The imbalance requirements in the magnitude and the phase of I and Q signals depend on the modulation scheme and the system protocol.

In the DCR, the received RF signal after amplifying in the RF front-end is directly down-converted into I and Q BB signals. These two signals propagate and are amplified in separate I and Q paths or referred to as I and Q channels. The gain in the both analog BB paths may vary more than 80 dB. In addition, the I and Q BB signals pass through the low-pass channel filter in their own channel separately. Generally speaking, it is difficult to keep the I and Q BB signals having perfect balance in their magnitude and phase when they pass through two completely separate paths even if using the-state-of-the-art highly integrated RF circuits. A wide range of the gain control in both BB channels makes it even more difficult to maintain the balance between these two signals.

Conclusion

The direct conversion receiver is an attractive yet challenging receiving technique. It has been successfully applied to devices such as pagers, mobile phones, PC and internet wireless connectivity cards, and satellite receivers, etc. in a variety of process technologies and increasing integration levels.

Low-IF Receiver

The main advantage of the low IF receiver over the direct conversion one is that this architecture has no DC offset problem because the desired signal is off the

DC by the IF. Properly choosing the low IF can remove the low-frequency interference product that resulted from the down-converter second-order nonlinearity demodulating AM interferer out of the desired signal bandwidth. In addition, the low IF architecture is also able to significantly reduce the near DC flicker noise impact on the receiver performance. This architecture, thus, is quite attractive for the highly integrated transceivers based on the CMOS technology since the flicker noise in the CMOS circuits is much higher than that in the GaAs, BiCMOS, and SiGe circuits. The main drawback of this architecture is the image rejection problem because the IF is so low that the image band interference is close to the desired signal and it is difficult to separate the image from the desired signal by using any passive UHF BPF without degrading the receiver sensitivity.

Direct RF Sampling Receiver

Performance advances in analog-to-digital converters (ADCs) over the past few years have enabled them to directly digitize signals at RF frequencies. While operating at high frequencies, they can maintain low noise and good linearity.

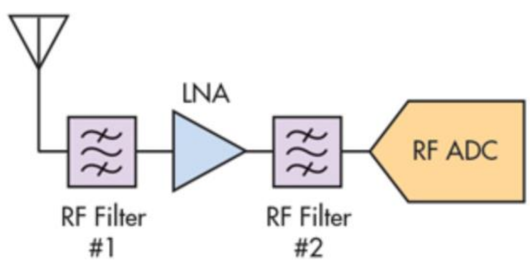


Figure 12: Direct RF Sampling Receiver

These advanced ADCs have enabled the direct RF-sampling receiver shown in Figure 12. RF ADC can directly sample signals at RF frequencies, thereby eliminating the need to convert an RF input signal to a lower frequency. This technique obviously represents a significant change from the traditional superheterodyne receiver. Now the promise of true software-defined radios (SDRs) can be fulfilled because of the capabilities of this new generation of ADCs.

A direct RF–sampling receiver basically consists of a low–noise amplifier (LNA), the required filters, and the ADC. The ADC digitizes the RF signal directly and sends it to a processor. Because frequency conversion is not required, the overall design of a direct RF–sampling receiver is much simpler in comparison to other receiver architectures. The ADC essentially replaces the mixer, oscillator, and the entire IF signal chain found in superheterodyne, direct conversion, low IF configurations. Because they have fewer components, direct RF–sampling receivers can be built in smaller form factors.

Performance Tradeoffs

Direct–conversion receivers do offer some advantages versus their superheterodyne counterparts. For example, they avoid the image–frequency problem. They also require fewer components, leading to simpler, lower–cost solutions. In fact, direct–conversion receivers are available as integrated circuits (ICs). However, direct–conversion receivers have also several disadvantages. One of the biggest problems surrounds dc offset issues. For example, LO leakage mixing with the actual LO signal may inflict large dc offset errors on the output signal. LO leakage can be minimized by maintaining high isolation between a mixer's LO and RF ports. Another way to resolve dc offset problems is by converting the input signal to a frequency that is close to, but not exactly, dc. This implementation is known as a low–IF receiver. However, using such an approach reintroduces the image–frequency problem. Superheterodyne receivers offer a number of benefits, such as the ability to achieve unmatched selectivity and sensitivity. They also are immune to the dc problems that occur in direct–conversion receivers. However, as mentioned, superheterodyne receivers have to cope with the image–frequency band. In addition, they are generally much larger in size because they require components like bulky filters. Therefore, these receivers typically are not suitable to be designed as ICs.

