1

Radio Communications: System Concepts, Propagation and Noise

A critical point for the development of radio communications and related applications was the invention of the 'super-heterodyne' receiver by Armstrong in 1917. This system was used to receive and demodulate radio signals by down-converting them in a lower intermediate frequency (IF). The demodulator followed the IF amplification and filtering stages and was used to extract the transmitted voice signal from a weak signal impaired by additive noise. The super-heterodyne receiver was quickly improved to demodulate satisfactorily very weak signals buried in noise (high sensitivity) and, at the same time, to be able to distinguish the useful signals from others residing in neighbouring frequencies (good selectivity). These two properties made possible the development of low-cost radio transceivers for a variety of applications. AM and FM radio were among the first popular applications of radio communications. In a few decades packet radios and networks targeting military communications gained increasing interest. Satellite and deep-space communications gave the opportunity to develop very sophisticated radio equipment during the 1960s and 1970s. In the early 1990s, cellular communications and wireless networking motivated a very rapid development of low-cost, low-power radios which initiated the enormous growth of wireless communications.

The biggest development effort was the cellular telephone network. Since the early 1960s there had been a considerable research effort by the AT&T Bell Laboratories to develop a cellular communication system. By the end of the 1970s the system had been tested in the field and at the beginning of the 1980s the first commercial cellular systems appeared. The increasing demand for higher capacity, low cost, performance and efficiency led to the second generation of cellular communication systems in the 1990s. To fulfill the need for high-quality bandwidth-demanding applications like data transmission, Internet, web browsing and video transmission, 2.5G and 3G systems appeared 10 years later.

Along with digital cellular systems, wireless networking and wireless local area networks (WLAN) technology emerged. The need to achieve improved performance in a harsh propagation environment like the radio channel led to improved transmission technologies like spread spectrum and orthogonal frequency division multiplexing (OFDM). These technologies were

put to practice in 3G systems like wideband code-division multiple access (WCDMA) as well as in high-speed WLAN like IEEE 802.11a/b/g.

Different types of digital radio system have been developed during the last decade that are finding application in wireless personal area networks (WPANs). These are Bluetooth and Zigbee, which are used to realize wireless connectivity of personal devices and home appliances like cellular devices and PCs. Additionally, they are also suitable for implementing wireless sensor networks (WSNs) that organize in an ad-hoc fashion. In all these, the emphasis is mainly on short ranges, low transmission rates and low power consumption.

Finally, satellite systems are being constantly developed to deliver high-quality digital video and audio to subscribers all over the world.

The aims of this chapter are twofold. The first is to introduce the variety of digital radio systems and their applications along with fundamental concepts and challenges of the basic radio transceiver blocks (the radio frequency, RF, front-end and baseband parts). The second is to introduce the reader to the technical background necessary to address the main objective of the book, which is the design of RF and baseband transmitters and receivers. For this purpose we present the basic concepts of linear systems, stochastic processes, radio propagation and channel models. Along with these we present in some detail the basic limitations of radio electronic systems and circuits, noise and nonlinearities. Finally, we introduce one of the most frequently used blocks of radio systems, the phase-locked loop (PLL), which finds applications in a variety of subsystems in a transmitter/receiver chain, such as the local oscillator, the carrier recovery and synchronization, and coherent detection.

1.1 Digital Radio Systems and Wireless Applications

The existence of a large number of wireless systems for multiple applications considerably complicates the allocation of frequency bands to specific standards and applications across the electromagnetic spectrum. In addition, a number of radio systems (WLAN, WPAN, etc.) operating in unlicensed portions of the spectrum demand careful assignment of frequency bands and permitted levels of transmitted power in order to minimize interference and permit the coexistence of more than one radio system in overlapping or neighbouring frequency bands in the same geographical area.

Below we present briefly most of the existing radio communication systems, giving some information on the architectures, frequency bands, main characteristics and applications of each one of them.

1.1.1 Cellular Radio Systems

A cellular system is organized in hexagonal cells in order to provide sufficient radio coverage to mobile users moving across the cell. A base station (BS) is usually placed at the centre of the cell for that purpose. Depending on the environment (rural or urban), the areas of the cells differ. Base stations are interconnected through a high-speed wired communications infrastructure. Mobile users can have an uninterrupted session while moving through different cells. This is achieved by the MTSOs acting as network controllers of allocated radio resources (physical channels and bandwidth) to mobile users through the BS. In addition, MTSOs are responsible for routing all calls associated with mobile users in their area.

Second-generation (2G) mobile communications employed digital technology to reduce cost and increase performance. Global system for mobile communications (GSM) is a very

successful 2G system that was developed and deployed in Europe. It employs Gaussian minimum shift keying (MSK) modulation, which is a form of continuous-phase phase shift keying (PSK). The access technique is based on time-division multiple access (TDMA) combined with slow frequency hopping (FH). The channel bandwidth is 200 kHz to allow for voice and data transmission.

IS-95 (Interim standard-95) is a popular digital cellular standard deployed in the USA using CDMA access technology and binary phase-shift keying (BPSK) modulation with 1.25 MHz channel bandwidth. In addition, IS-136 (North American Digital Cellular, NADC) is another standard deployed in North America. It utilizes 30 kHz channels and TDMA access technology.

2.5G cellular communication emerged from 2G because of the need for higher transmission rates to support Internet applications, e-mail and web browsing. General Packet Radio Service (GPRS) and Enhanced Data Rates for GSM Evolution (EGDE) are the two standards designed as upgrades to 2G GSM. GPRS is designed to implement packet-oriented communication and can perform network sharing for multiple users, assigning time slots and radio channels [Rappaport02]. In doing so, GPRS can support data transmission of 21.4 kb/s for each of the eight GSM time slots. One user can use all of the time slots to achieve a gross bit rate of $21.4 \times 8 = 171.2$ kb/s.

EDGE is another upgrade of the GSM standard. It is superior to GPRS in that it can operate using nine different formats in air interface [Rappaport02]. This allows the system to choose the type and quality of error control. EDGE uses 8-PSK modulation and can achieve a maximum throughput of 547.2 kb/s when all eight time slots are assigned to a single user and no redundancy is reserved for error protection. 3G cellular systems are envisaged to offer high-speed wireless connectivity to implement fast Internet access, Voice-over-Internet Protocol, interactive web connections and high-quality, real-time data transfer (for example music).

UMTS (Universal Mobile Telecommunications System) is an air interface specified in the late 1990s by ETSI (European Telecommunications Standards Institute) and employs WCDMA, considered one of the more advanced radio access technologies. Because of the nature of CDMA, the radio channel resources are not divided, but they are shared by all users. For that reason, CDMA is superior to TDMA in terms of capacity. Furthermore, each user employs a unique spreading code which is multiplied by the useful signal in order to distinguish the users and prevent interference among them. WCDMA has 5 MHz radio channels carrying data rates up to 2 Mb/s. Each 5 MHz channel can offer up to 350 voice channels [Rappaport02].

1.1.2 Short- and Medium-range Wireless Systems

The common characteristic of these systems is the range of operation, which is on the order of 100 m for indoor coverage and 150–250 m for outdoor communications. These systems are mostly consumer products and therefore the main objectives are low prices and low energy consumption.

1.1.2.1 Wireless Local Area Networks

Wireless LANs were designed to provide high-data-rate, high-performance wireless connectivity within a short range in the form of a network controlled by a number of central points (called access points or base stations). Access points are used to implement communication between two users by serving as up-link receivers and down-link transmitters. The geographical area

of operation is usually confined to a few square kilometres. For example, a WLAN can be deployed in a university campus, a hospital or an airport.

The second and third generation WLANs proved to be the most successful technologies. IEEE 802.11b (second generation) operates in the 2.4 GHz ISM (Industral, Scientific and Medical) band within a spectrum of 80 MHz. It uses direct sequence spread spectrum (DSSS) transmission technology with gross bit rates of 1, 2, 5 and 11 Mb/s. The 11 Mb/s data rate was adopted in late 1998 and modulates data by using complementary code keying (CCK) to increase the previous transmission rates. The network can be formulated as a centralized network using a number of access points. However, it can also accommodate peer-to-peer connections.

The IEEE 802.11a standard was developed as the third-generation WLAN and was designed to provide even higher bit rates (up to 54 Mb/s). It uses OFDM transmission technology and operates in the 5 GHz ISM band. In the USA, the Federal Communications Commission (FCC) allocated two bands each 100 MHz wide (5.15–5.25 and 5.25–5.35 GHz), and a third one at 5.725–5.825 GHz for operation of 802.11a. In Europe, HIPERLAN 2 was specified as the standard for 2G WLAN. Its physical layer is very similar to that of IEEE 802.11a. However, it uses TDMA for radio access instead of the CSMA/CA used in 802.11a.

The next step was to introduce the 802.11g, which mostly consisted of a physical layer specification at 2.4 GHz with data rates matching those of 802.11a (up to 54 Mb/s). To achieve that, OFDM transmission was set as a compulsory requirement. 802.11g is backward-compatible to 802.11b and has an extended coverage range compared with 802.11a. To cope with issues of quality of service, 802.11e was introduced, which specifies advanced MAC techniques to achieve this.

1.1.2.2 WPANs and WSNs

In contrast to wireless LANs, WPAN standardization efforts focused primarily on lower transmission rates with shorter coverage and emphasis on low power consumption. Bluetooth (IEEE 802.15.1), ZigBee (IEEE 802.15.4) and UWB (IEEE 802.15.3) represent standards designed for personal area networking. Bluetooth is an open standard designed for wireless data transfer for devices located a few metres apart. Consequently, the dominant application is the wireless interconnection of personal devices like cellular phones, PCs and their peripherals. Bluetooth operates in the 2.4 GHz ISM band and supports data and voice traffic with data rates of 780 kb/s. It uses FH as an access technique. It hops in a pseudorandom fashion, changing frequency carrier 1600 times per second (1600 hops/s). It can hop to 80 different frequency carriers located 1 MHz apart. Bluetooth devices are organized in groups of two to eight devices (one of which is a master) constituting a piconet. Each device of a piconet has an identity (device address) that must be known to all members of the piconet. The standard specifies two modes of operation: asynchronous connectionless (ACL) in one channel (used for data transfer at 723 kb/s) and synchronous connection-oriented (SCO) for voice communication (employing three channels at 64 kb/s each).

A scaled-down version of Bluetooth is ZigBee, operating on the same ISM band. Moreover, the 868/900 MHz band is used for ZigBee in Europe and North America. It supports transmission rates of up to 250 kb/s covering a range of 30 m.

During the last decade, WSNs have emerged as a new field for applications of low-power radio technology. In WSN, radio modules are interconnected, formulating ad-hoc networks.

WSN find many applications in the commercial, military and security sectors. Such applications concern home and factory automation, monitoring, surveillance, etc. In this case, emphasis is given to implementing a complete stack for ad hoc networking. An important feature in such networks is multihop routing, according to which information travels through the network by using intermediate nodes between the transmitter and the receiver to facilitate reliable communication. Both Bluetooth and ZigBee platforms are suitable for WSN implementation [Zhang05], [Wheeler07] as they combine low-power operation with network formation capability.

1.1.2.3 Cordless Telephony

Cordless telephony was developed to satisfy the needs for wireless connectivity to the public telephone network (PTN). It consists of one or more base stations communicating with one or more wireless handsets. The base stations are connected to the PTN through wireline and are able to provide coverage of approximately 100 m in their communication with the handsets. CT-2 is a second-generation cordless phone system developed in the 1990s with extended range of operation beyond the home or office premises.

On the other hand, DECT (Digital European Cordless Telecommunications) was developed such that it can support local mobility in an office building through a private branch exchange (PBX) system. In this way, hand-off is supported between the different areas covered by the base stations. The DECT standard operates in the 1900 MHz frequency band. Personal handyphone system (PHS) is a more advanced cordless phone system developed in Japan which can support both voice and data transmission.

1.1.2.4 Ultra-wideband Communications

A few years ago, a spectrum of 7.5 GHz (3.1–10.6 GHz) was given for operation of ultra-wideband (UWB) radio systems. The FCC permitted very low transmitted power, because the wide area of operation of UWB would produce interference to most commercial and even military wireless systems. There are two technology directions for UWB development. Pulsed ultra-wideband systems (P-UWB) convey information by transmitting very short pulses (of duration in the order of 1 ns). On the other hand, multiband-OFDM UWB (MB-OFDM) transmits information using the OFDM transmission technique.

P-UWB uses BPSK, pulse position modulation (PPM) and amplitude-shift keying (ASK) modulation and it needs a RAKE receiver (a special type of receiver used in Spread Spectrum systems) to combine energy from multipath in order to achieve satisfactory performance. For very high bit rates (on the order of 500 Mb/s) sophisticated RAKE receivers must be employed, increasing the complexity of the system. On the other hand, MB-UWB uses OFDM technology to eliminate intersymbol interference (ISI) created by high transmission rates and the frequency selectivity of the radio channel.

Ultra-wideband technology can cover a variety of applications ranging from low-bit-rate, low-power sensor networks to very high transmission rate (over 100 Mb/s) systems designed to wirelessly interconnect home appliances (TV, PCs and consumer electronic appliances). The low bit rate systems are suitable for WSN applications.

P-UWB is supported by the UWB Forum, which has more than 200 members and focuses on applications related to wireless video transfer within the home (multimedia, set-top boxes,

DVD players). MB-UWB is supported by WiMedia Alliance, also with more than 200 members. WiMedia targets applications related to consumer electronics networking (PCs TV, cellular phones). UWB Forum will offer operation at maximum data rates of 1.35 Gb/s covering distances of 3 m [Geer06]. On the other hand, WiMedia Alliance will provide 480 Mb/s at distances of 10 m.

1.1.3 Broadband Wireless Access

Broadband wireless can deliver high-data-rate wireless access (on the order of hundreds of Mb/s) to fixed access points which in turn distribute it in a local premises. Business and residential premises are served by a backbone switch connected at the fixed access point and receive broadband services in the form of local area networking and video broadcasting.

LMDS (local multipoint distribution system) and MMDS (multichannel multipoint distribution services) are two systems deployed in the USA operating in the 28 and 2.65 GHz bands. LMDS occupies 1300 MHz bandwidth in three different bands around 28, 29 and 321 GHz and aims to provide high-speed data services, whereas MMDS mostly provides telecommunications services [Goldsmith05] (hundreds of digital television channels and digital telephony). HIPERACCESS is the European standard corresponding to MMDS.

On the other hand, 802.16 standard is being developed to specify fixed and mobile broadband wireless access with high data rates and range of a few kilometres. It is specified to offer 40 Mb/s for fixed and 15 Mb/s for mobile users. Known as WiMAX, it aims to deliver multiple services in long ranges by providing communication robustness, quality of service (QoS) and high capacity, serving as the 'last mile' wireless communications. In that capacity, it can complement WLAN and cellular access. In the physical layer it is specified to operate in bands within the 2–11 GHz frequency range and uses OFDM transmission technology combined with adaptive modulation. In addition, it can integrate multiple antenna and smart antenna techniques.

1.1.4 Satellite Communications

Satellite systems are mostly used to implement broadcasting services with emphasis on high-quality digital video and audio applications (DVB, DAB). The Digital Video Broadcasting (DVB) project specified the first DVB-satellite standard (DVB-S) in 1994 and developed the second-generation standard (DVB-S2) for broadband services in 2003. DVB-S3 is specified to deliver high-quality video operating in the 10.7–12.75 GHz band. The high data rates specified by the standard can accommodate up to eight standard TV channels per transponder. In addition to standard TV, DVB-S provides HDTV services and is specified for high-speed Internet services over satellite.

In addition to DVB, new-generation broadband satellite communications have been developed to support high-data-rate applications and multimedia in the framework of fourth-generation mobile communication systems [Ibnkahla04].

Direct-to-Home (DTH) satellite systems are used in North America and constitute two branches: the Broadcasting Satellite Service (BSS) and the Fixed Satellite Service (FSS). BSS operates at 17.3–17.8 GHz (uplink) and 12.2–12.7 GHz (downlink), whereas the bands for FSS are 14–14.5 and 10.7–11.2 GHz, respectively.

Finally, GPS (global positioning satellite) is an ever increasing market for providing localization services (location finding, navigation) and operates using DSSS in the 1500 MHz band.

1.2 Physical Layer of Digital Radio Systems

Radio receivers consist of an RF front-end, a possible IF stage and the baseband platform which is responsible for the detection of the received signal after its conversion from analogue to digital through an A/D converter. Similarly, on the transmitter side, the information signal is digitally modulated and up-converted to a radio frequency band for subsequent transmission.

In the next section we use the term 'radio platform' to loosely identify all the RF and analogue sections of the transmitter and the receiver.

1.2.1 Radio Platform

Considering the radio receiver, the main architectures are the super-heterodyne (SHR) and the direct conversion receiver (DCR). These architectures are examined in detail in Chapter 3, but here we give some distinguishing characteristics as well as their main advantages and disadvantages in the context of some popular applications of radio system design. Figure 1.1 illustrates the general structure of a radio transceiver. The SHR architecture involves a mixing stage just after the low-noise amplifier (LNA) at the receiver or prior to the transmitting medium-power and high-power amplifiers (HPA). Following this stage, there is quadrature mixing bringing the received signal down to the baseband. Following mixers, there is variable gain amplification and filtering to increase the dynamic range (DR) and at the same time improve selectivity.

When the local oscillator (LO) frequency is set equal to the RF input frequency, the received signal is translated directly down to the baseband. The receiver designed following this approach is called Direct conversion Receiver or zero-IF receiver. Such an architecture eliminates the IF and the corresponding IF stage at the receiver, resulting in less hardware but, as we will see in Chapter 3, it introduces several shortcomings that can be eliminated with careful design.

Comparing the two architectures, SHR is advantageous when a very high dynamic range is required (as for example in GSM). In this case, by using more than one mixing stage, amplifiers with variable gain are inserted between stages to increase DR. At the same time, filtering inserted between two mixing stages becomes narrower, resulting in better selectivity [Schreir02].

Furthermore, super-heterodyne can be advantageous compared with DCR when large in-band blocking signals have to be eliminated. In DCR, direct conversion (DC) offset would change between bursts, requiring its dynamic control [Tolson99].

Regarding amplitude and phase imbalances of the two branches, In-phase (I-phase) and Q-phase considerably reduce the image rejection in SHR. In applications where there can be no limit to the power of the neighbouring channels (like the ISM band), it is necessary to have an image rejection (IR) on the order of 60 dB. SHR can cope with the problem by suitable choice of IF frequencies [Copani05]. At the same time, more than one down-converting stage relaxes the corresponding IR requirements. On the other hand, there is no image band in DCR and hence no problem associated with it. However, in DCR, imbalances at the I–Q

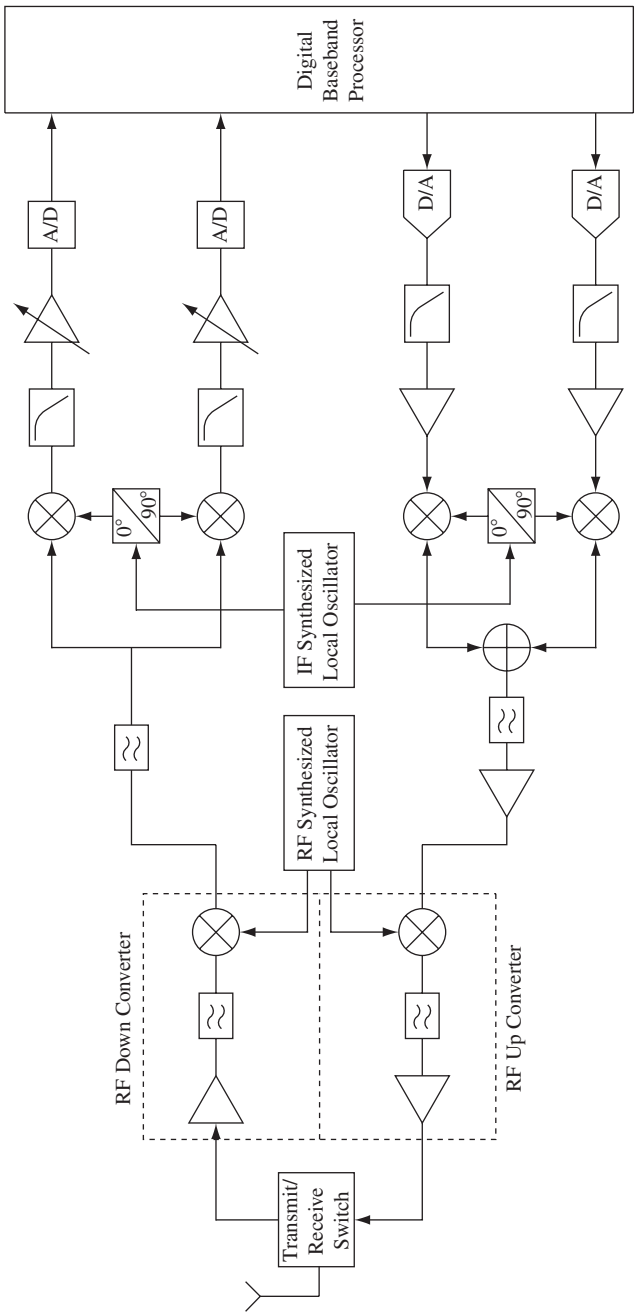

Figure 1.1 General structure of a radio transmitter and receiver

mixer create problems from the self-image and slightly deteriorate the receiver signal-to-noise ratio (SNR) [Razavi97]. This becomes more profound in high-order modulation constellations (64-QAM, 256-QAM, etc.)