Receiver Dynamic Range

The dynamic range of an RF receiver is the input power range at its antenna port over which the data error rate does not exceed a specified value. The lower end of this range depends on the receiver sensitivity level, and the upper end is determined by the allowed maximum input power at which the data error does not exceed the specified value.

As an example, the minimum requirements of the maximum input power at the receiver antenna port and dynamic range for different mobile systems are presented in Table 2.

Systems	Maximum Input Power (dBm)	Minimum Dynamic Range (dB)
AMPS	N/A	> 96
CDMA 800	-25	> 79
CDMA 1900	-25	> 79
EDGE	-26	> 72
GPRS	-26	> 73
GSM 900	-15	> 87
GSM 1800	-23	> 79
PHS	-21	> 76
TDMA	-25	> 85
WCDMA	-25	> 81.7

Table 2: Maximum input signal and minimum dynamic range for different mobile systems

To be able to operate over such a wide dynamic range, a receiver commonly employs an automatic gain control (AGC) system. The AGC range is usually wider than the receiver dynamic range. It must also cover the receiver gain variation resulting from device production processing deviation, temperature, and voltage variations. The minimum dynamic range of a CDMA mobile receiver, for instance, is 79 dB, but an AGC system may need a 100 dB control range to cover the gain possible variations and the dynamic range design margin.

The main portion of a receiver AGC system used in the mobile receiver is in the digital base-band and the DSP. The controlled objects of the AGC system are in the RF analog section of the receiver, and they are: LNA, IF variable gain amplifier (VGA), and BB VGA. A block diagram of the CDMA receiver AGC system is shown in Figure 13. This system consists of a gain step-controlled front-end, an IF VGA, ADCs, SINC filters, a CDMA core, an AGC algorithm stored in the DSP, and a 10 bit PDM DAC.

In the CDMA mobile receiver, the receiver AGC system function is not only to maintain the receiver chain properly operating over the dynamic range and keeping the level at the ADC input constant, but also to measure the received signal strength through the receiver signal strength indicator (RSSI) and then to determine the transmission power of the transmitter.