On the other hand, DCR is preferred when implementation cost and high integration are the most important factors. For example, 3G terminals and multimode transceivers frequently employ the direct conversion architecture. DC offset and $1/f$ noise close to the carrier are the most frequent deficiencies of homodyne receivers, as presented in detail in Chapter 3. Furthermore, second-order nonlinearities can also create a problem at DC. However, digital and analogue processing techniques can be used to eliminate these problems.

Considering all the above and from modern transceiver design experience, SHR is favoured in GSM, satellite and millimetre wave receivers, etc. On the other hand, DCR is favoured in 3G terminals, Bluetooth and wideband systems like WCDMA, 802.11a/b/g, 802.16 and UWB.

1.2.2 Baseband Platform

The advent of digital signal processors (DSP) and field-programmable gate arrays (FPGAs), dramatically facilitated the design and implementation of very sophisticated digital demodulators and detectors for narrowband and wideband wireless systems. 2G cellular radio uses GMSK, a special form of continuous-phase frequency-shift keying (CPFSK). Gaussian minimum-shift keying (GMSK) modem (modulator–demodulator) implementation can be fully digital and can be based on simple processing blocks like accumulators, correlators and look-up tables (LUTs) [Wu00], [Zervas01]. FIR (Finite Impulse Response) filters are always used to implement various forms of matched filters. Coherent demodulation in modulations with memory could use more complex sequential receivers implementing the Viterbi algorithm.

3G cellular radios and modern WLAN transceivers employ advanced transmission techniques using either spread spectrum or OFDM to increase performance. Spread spectrum entails multiplication of the information sequence by a high-bit-rate pseudorandom noise (PN) sequence operating at speeds which are multiples of the information rate. The multiple bandwidth of the PN sequence spreads information and narrowband interference to a band with a width equal to that of the PN sequence. Suitable synchronization at the receiver restores information at its original narrow bandwidth, but interference remains spread due to lack of synchronization. Consequently, passing the received signal plus spread interference through a narrow band filter corresponding to the information bandwidth reduces interference considerably. In a similar fashion, this technique provides multipath diversity at the receiver, permitting the collection and subsequent constructive combining of the main and the reflected signal components arriving at the receiver. This corresponds to the RAKE receiver principle, resembling a garden rake that is used to collect leaves. As an example, RAKE receivers were used to cope with moderate delay spread and moderate bit rates (60 ns at the rate of 11 Mb/s [VanNee99]. To face large delay spreads at higher transmission rates, the RAKE receiver was combined with equalization. On the other hand, OFDM divides the transmission bandwidth into many subchannels, each one occupying a narrow bandwidth. In this way, owing to the increase in symbol duration, the effect of dispersion in time of the reflected signal on the receiver is minimized. The effect of ISI is completely eliminated by inserting a guard band in the resulting composite OFDM symbol. Fast Fourier transform (FFT) is an efficient way to produce (in the digital domain) the required subcarriers over which the information will be embedded. In practice, OFDM is used in third-generation WLANs, WiMAX and DVB to eliminate ISI.

From the above discussion it is understood that, in modern 3G and WLAN radios, advanced digital processing is required to implement the modem functions which incorporate transmission techniques like spread spectrum and OFDM. This can be performed using DSPs [Jo04], FPGAs [Chugh05], application-specific integrated circuits (ASICs) or a combination of them all [Jo04].

1.2.3 Implementation Challenges

Many challenges to the design and development of digital radio systems come from the necessity to utilize the latest process technologies (like deep submicron complementary metal-oxide semiconductor, CMOS, processes) in order to save on chip area and power consumption. Another equally important factor has to do with the necessity to develop multistandard and multimode radios capable of implementing two or more standards (or more than one mode of the same standard) in one system. For example, very frequently a single radio includes GSM/GPRS and Bluetooth. In this case, the focus is on reconfigurable radio systems targeting small, low-power-consumption solutions.

Regarding the radio front-end and related to the advances in process technology, some technical challenges include:

- reduction of the supply voltage while dynamic range is kept high [Muhammad05];
- elimination of problems associated with integration-efficient architectures like the direct conversion receiver; such problems include DC offset, $1/f$ noise and second order nonlinearities;
- low-phase-noise local oscillators to accommodate for broadband and multistandard system applications;
- wideband passive and active components (filters and low-noise amplifiers) just after the antenna to accommodate for multistandard and multimode systems as well as for emerging ultrawideband receivers;

For all the above RF front-end-related issues a common target is to minimize energy dissipation.

Regarding the baseband section of the receiver, reconfigurability poses considerable challenges as it requires implementation of multiple computationally intensive functions (like FFT, spreading, despreading and synchronization and decoding) in order to:

- perform hardware/software partition that results in the best possible use of platform resources;
- define the architecture based on the nature of processing; for example, parallel and computationally intensive processing vs algorithmic symbol-level processing [Hawwar06];
- implement the multiple functionalities of the physical layer, which can include several kinds of physical channels (like dedicated channels or synchronization channels), power control and parameter monitoring by measurement (e.g. BER, SNR, signal-to-interference ratio, SIR).

The common aspect of all the above baseband-related problems is to design the digital platform such that partition of the functionalities in DSP, FPGAs and ASICs is implemented in the most efficient way.

1.3 Linear Systems and Random Processes

1.3.1 Linear Systems and Expansion of Signals in Orthogonal Basis Functions

A periodic signal $s(t)$ of bandwidth B_S can be fully reproduced by N samples per period T, spaced $1/(2B_S)$ seconds apart (Nyquist's theorem). Hence, $s(t)$ can be represented by a vector of dimension $N = 2B_S T$. Consequently, most of the properties of vector spaces are true for time waveforms like $s(t)$. Hence, we can define the inner product of two signals $s(t)$ and $y(t)$ in an interval $[c_1, c_2]$ as:

$$\langle s(t), y(t)\rangle = \int_{c_1}^{c_2} s(t)y^*(t)\mathrm{d}t \tag{1.1}$$

Using this, a group of signals $\psi_n(t)$ is defined as orthonormal basis if the following is satisfied:

$$\langle \psi_n(t),\ \psi_m(t)\rangle = \delta_{mn} = \begin{cases} 1, & m = n \\ 0, & m \neq n \end{cases} \tag{1.2}$$

This is used to expand all signals $\{s_m(t),\ m = 1, 2, \ldots M\}$ in terms of the functions $\psi_n(t)$. In this case, $\psi_n(t)$ is defined as a complete basis for $\{s_m(t),\ m = 1, 2, \ldots M\}$ and we have:

$$s_m(t) = \sum_k s_{mk}\psi_k(t), \quad s_{mk} = \int_0^T s_m(t)\psi_k(t)\mathrm{d}t \tag{1.3}$$

This kind of expansion is of great importance in digital communications because the group $\{s_m(t),\ m = 1, 2, \ldots M\}$ represents all possible transmitted waveforms in a transmission system.

Furthermore, if $y_m(t)$ is the output of a linear system with input $s_m(t)$ which performs operation $H[\cdot]$, then we have:

$$y_m(t) = H[s_m(t)] = \sum_k s_{mk}H[\psi_k(t)] \tag{1.4}$$

The above expression provides an easy way to find the response of a system by determining the response of it, when the basis functions $\psi_n(t)$ are used as inputs.

In our case the system is the composite transmitter–receiver system with an overall impulse response $h(t)$ constituting, in most cases, the cascading three filters, the transmitter filter, the channel response and the receiver filter. Hence the received signal will be expressed as the following convolution:

$$y_m(t) = s_m(t) * h(t) \tag{1.5}$$

For example, as shown in Section 2.1, in an ideal system where the transmitted signal is only corrupted by noise we have:

$$r(t) = s_m(t) + n(t), \quad r_k = \int_0^T r(t)\psi_k(t)\mathrm{d}t = s_{mk} + n_k \tag{1.6}$$

Based on the orthonormal expansion

$$s_m(t) = \sum_k s_{mk}\psi_k(t)$$

of signal $s_m(t)$ as presented above, it can be shown [Proakis02] that the power content of a periodic signal can be determined by the summation of the power of its constituent harmonics. This is known as the Parseval relation and is mathematically expressed as follows:

$$\frac{1}{T_0}\int_{t_C}^{t_C+T_0} |s_m(t)|^2 \mathrm{d}t = \sum_{k=-\infty}^{+\infty} |s_{mk}|^2 \tag{1.7}$$

1.3.2 Random Processes

Figure 1.2 shows an example for a random process $X(t)$ consisting of sample functions $x_{Ei}(t)$. Since, as explained above, the random process at a specific time instant t_C corresponds to a random variable, the mean (or expectation function) and autocorrelation function can be defined as follows:

$$E\{X(t_C)\} = m_X(t_C) = \int_{-\infty}^{\infty} x p_{X(t_C)}(x)\mathrm{d}x \tag{1.8}$$

$$R_{XX}(t_1, t_2) = E[X(t_1)X(t_2)] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x_1 x_2 p_{X(t_1)X(t_2)}(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{1.9}$$

Wide-sense stationary (WSS) is a process for which the mean is independent of t and its autocorrelation is a function of the time difference $t_1 - t_2 = \tau$ and not of the specific values of t_1 and t_2 $[R_X(t_1 - t_2) = R_X(\tau)]$.

A random process is stationary if its statistical properties do not depend on time. Stationarity is a stronger property compared with wide-sense stationarity.

Two important properties of the autocorrelation function of stationary processes are:

(1) $R_X(-\tau) = R_X(\tau)$, which means that it is an even function;
(2) $R_X(\tau)$ has a maximum absolute value at $\tau = 0$, i.e. $|R_X(\tau)| \leq R_X(0)$.