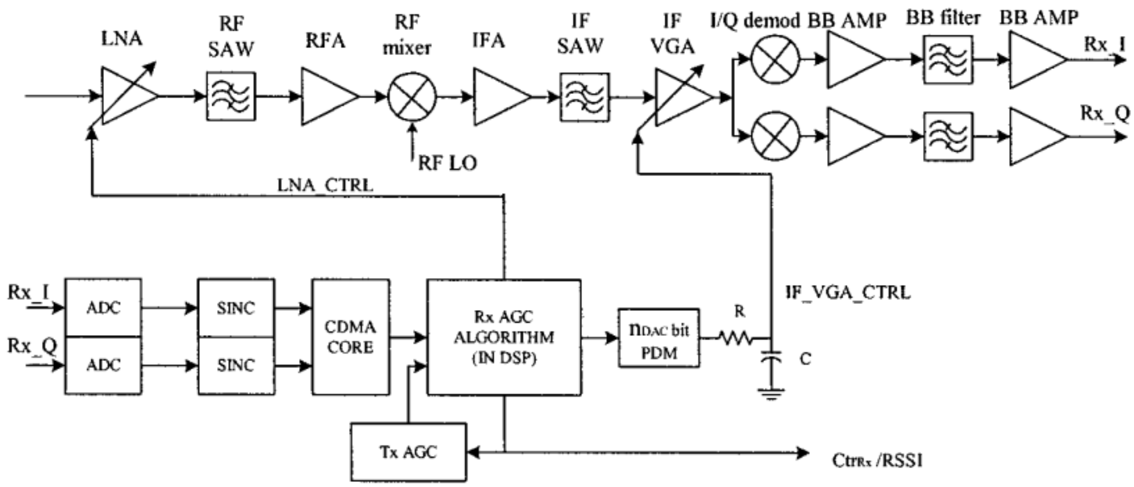


Figure 13: block diagram of a CDMA receiver AGC system

Conclusion

It is obvious that receivers are critical to the performance of any communications system. Many factors must be taken into account during the development process. Performance characteristics like noise figure, selectivity, and sensitivity must be examined closely.

The superheterodyne receiver is a traditional implementation found in many applications. However, direct conversion receivers are good fits in certain situations, too. Lastly, the direct-RF sampling technique, the most recent receiver implementation, is a technology to keep an eye on. It holds a great amount of promise for future wireless systems.

Test

10 marks for each question, Successful test mark is 60/100

Choose the correct answer.

1. A superheterodyne receiver with an IF of 450 kHz is tuned to a signal at 1200 kHz. The image frequency is:
   - **a.** 750 kHz
   - **b.** 900 kHz
   - **c.** 1650 kHz
   - **d.** 2100 kHz
2. The image frequency of a superheterodyne receiver is
   - **a.** created within the receiver itself
   - **b.** due to insufficient adjacent channel rejection
   - **c.** not rejected by the IF tuned circuits
   - **d.** independent of the frequency to which the receiver is tuned
3. In which radio architecture does image rejection become of concern?
   - **a.** In any IF architecture (superheterodyne or low IF).
   - **b.** In superheterodyne architecture only.
   - **c.** In low IF architecture only.
   - **d.** In direct−conversion architecture.
4. In which radio architecture does DC offset become of concern?
   - **a.** In any IF architecture (superheterodyne or low IF).
   - **b.** In superheterodyne architecture.
   - **c.** In low IF architecture.
   - **d.** In direct−conversion architecture.
5. Why are DCR receiver ICs preferably made using SiGe or BiCMOS?
   - **a.** Because the Flicker noise is much lower for ICs using SiGe or BiCMOS.
   - **b.** Because the Flicker noise is much higher for ICs using SiGe or BiCMOS.
   - **c.** Because ICs using SiGe or BiCMOS are low cost.
   - **d.** Because ICs using SiGe or BiCMOS increase integration level.

6. In full–duplex channel

   a. the channel is carrying signals in one direction at the same time.

   b. signals are transmitted in only one direction

   c. signals are transmitted and received simultaneously.

   d. signals are transmitted and received but only one at a time.

7. Minimum discernible signal (MDS) is defined for

   a. input signal power equal to the noise power.

   b. output signal power equal to the noise power.

   c. input signal–to–noise ratio of 0 dB

   d. a specified minimum signal–to–noise ratio

8. Receiver sensitivity is defined as

   a. the input power for output signal–to–noise ratio of 0 dB

   b. the output power for input signal–to–noise ratio of 0 dB

   c. the input power for a specified minimum output signal–to–noise ratio

   d. the output power for a specified minimum input signal–to–noise ratio

9. High stable local oscillator is required for

   a. superheterodyne architecture.

   b. low IF.

   c. direct RF sampling architecture.

   d. direct–conversion architecture.

10. In IF architectures, lower IF is better for

   a. higher receiver selectivity performance

   b. higher image rejection

   c. higher intermodulation distortion (IMD) performance.

   d. higher receiver sensitivity

Question number	Answer	Feedback
1	d	Revise superheterodyne receiver
2	c	Revise superheterodyne receiver
3	a	Revise receiver architectures
4	d	Revise receiver architectures
5	a	Revise direct–conversion receiver (DCR).
6	c	Revise types of communication channels
7	b	Revise receiver considerations
8	c	Revise receiver considerations
9	d	Revise receiver architectures.
10	a	Revise receiver considerations