Ergodicity is a very useful concept in random signal analysis. A stationary process is ergodic if, for all outcomes E_i and for all functions $f(x)$, the statistical averages are equal to time averages:

$$E\{f[X(t)]\} = \lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} f[x_{Ei}(t)]\mathrm{d}t \tag{1.10}$$

1.3.2.1 Power Spectral Density of Random Processes

It is not possible to define a Fourier transform for random signals. Thus, the concept of power spectral density (PSD) is introduced for random processes. To do that the following steps are taken:

(1) Truncate the sample functions of a random process to be nonzero for $t < T$:

$$x_{Ei}(t;T) = \begin{cases} x_{Ei}(t), & 0 \leq t \leq T \\ 0, & \text{otherwise} \end{cases} \tag{1.11}$$

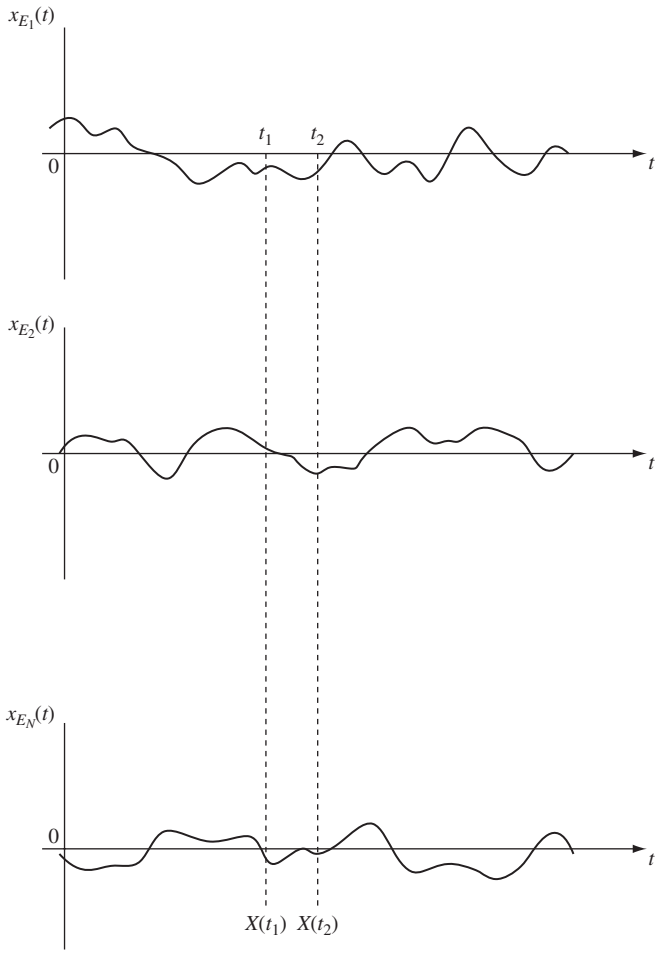

Figure 1.2 Sample functions of random process $X(t)$

(2) Determine $|X_{Ti}(f)|^2$ from the Fourier transform $X_{Ti}(f)$ of the truncated random process $x_{Ei}(t;T)$. The power spectral density $S_{x_{Ei}}(f)$ for $x_{Ei}(t;T)$ is calculated by averaging over a large period of time T:

$$S_{x_{Ei}(f)} = \lim_{T\to\infty} \frac{|X_{Ti}(f)|^2}{T} \tag{1.12}$$

(3) Calculate the average $E\left\{|X_{Ti}(f)|^2\right\}$ over all sample functions $x_{Ei}(t;T)$ [Proakis02]:

$$S_X(f) = \underset{i}{E}\left\{\lim_{T\to\infty} \frac{|X_{Ti}(f)|^2}{T}\right\} = \lim_{T\to\infty} \frac{E\left\{|X_{Ti}(f)|^2\right\}}{T} \tag{1.13}$$

The above procedure converts the power-type signals to energy-type signals by setting them to zero for $t > T$. In this way, power spectral density for random processes defined as above corresponds directly to that of deterministic signals [Proakis02].

In practical terms, $S_X(f)$ represents the average power that would be measured at frequency f in a bandwidth of 1 Hz.

Extending the definitions of energy and power of deterministic signals to random processes, we have for each sample function $x_{Ei}(t)$:

$$E_i = \int x_{Ei}^2(t)\mathrm{d}t, \quad P_i = \lim_{T\to\infty}\frac{1}{T}\int x_{Ei}^2(t)\mathrm{d}t \tag{1.14}$$

Since these quantities are random variables the energy and power of the random process $X(t)$ corresponding to sample functions $x_{Ei}(t)$ are defined as:

$$E_X = E\left\{\int X^2(t)\mathrm{d}t\right\} = \int R_X(t,t)\mathrm{d}t \tag{1.15}$$

$$P_X = E\left[\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} X^2(t)\mathrm{d}t\right] = \frac{1}{T}\int_{-T/2}^{T/2} R_X(t,t)\mathrm{d}t \tag{1.16}$$

For stationary processes, the energy and power are:

$$\begin{aligned} P_X &= R_X(0) \\ E_X &= \int_{-\infty}^{+\infty} R_X(0)\mathrm{d}t \end{aligned} \tag{1.17}$$

1.3.2.2 Random Processes Through Linear Systems

If $Y(t)$ is the output of a linear system with input the stationary random process $X(t)$ and impulse response $h(t)$, the following relations are true for the means and correlation (crosscorrelation and autocorrelation) functions:

$$m_Y = m_X\int_{-\infty}^{+\infty} h(t)\mathrm{d}t \tag{1.18}$$

$$R_{XY}(\tau) = R_X(\tau) * h(-\tau) \tag{1.19}$$

$$R_Y(\tau) = R_X(\tau) * h(\tau) * h(-\tau) \tag{1.20}$$

Furthermore, translation of these expressions in the frequency domain [Equations (1.21)–(1.23)] provides powerful tools to determine spectral densities along the receiver chain in the presence of noise.

$$m_Y = m_X H(0) \tag{1.21}$$

$$S_Y(f) = S_X|H(f)|^2 \tag{1.22}$$

$$S_{YX}(f) = S_X(f)H^*(f) \tag{1.23}$$

1.3.2.3 Wiener–Khinchin Theorem and Applications

The power spectral density of a random process $X(t)$ is given as the following Fourier transform:

$$S_X(f) = \mathrm{F}\left[\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} R_X(t+\tau,t)\mathrm{d}t\right] \tag{1.24}$$

provided that the integral within the brackets takes finite values.

If $X(t)$ is stationary then its PSD is the Fourier transform of the autocorrelation function:

$$S_X(f) = \mathcal{F}[R_X(\tau)] \tag{1.25}$$

An important consequence of the Wiener–Khinchin is that the total power of the random process is equal to the integral of the power spectral density:

$$P_X = E\left[\lim_{T\to\infty}\frac{1}{T}\int_{-T/2}^{T/2} X^2(t)\mathrm{d}t\right] = \int_{-\infty}^{\infty} S_X(f)\mathrm{d}f \tag{1.26}$$

Another useful outcome is that, when the random process is stationary and ergodic, its power spectral density is equal to the PSD of each sample function of the process $x_{Ei}(t)$:

$$S_X(f) = S_{x_{Ei}}(f) \tag{1.27}$$

1.3.3 White Gaussian Noise and Equivalent Noise Bandwidth

White noise is a random process $N(t)$ having a constant power spectral density over all frequencies. Such a process does not exist but it was experimentally shown that thermal noise can approximate $N(t)$ well in reasonably wide bandwidth and has a PSD of value $kT/2$ [Proakis02]. Because the PSD of white noise has an infinite bandwidth, the autocorrelation function is a delta function:

$$R_N(\tau) = \frac{N_0}{2}\delta(\tau) \tag{1.28}$$

where $N_0 = kT$ for white random process. The above formula shows that the random variables associated with white noise are uncorrelated [because $R_N(\tau) = 0$ for $\tau \neq 0$]. If white noise is also a Gaussian process, then the resulting random variables are independent. In practical terms the noise used for the analysis of digital communications is considered white Gaussian, stationary and ergodic process with zero mean. Usually, this noise is additive and is called additive white Gaussian noise (AWGN).

If the above white noise passes through an ideal bandpass filter of bandwidth B, the resulting random process is a bandpass white noise process. Its spectral density and autocorrelation function are expressed as follows:

$$S_{BN}(f) = \begin{cases} N_0, & |f| \le B/2 \\ 0, & |f| \ge B/2 \end{cases}, \quad R_{BN}(\tau) = N_0\frac{\sin(\pi B\tau)}{\pi\tau} \tag{1.29}$$

Noise equivalent bandwidth of a specific system refers to the bandwidth of an ideal reference filter that will produce the same noise power at its output with the given system. More specifically, let the same white noise with PSD equal to $N_0/2$ pass through a filter F_r with a given frequency response $|H_F(f)|$ and a fictitious ideal (rectangular) filter, as shown in Figure 1.3. We can define the constant magnitude of the ideal filter equal to the magnitude $|H(f)|_{\mathrm{ref}}$ of

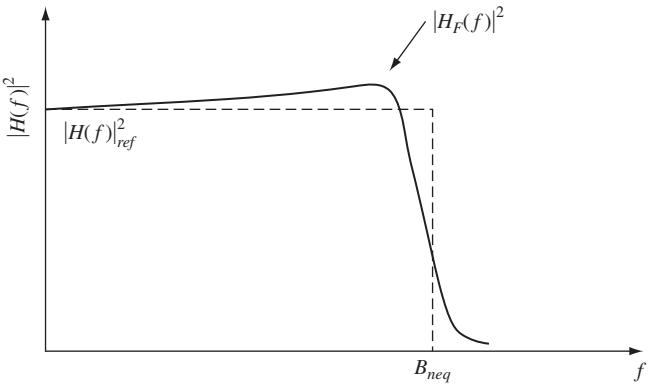

Figure 1.3 Equivalence between frequency response $|H_F(f)|^2$ and ideal brick-wall filter with bandwidth B_{neq}

F_r at a reference frequency f_{ref}, which in most cases represents the frequency of maximum magnitude or the 3 dB frequency.

In this case, noise equivalent bandwidth is the bandwidth of the ideal brick-wall filter, which will give the same noise power at its output as filter F_r. The output noise power of the given filter and the rectangular filter is:

$$P_N = N_0 \int_0^\infty |H(f)|^2 \mathrm{d}f, \quad P_{Nr} = N_0 |H(f)|^2_{f_{\text{ref}}} B_{\text{neq}} \tag{1.30}$$

To express B_{neq} we equate the two noise powers. Thus, we get:

$$B_{\text{neq}} = \frac{\int_0^\infty |H(f)|^2 \mathrm{d}f}{|H(f)|^2_{f_{\text{ref}}}} \tag{1.31}$$

1.3.4 Deterministic and Random Signals of Bandpass Nature

In communications, a high-frequency carrier is used to translate the information signal into higher frequency suitable for transmission. For purposes of analysis and evaluation of the performance of receivers, it is important to formulate such signals and investigate their properties.

A bandpass signal is defined as one for which the frequency spectrum $X(f)$ is nonzero within a bandwidth W around a high frequency carrier f_C:

$$X(f) = \begin{cases} \text{nonzero}, & |f - f_C| \le W \\ \text{zero}, & |f - f_C| \ge W \end{cases} \tag{1.32}$$

It is customary to express high frequency modulated signals as:

$$x(t) = A(t) \cos [2\pi f_C t + \theta(t)] = \text{Re}[A(t) \exp (j2\pi f_C t) \exp (j\theta(t))] \tag{1.33}$$

$A(t)$ and $\theta(t)$ correspond to amplitude and phase which can contain information. Since the carrier f_C does not contain information, we seek expressions for the transmitted signal in

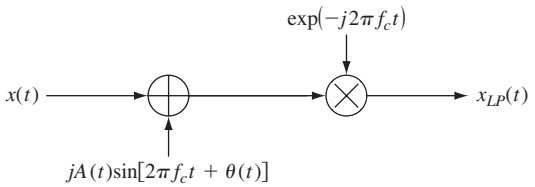

Figure 1.4 The lowpass complex envelope signal $x_{\mathrm{LP}}(t)$ produced from $x(t)$

which dependence on f_C is eliminated. Inevitably, this signal will be of lowpass (or baseband) nature and can be expressed in two ways:

$$x_{LP}(t) = A(t)\exp\left(j\theta(t)\right), \quad x_{LP}(t) = x_I(t) + jx_Q(t) \tag{1.34}$$

To obtain the lowpass complex envelope signal $x_{\mathrm{LP}}(t)$ from $x(t)$, the term $jA(t)\sin\left[2\pi f_C t + \theta(t)\right]$ must be added to the passband signal $x(t)$ and the carrier must be removed by multiplying by $\exp\left(-2\pi f_C t\right)$. This is depicted in Figure 1.4.

Consequently, $x_{\mathrm{LP}}(t)$ can be expressed as [Proakis02]:

$$x_{\mathrm{LP}}(t) = \left[x(t) + \breve{x}(t)\right]\exp\left(-j2\pi f_C t\right) \tag{1.35}$$

where $\breve{x}(t)$ is defined as the Hilbert transform of $x(t)$ and is analytically expressed in time and frequency domains as:

$$\breve{x}(t) = \frac{1}{\pi t} * x(t) \tag{1.36}$$

From the above we realize that the Hilbert transform is a simple filter which shifts by $-\pi/2$ the phase of the positive frequencies and by $+\pi/2$ the phase of the negative frequencies. It is straightforward to show that the relation of the bandpass signal $x(t)$ and quadrature lowpass components $x_I(t), x_Q(t)$ is:

$$x(t) = x_I(t)\cos\left(2\pi f_C t\right) - x_Q(t)\sin\left(2\pi f_C t\right) \tag{1.37a}$$

$$\breve{x}(t) = x_I(t)\sin\left(2\pi f_C t\right) + x_Q(t)\cos\left(2\pi f_C t\right) \tag{1.37b}$$

The envelope and phase of the passband signal are:

$$A(t) = \sqrt{x_I^2(t) + x_Q^2(t)}, \quad \theta(t) = \tan^{-1}\left[\frac{x_Q(t)}{x_I(t)}\right] \tag{1.38}$$

Considering random processes, we can define that a random process $X_N(t)$ is bandpass if its power spectral density is confined around the centre frequency f_C:

$$X_N(t) \text{ is a bandpass process if : } S_{X_N}(f) = 0 \text{ for } |f - f_C| \geq W, \quad W < f_C \tag{1.39}$$

It is easy to show that the random process along with its sample functions $x_{Ei}(t)$ can be expressed in a similar way as for deterministic signals in terms of two new processes $X_{nI}(t)$ and $X_{nQ}(t)$, which constitute the in-phase and quadrature components:

$$X_N(t) = A(t)\cos\left[2\pi f_C t + \theta(t)\right] = X_{nI}(t)\cos\left(2\pi f_C t\right) - X_{nQ}(t)\sin\left(2\pi f_C t\right) \tag{1.40}$$

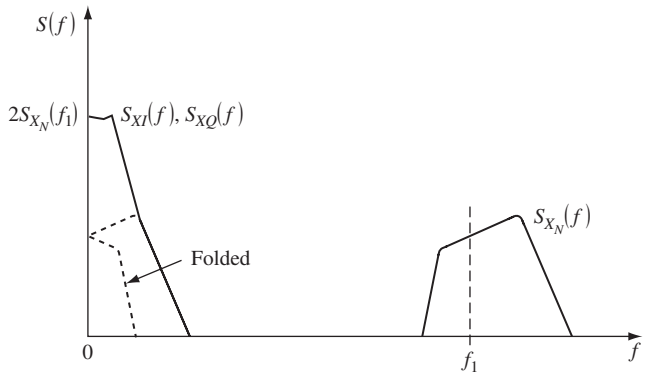

Figure 1.5 Lowpass nature of the PSDs $S_{XI}(f)$, $S_{XQ}(f)$ of quadrature components $x_I(t)$, $x_Q(t)$

If $X_N(t)$ is a stationary bandpass process of zero mean, processes $X_{nI}(t)$ and $X_{nQ}(t)$ are also zero mean [Proakis02].

Considering autocorrelation functions $R_{nI}(\tau)$, $R_{nQ}(\tau)$ of $X_{nI}(t)$ and $X_{nQ}(t)$, it can be shown that:

$$R_{nI}(\tau) = R_{nQ}(\tau) \tag{1.41}$$

The spectra $S_{XI}(f)$ and $S_{XQ}(f)$ of processes $X_{nI}(t)$ and $X_{nQ}(t)$ become zero for $|f| \geq W$ and consequently they are lowpass processes. Furthermore, their power spectral densities can be calculated and are given as [Proakis02]:

$$S_{XI}(f) = S_{XQ}(f) = \frac{1}{2}[S_{X_N}(f - f_C) + S_{X_N}(f + f_C)] \tag{1.42}$$

Figure 1.5 gives the resulting lowpass spectrum of $X_{nI}(t)$ and $X_{nQ}(t)$. Similarly, as for deterministic signals, the envelope and phase processes $A(t)$ and $\theta(t)$ are defined as:

$$X_{LP}(t) = A(t)\exp[\,j\theta(\mathrm{t})], \quad A(t) = \sqrt{X_{nI}^2(t) + X_{nQ}^2(t)}, \quad \theta(\mathrm{t}) = \tan^{-1}\left[\frac{X_{nQ}(t)}{X_{nI}(t)}\right] \tag{1.43}$$

where $X_{LP}(t)$ is the equivalent lowpass process for $X_N(t)$, which now can be expressed as:

$$X_N(t) = A(t)\cos[2\pi f_C t + \theta(t)] \tag{1.44}$$

The amplitude p.d.f. follows the Rayleigh distribution with mean $\overline{A}$ and variance $\overline{A^2}$ [Gardner05]:

$$E[A] = \overline{A} = \sigma_n\sqrt{\pi/2}, \quad E[A^2] = \overline{A^2} = 2\sigma_n^2 \tag{1.45}$$

Regarding the phase, if we assume that it takes values in the interval $[-\pi, \pi]$, $\theta(t)$ follows a uniform distribution with p.d.f. $p(\theta) = 1/(2\pi)$ within the specified interval. Furthermore, its mean value is equal to zero and its variance is $\overline{\theta^2} = \pi^2/3$.

1.4 Radio Channel Characterization

Transmission of high frequency signals through the radio channel experiences distortion and losses due to reflection, absorption, diffraction and scattering. One or more of these mechanisms is activated depending on the transceiver position. Specifically, in outdoor environments important factors are the transmitter–receiver (Tx–Rx) distance, mobility of the transmitter or the receiver, the formation of the landscape, the density and the size of the buildings. For indoor environments, apart from the Tx–Rx distance and mobility, important factors are the floor plan, the type of partitions between different rooms and the size and type of objects filling the space.

A three-stage model is frequently used in the literature to describe the impact of the radio channel [Pra98], [Rappaport02], [Proakis02], [Goldsmith05]:

- large-scale path loss;
- medium-scale shadowing;
- small-scale multipath fading.

Large-scale attenuation (or path loss) is associated with loss of the received power due to the distance between the transmitter and the receiver and is mainly affected by absorption, reflection, refraction and diffraction.

Shadowing or shadow fading is mainly due to the presence of obstacles blocking the line-of-sight (LOS) between the transmitter and the receiver. The main mechanisms involved in shadowing are reflection and scattering of the radio signal.

Small-scale multipath fading is associated with multiple reflected copies of the transmitted signal due to scattering from various objects arriving at the receiver at different time instants. In this case, the vector summation of all these copies with different amplitude and phase results in fading, which can be as deep as a few tens of decibels. Successive fades can have distances smaller than $\lambda/2$ in a diagram presenting received signal power vs. distance. In addition, the difference in time between the first and the last arriving copy of the received signal is the time spread of the delay of the time of arrival at the receiver. This is called *delay spread of the channel* for the particular Tx–Rx setting. Figure 1.6 depicts the above three attenuation and fading mechanisms.

1.4.1 Large-scale Path Loss

The ratio between the transmitted power P_{T} and the locally-averaged receiver signal power P_{Rav} is defined as the path loss of the channel:

$$P_{\mathrm{L}} = \frac{P_{\mathrm{T}}}{P_{\mathrm{Rav}}} \tag{1.46}$$

The receiver signal power is averaged within a small area (with a radius of approximately 10 wavelengths) around the receiver in order to eliminate random power variations due to shadow fading and multipath fading.

The free-space path loss for a distance d between transmitter and receiver, operating at a frequency $f = c/\lambda$, is given by [Proakis02]:

$$L_S = \left(\frac{4\pi d}{\lambda}\right)^2 \tag{1.47}$$

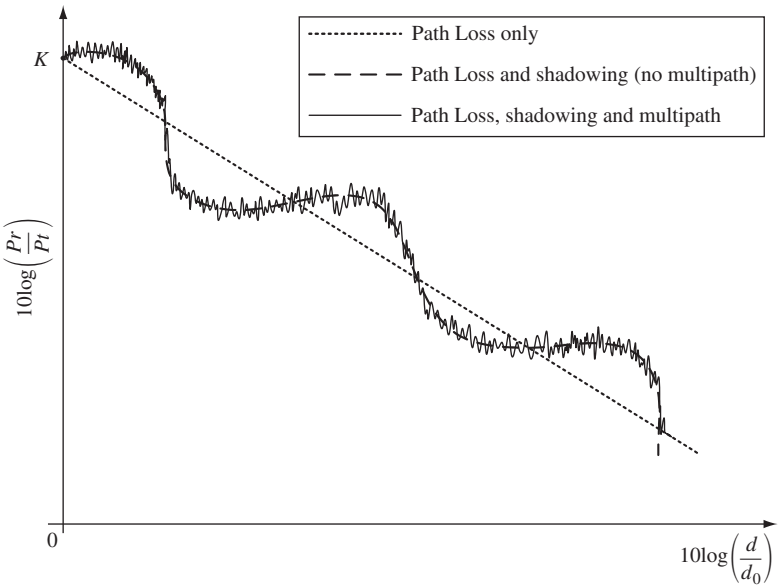

Figure 1.6 The three mechanisms contributing to propagation losses (reprinted from A. Goldsmith, 'Wireless Communications', copyright © 2005 by Cambridge Academic Press)

whereas the power at the input of the receiver for antenna gains of the transmitter and the receiver $G_{\mathrm{T}}, G_{\mathrm{R}}$, respectively, is:

$$P_{\mathrm{Rav}} = \frac{P_{\mathrm{T}} G_{\mathrm{T}} G_{\mathrm{R}}}{(4\pi d/\lambda)^2} \tag{1.48}$$

With these in mind, the free-space path loss is given as:

$$P_{\mathrm{L}}(\mathrm{dB}) = -10 \log_{10} \left[\frac{G_{\mathrm{T}} G_{\mathrm{R}} \lambda^2}{(4\pi d)^2} \right] \tag{1.49}$$

However, in most radio systems the environment within which communication between the transmitter and the receiver takes place is filled with obstacles which give rise to phenomena like reflection and refraction, as mentioned above. Consequently, the free-space path loss formula cannot be used to accurately estimate the path losses. For this reason, empirical path loss models can be used to calculate path loss in macrocellular, microcellular and picocellular environments. The most important of these models are the Okumura model and the Hata model [Rappaport02], [Goldsmith05], which are based on attenuation measurements recorded in specific environments as a function of distance.

The Okumura model refers to large urban macrocells and can be used for distances of 1–100 km and for frequency ranges of 150–1500 MHz. The Okumura path-loss formula is associated with the free-space path loss and also depends on a mean attenuation factor $A_M(f_C d)$ and gain factors $G_{\mathrm{T}}(h_{\mathrm{T}})$, $G_{\mathrm{R}}(h_{\mathrm{R}})$ and G_{ENV} related to base station antenna, mobile antenna and type of environment respectively [Okumura68], [Rappaport02], [Goldsmith05]:

$$P_{\mathrm{L}}(d) = L_F(f_C, d) + A_M(f_C, d) - G_{\mathrm{T}}(h_{\mathrm{T}}) - G_{\mathrm{R}}(h_{\mathrm{R}}) - G_{\mathrm{ENV}} \tag{1.50}$$

$$G_{\mathrm{T}}(h_{\mathrm{T}}) = 20\log_{10}(h_{\mathrm{T}}/200),\ \text{for } 30\,\mathrm{m} < h_{\mathrm{T}} < 1000\,\mathrm{m} \tag{1.51}$$

$$G_{\mathrm{R}}(h_{\mathrm{R}}) = \begin{cases} 10\log_{10}(h_{\mathrm{R}}/3) & h_{\mathrm{R}} \le 3\,\mathrm{m} \\ 20\log_{10}(h_{\mathrm{R}}/3) & 3\,\mathrm{m} < h_{\mathrm{R}} < 10\,\mathrm{m} \end{cases} \tag{1.52}$$

The Hata model [Hata80] is a closed form expression for path loss based on the Okumura data and is valid for the same frequency range (150–1500 MHz):

$$P_{\mathrm{L}u}(d) = 69.55 + 26.16\log_{10}(f_C) - 13.82\log_{10}(h_{\mathrm{T}}) - C(h_{\mathrm{R}}) + [44.9 - 6.55\log_{10}(h_{\mathrm{T}})]\log_{10}(d)\ \mathrm{dB} \tag{1.53}$$

h_{T} and h_{R} represent the base station and mobile antenna heights as previously whereas $C(h_{\mathrm{R}})$ is a correction factor associated with the antenna height of the mobile and depends on the cell radius. For example, for small or medium size cities it is [Goldsmith05]:

$$C(h_{\mathrm{R}}) = [1.1\log_{10} f_C - 0.7]h_{\mathrm{R}} - [1.56\log_{10}(f_C) - 0.8]\ \mathrm{dB} \tag{1.54}$$

There is a relation associating the suburban and rural models to the urban one. For example, the suburban path-loss model is:

$$P_{\mathrm{L,sub}}(d) = P_{\mathrm{L},u}(d) - 2(\log_{10}(f_C/28))^2 - 5.4 \tag{1.55}$$

COST 231 [Euro-COST 231-1991] is an extension of the Hata model for specific ranges of antenna heights and for frequencies between 1.5 and 2.0 GHz.

An empirical model for path loss in a microcellular environment (outdoor and indoor) is the so-called 'piecewise linear' model [Goldsmith05]. It can consist of N linear sections (segments) of different slopes on a path loss (in decibels) vs. the logarithm of normalized distance $[\log_{10}(d/d_0)]$ diagram.

The most frequently used is the dual-slope model, giving the following expression for the received power [Goldsmith05]:

$$P_{\mathrm{R}}(d) = \begin{cases} P_T + K - 10\gamma_1\log_{10}(d/d_0)\ \mathrm{dB} & d_0 \le d \le d_{\mathrm{B}} \\ P_T + K - 10\gamma_1\log_{10}(d_B/d_0) - 10\gamma_2\log_{10}(d/d_C)\ \mathrm{dB} & d > d_{\mathrm{B}} \end{cases} \tag{1.56}$$

where K is an attenuation factor depending on channel attenuation and antenna patterns. K is usually less than 1, corresponding to negative values in decibels; d_0 is a reference distance marking the beginning of the antenna far field; d_{B} is a breakpoint beyond which the diagram of P_{R} vs the logarithm of normalized distance changes slope; and γ_1 and γ_2 represent the two different slopes (for distances up to d_{B} and beyond d_{B}) of the P_{R} vs $\log_{10}(d/d_0)$ diagram.

For system design and estimation of the coverage area, it is frequently very useful to employ a simplified model using a single path loss exponent γ covering the whole range of transmitter–receiver distances. Hence the corresponding formula is:

$$P_{\mathrm{R}}(d) = P_{\mathrm{T}} + K - 10\log_{10}(d/d_0)\,\mathrm{dBm}, \quad d > d_0 \tag{1.57}$$

The above model is valid for both indoor environments ($d_0 = 1$–10 m) and outdoor environments ($d_0 = 10$–100 m). In general, the path loss exponent is between 1.6 and 6 in most applications depending on the environment, the type of obstructions and nature of the walls in indoor communication. For example, in an indoor environment $1.6 \le \gamma \le 3.5$, when transmitter and receiver are located on the same floor [Rappaport02].

Finally, a more detailed model for indoor propagation can be produced by taking into account specific attenuation factors for each obstacle that the signal finds in its way from the transmitter to the receiver. Hence, the above formulas can be augmented as follows:

$$P_{\mathrm{R}}(d) = P_{\mathrm{T}} - P_{\mathrm{L}}(d) - \sum_{i=1}^{N} AF_i \text{ dB} \tag{1.58}$$

where $P_{\mathrm{L}}(d)$ is the losses using a path loss model and AF_i is the attenuation factor of the ith obstacle. For example, if the obstacle is a concrete wall, AF is equal to 13 dB.

1.4.2 Shadow Fading

As mentioned in the beginning of this section, shadow fading is mainly due to the presence of objects between the transmitter and the receiver. The nature, size and location of the objects are factors that determine the amount of attenuation due to shadowing. Hence, the randomness due to shadow fading stems from the size and location of the objects and not from the distance between the transmitter and the receiver. The ratio of the transmitted to the received power $\psi = P_T/P_R$ is a random variable with log–normal distribution [Goldsmith05]:

$$p(\psi) = \frac{10/\ln 10}{\psi\sqrt{2\pi}\sigma_{\psi\mathrm{dB}}} \exp\left[-\frac{(10\log_{10}\psi - m_{\psi\mathrm{dB}})^2}{2\sigma_{\psi\mathrm{dB}}^2}\right] \tag{1.59}$$

where $m_{\psi\mathrm{dB}}$ and $\sigma_{\psi\mathrm{dB}}$ are the mean and variance (both in decibels) of the random variable $\Psi = 10\log_{10}\psi$. The mean $m_{\psi\mathrm{dB}}$ represents the empirical or analytical path loss, as calculated in the 'large-scale path-loss' subsection above.

1.4.3 Multipath Fading in Wideband Radio Channels

1.4.3.1 Input–Output Models for Multipath Channels

The objective of this section is to obtain simple expressions for the impulse response of a radio channel dominated by multipath. For this purpose, we assume a transmitting antenna, a receiving antenna mounted on a vehicle and four solid obstructions (reflectors) causing reflected versions of the transmitted signal to be received by the vehicle antenna, as illustrated in Figure 1.7. We examine two cases: one with static vehicle and one with moving vehicle.

Taking into account that we have discrete reflectors (scatterers), we represent by a_n, x_n, τ_n and φ_n, the attenuation factor, the length of the path, the corresponding delay and the phase change due to the nth arriving version of the transmitted signal (also called the nth path), respectively.

Let the transmitted signal be:

$$s(t) = \mathrm{Re}\left\{s_{\mathrm{L}}(t)\mathrm{e}^{j2\pi f_C t}\right\} = \mathrm{Re}\left\{|s_{\mathrm{L}}(t)|\mathrm{e}^{j(2\pi f_C t + \varphi_{s_{\mathrm{L}}}(t))}\right\} = |s_{\mathrm{L}}(t)|\cos\left(2\pi f_C t + \varphi_{s_{\mathrm{L}}}(t)\right) \tag{1.60}$$

where $s_L(t)$ represents the baseband equivalent received signal.

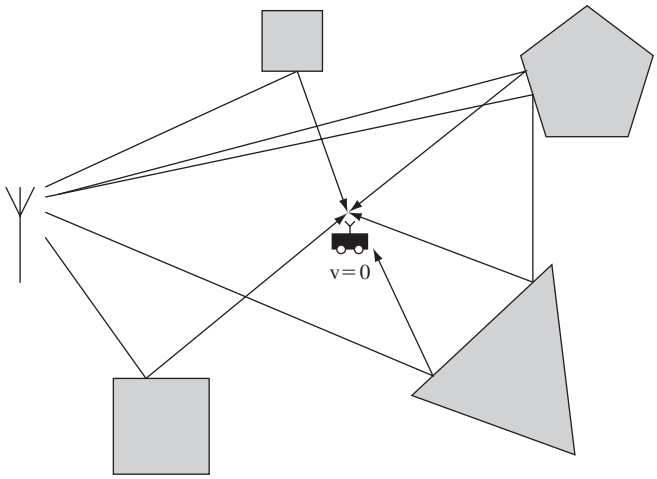

Figure 1.7 Signal components produced by reflections on scatterers arrive at the mobile antenna

Static Transmitter and Reflectors (Scatterers), Static Vehicle

The received bandpass and baseband signals are given as:

$$r(t) = \mathrm{Re}\left\{r_{\mathrm{L}}(t)\mathrm{e}^{j2\pi f_C t}\right\} = \mathrm{Re}\left\{\left[\sum_n a_n \mathrm{e}^{-j2\pi f_C \tau_n} s_{\mathrm{L}}(t-\tau_n)\right]\mathrm{e}^{j2\pi f_C t}\right\} \tag{1.61}$$

with

$$a_n = \overline{a_n}\mathrm{e}^{j\varphi_n} \tag{1.62}$$

The attenuation coefficient a_n is complex so as to include the magnitude of the attenuation factor $\overline{a_n}$ and the effect of the change of phase φ_n due to reflections. τ_n is related to x_n by $\tau_n = x_n/c = x_n/(\lambda f_C)$. Consequently, the lowpass channel impulse response is given by:

$$c(\tau;t) = \sum_n a_n \mathrm{e}^{-j2\pi f_C \tau_n}\delta(\tau - t_n) \tag{1.63}$$

Static Transmitter and Reflectors, Moving Vehicle

In this case we have to briefly present the impact of the Doppler effect. For a vehicle moving in the horizontal direction, Doppler is associated with the small difference Δx in the distance that the transmitter signal must cover in the two different positions of the vehicle. As shown in Figure 1.8, when the vehicle is located at point X, it receives the signal from point S in an angle θ.

After time t, the vehicle has moved to point Y where we assume that the angle of arrival is still θ (a valid assumption if the transmitter S is far away from the vehicle). Hence, the distance difference Δx is:

$$\Delta x = -vt\cos\theta \tag{1.64}$$

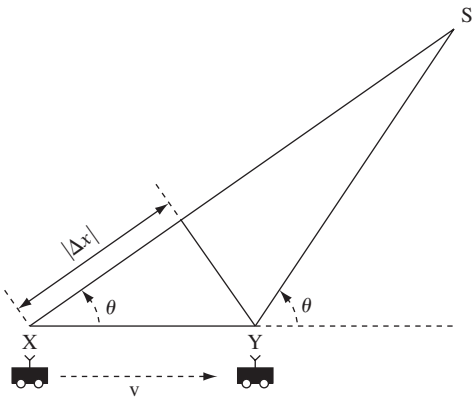

Figure 1.8 Generation of Doppler effect due to a moving vehicle

This change results in a phase change and consequently in a change of instantaneous frequency which is:

$$v_d = \frac{f_C}{c} v \cos\theta \tag{1.65}$$

whereas the maximum frequency change (for $\theta = 0$) is:

$$v_D = \frac{f_C}{c} v \tag{1.66}$$

After some elementary calculations one can show that the received equivalent baseband signal is the summmation of n different paths:

$$\begin{aligned} r_{\mathrm{L}}(t) &= \sum_n a_n \mathrm{e}^{-j2\pi x_n/\lambda} \mathrm{e}^{j2\pi \frac{v}{\lambda} \cos\theta_n t} s_L\left(t - \frac{x_n}{c} + \frac{v\cos\theta_n t}{c}\right) \\ &= \sum_n a_n \mathrm{e}^{-j2\pi x_n/\lambda} \mathrm{e}^{j2\pi f_{Dn} t} s_L\left(t - \frac{x_n}{c} + \frac{v\cos\theta_n t}{c}\right) \end{aligned} \tag{1.67}$$

Disregarding $v\cos\theta_n t/c$ because it is very small, Equation (1.67) gives:

$$r_{\mathrm{L}}(t) = \sum_n a_n \mathrm{e}^{j2\pi v_D \cos\theta_n t} s_{\mathrm{L}}(t - \tau_n) \tag{1.68}$$

Consequently:

$$c(\tau; t) = \sum_n a_n \mathrm{e}^{j2\pi v_D \cos\theta_n t} \delta(t - \tau_n) \tag{1.69}$$

The next step is to assume that there is a 'continuous' set of scatterers instead of discrete scatterers located in the surrounding area. In that case, summations are replaced with integrals and it can be shown that the received signals (passband and baseband) and impulse responses are [Proakis02], [Goldsmith05]:

$$r(t) = \int_{-\infty}^{+\infty} \alpha(\tau; t) s(t - \tau) \mathrm{d}\tau \tag{1.70}$$

$$r_{\mathrm{L}}(t) = \int_{-\infty}^{+\infty} \alpha(\tau; t)\mathrm{e}^{-j2\pi f_C \tau} s_{\mathrm{L}}(t-\tau)\mathrm{d}\tau \tag{1.71}$$

where $\alpha(\tau; t)$ represents the attenuation at delay equal to τ at time instant t.

The lowpass channel impulse response in this case is:

$$c(\tau; t) = \alpha(\tau; t)\mathrm{e}^{-j2\pi f_C \tau} \tag{1.72}$$

The input–output relations between $r_{\mathrm{L}}(t)$ and $s_{\mathrm{L}}(t)$ are given [Fleury00], [Goldsmith05]:

$$r_L(t) = \int_{-\infty}^{+\infty} c(\tau; t)s_{\mathrm{L}}(t-\tau)\mathrm{d}\tau \tag{1.73}$$

$$r_L(t) = \int_{-\infty}^{+\infty} C(f; t)S_{\mathrm{L}}(f)\mathrm{e}^{-j2\pi f t}\mathrm{d}f \tag{1.74}$$

where $C(f; t)$ represents the Fourier transform of $c(\tau; t)$ with respect to variable τ. It is called the time-variant transfer function and is given by:

$$C(f; t) = \int_{-\infty}^{+\infty} c(\tau; t)\mathrm{e}^{-j2\pi f \tau}\mathrm{d}\tau \tag{1.75}$$

which, for discrete impulse response, becomes:

$$C(f; t) = \sum_n \alpha_n(t)\mathrm{e}^{-j2\pi f_C \tau_n(t)}\mathrm{e}^{-j2\pi f \tau_n(t)} \tag{1.76}$$

In addition, $S_L(f)$ represents the power spectrum of $s_L(t)$.

Another expression for $r_{\mathrm{L}}(t)$ is [Fleury00]:

$$r_L(t) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{e}^{j2\pi\nu t} s_{\mathrm{L}}(t-\tau)h(\tau, \nu)\mathrm{d}\tau\mathrm{d}\nu \tag{1.77}$$

where $h(\tau, \nu)$ is called the delay-Doppler spread function expressed as:

$$h(\tau, \nu) = \sum_n \alpha_n \delta(\nu - \nu_n)\delta(\tau - \tau_n) \tag{1.78}$$

and consequently $h(\tau, \nu)$ is the Fourier transform of $c(\tau; t)$ with respect to variable t:

$$h(\tau, \nu) = \int_{-\infty}^{+\infty} c(\tau; t)\mathrm{e}^{-j2\pi\nu t}\mathrm{d}t \tag{1.79}$$

Furthermore, $h(\tau, \nu)$ is the two-dimensional Fourier transform of $C(f; t)$:

$$h(\tau, \nu) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} C(f; t)\mathrm{e}^{j2\pi f \tau}\mathrm{e}^{-j2\pi\nu t}\mathrm{d}t\mathrm{d}f \tag{1.80}$$

1.4.3.2 Spectral Densities and Autocorrelation Functions

It is now necessary to produce quantities which can be used to determine the distribution of power with respect to time delay τ and Doppler frequency ν. These quantities are associated

with autocorrelation functions of the impulse response and the frequency response of the radio channel.

With respect to the distribution of power the cross-power delay spectrum $\varphi_c(\tau;\Delta t)$ is needed, which is given by (for wide-sense stationary uncorrelated scattering, WSSUS):

$$\varphi_c(\tau_1;\Delta t)\delta(\tau_1-\tau_2)=R_c(\tau_1,\tau_2;t_1,t_2)\equiv\frac{1}{2}E\left[c^*(\tau_1,t_1)c(\tau_2,t_2)\right] \tag{1.81}$$

where $\Delta t=t_2-t_1$.

$\varphi_c(\tau_1;\Delta t)$ gives the average power of the output of the channel as a function of τ and Δt. For $\Delta t=0$ the resulting autocorrelation $\varphi_c(\tau;0)\equiv\varphi_c(\tau)$ is called the power delay profile (PDP) and illustrates how the power at the radio channel output is distributed in the delay τ domain.

Furthermore, we define the frequency–time correlation function $\varphi_C(\Delta f;\Delta t)$:

$$\begin{aligned}\varphi_C(\Delta f;\Delta t)&=\frac{1}{2}E\left[C^*(f;t)C(f+\Delta f;t+\Delta t)\right]\\&\equiv\int_{-\infty}^{\infty}\varphi_c(\tau_1;\Delta t)\mathrm{e}^{j2\pi\tau_1(f_2-f_1)}\mathrm{d}\tau_1=R_C(f_1,f_2;t_1,t_2)\end{aligned} \tag{1.82}$$

At this point, it is important to introduce the delay-Doppler power spectrum, or scattering function $S(\tau;\nu)$ which can be shown to be [Fleury00]:

$$S(\tau;\nu)=\int_{-\infty}^{\infty}\varphi_c(\tau;\Delta t)\mathrm{e}^{-j2\pi\nu\Delta t}\mathrm{d}(\Delta t)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\varphi_C(\Delta f;\Delta t)\mathrm{e}^{-j2\pi\nu\Delta t}\mathrm{e}^{j2\pi\tau\Delta f}\mathrm{d}(\Delta t)\mathrm{d}(\Delta f) \tag{1.83}$$

Also, its relation to the autocorrelation of $h(\tau,\nu)$ can be shown to be [Fleury00]:

$$R_h(\tau_1,\tau_2;\nu_1,\nu_2)=S(\tau_1;\nu_1)\delta(\nu_2-\nu_1)\delta(\tau_2-\tau_1) \tag{1.84}$$

The importance of the scattering function lies in the fact that it reveals the way the average power at the receiver is distributed in two domains, the delay domain and the Doppler frequency domain.

In addition the double integral of $S(\tau;\nu)$ can be expressed as:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}S(\tau;\nu)\mathrm{d}\tau\mathrm{d}\nu=\varphi_C(0;0)=\frac{1}{2}E\left[|C(f;t)|^2\right] \tag{1.85}$$

which implies that the bandpass output power is the same regardless of the fact that the input to the channel may be a narrowband or a wideband signal.

Finally, the Doppler cross-power spectrum $S_C(\Delta f;\Delta\nu)$ is the Fourier transform of $\varphi_C(\Delta f;\Delta t)$:

$$S_C(\Delta f;\nu)=\int_{-\infty}^{+\infty}\varphi_C(\Delta f;\Delta t)\mathrm{e}^{-j2\pi\nu\Delta t}\mathrm{d}\tau=\int_{-\infty}^{+\infty}S(\tau;\nu)\mathrm{e}^{-j2\pi\tau\Delta f}\mathrm{d}\tau \tag{1.86}$$

Letting $\Delta f=0$, we define $S_C(0;\Delta\nu)\equiv S_C(\nu)$ as the Doppler power spectrum which takes the form:

$$S_C(\nu)=\int_{-\infty}^{+\infty}\varphi_C(\Delta t)\mathrm{e}^{-j2\pi\nu\Delta t}\mathrm{d}\tau \tag{1.87}$$

It is important to note that $S_C(\nu)$ depicts the power distribution in Doppler frequency.

In addition we have:

$$\varphi_c(\tau) = \int_{-\infty}^{+\infty} S(\tau;\nu)\mathrm{d}\nu \tag{1.88}$$

$$S_C(\nu) = \int_{-\infty}^{+\infty} S(\tau;\nu)d\tau \tag{1.89}$$

In conclusion, all three functions $S(\tau;\nu)$, $S_C(\nu)$ and $\varphi_c(\tau)$ represent power spectral densities and will produce narrowband or wideband power after integration. Furthermore, function $\varphi_C(\Delta f;\Delta t)$ characterizes channel selectivity in frequency and time. By eliminating one of the variables, we obtain channel selectivity with respect to the other. Hence, $\varphi_C(\Delta f)$ gives the channel selectivity with respect to frequency and $\varphi_C(\Delta t)$ with respect to time.

For discrete scatterer positions, the power delay profile is given by:

$$\varphi_c(\tau) = \sum_n \frac{1}{2}E\left[|a_n|^2\right]\delta(\tau - \tau_n) \tag{1.90}$$

Furthermore, the Doppler power spectrum and the corresponding autocorrelation function are given by:

$$S_C(\nu) = \frac{\sigma^2}{\pi\nu_D}\frac{1}{\sqrt{1-\left(\frac{\nu}{\nu_D}\right)^2}}, \quad \text{for } \nu \in (-\nu_D, \nu_D) \tag{1.91}$$

$$\varphi_C(\Delta t) = \sigma^2 J_0(2\pi\nu_D\Delta t) \tag{1.92}$$

Figure 1.9 shows the Doppler power spectrum for omnidirectional receiving antenna and uniformly distributed scatterers.

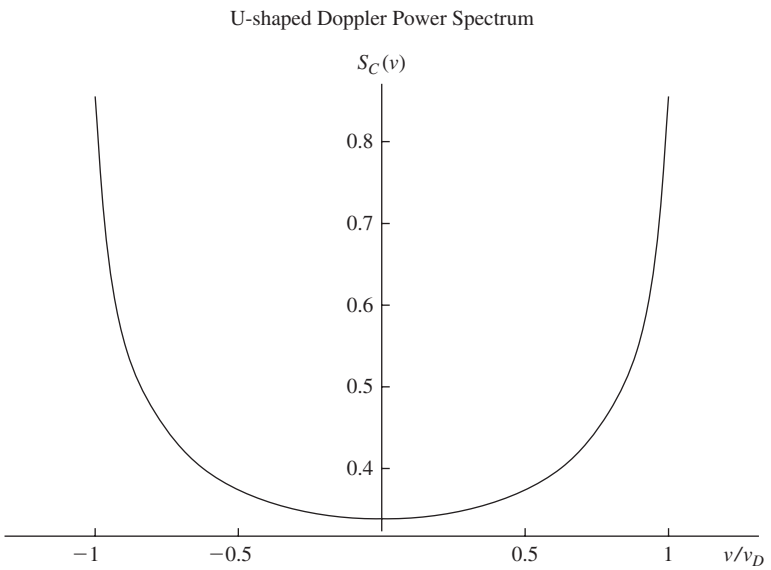

Figure 1.9 U-Shaped Doppler power spectrum for uniform distributed scatterers

1.4.3.3 Parameters of Multipath Radio Channels

Dispersion in Time Delay and Frequency Selectivity

By choosing the peak values in a continuous power delay profile, we obtain the corresponding discrete PDP. Furthermore, it is convenient to set the time of arrival of the first ray equal to zero ($\tau_1 = 0$).

Figure 1.10 illustrates the above notions. $\varphi_c(\tau)$ is used to depict time dispersion because it gives the way power is distributed as a function of time delay. For equivalence between the continuous and the corresponding discrete PDP, the following must hold:

$$\sigma_{a_n}^2 = \int_{\tau_n}^{\tau_{n+1}} \varphi_c(\tau)\mathrm{d}\tau, \quad \text{for } n = 1, \ldots, N-1$$
$$\sigma_{a_N}^2 = \int_{\tau_N}^{+\infty} \varphi_c(\tau)\mathrm{d}\tau, \quad \text{for } n = N \tag{1.93}$$

The mean delay for continuous and discrete power delay profiles, is given respectively by:

$$m_\tau = \frac{\int_0^\infty \tau\varphi_c(\tau)\mathrm{d}\tau}{\int_0^\infty \varphi_c(\tau)\mathrm{d}\tau} \tag{1.94}$$

$$m_\tau = \frac{\sum_n \tau_n \int_{\tau_n}^{\tau_{n+1}} \varphi_c(\tau)\mathrm{d}\tau}{\sum_n \int_{\tau_n}^{\tau_{n+1}} \varphi_c(\tau)\mathrm{d}\tau} = \frac{\sum_n \sigma_{a_n}^2 \tau_n}{\sum_n \sigma_{a_n}^2} \tag{1.95}$$

where the denominator is necessary for PDP normalization. If the PDP power is normalized to one, the denominators are equal to one and can be eliminated.

The rms delay spread for both continuous and discrete profiles is given by:

$$\sigma_\tau \equiv \sqrt{\frac{\int_0^{+\infty} (\tau - m_\tau)^2 \varphi_c(\tau)\mathrm{d}\tau}{\int_0^\infty \varphi_c(\tau)\mathrm{d}\tau}} \tag{1.96}$$

$$\sigma_\tau = \sqrt{\frac{\sum_n \sigma_{a_n}^2 \tau_n^2}{\sum_n \sigma_{a_n}^2} - m_\tau^2} \tag{1.97}$$

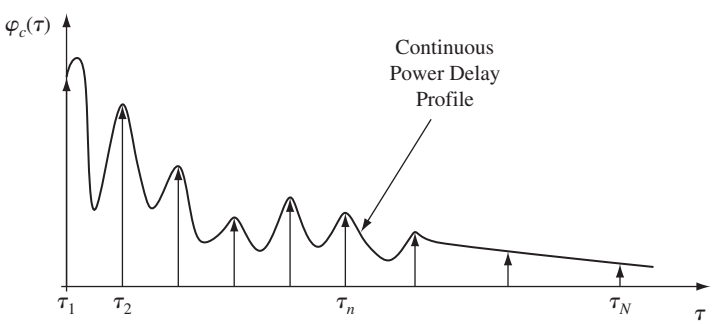

Figure 1.10 Transformation of continuous power delay profile into a discrete one

Mean excess delay and rms delay spread are associated with the power delay profile $\varphi_c(\tau)$ of the radio channel. In fact, assuming that $\varphi_c(\tau)$ represents a probability distribution function, mean excess delay corresponds to the mean value of the delay, in which a very narrow pulse is subjected. Furthermore, the rms delay spread gives the spread of the delays around the mean delay of the pulse.

Frequency selectivity is characterized by a parameter called coherence bandwidth of the channel $[\Delta f]_C$. This is defined as the bandwidth within which the channel frequency response is approximately flat. This implies that for $\Delta f \geq [\Delta f]_C$, $\varphi_C(\Delta f) \approx 0$ and consequently the channel responses for frequency difference exceeding the coherence bandwidth are uncorrelated.

As frequency correlation function $\varphi_C(\Delta f)$ is the Fourier transform of power delay profile $\varphi_c(\tau)$, the following relation holds between rms delay spread and coherence bandwidth:

$$(\Delta f)_{\text{coh}} \cong \frac{k_\tau}{\sigma_\tau} \tag{1.98}$$

where k_τ depends on the shape of $\varphi_c(\tau)$ and the value at which we use for correlation. Most of the times we use $k_\tau = 1$.

Figure 1.11 shows the shape of a power delay profile and its Fourier transform from where the coherence bandwidth can be calculated.

Dispersion in Doppler Frequency and Time Selectivity

In analogy with time delay, $S_C(\nu)$ is used to determine the spread in Doppler frequency. More specifically, when there is Doppler effect due to movement of the mobile unit, a single carrier f_C transmitted through the channel, produces a Doppler spectrum occupying a frequency band $[f_C - \nu_D, f_C + \nu_D]$, where ν_D indicates the maximum Doppler frequency. The U-shaped Doppler is an example of this distribution. The mean Doppler frequency and rms Doppler spread are given respectively by:

$$m_\nu \equiv \int_{-\infty}^{+\infty} \nu S_C(\nu) \mathrm{d}\nu \tag{1.99}$$

$$\sigma_\nu \equiv \sqrt{\int_{-\infty}^{+\infty} (\nu - m_\nu)^2 S_C(\nu) \mathrm{d}\nu} \tag{1.100}$$

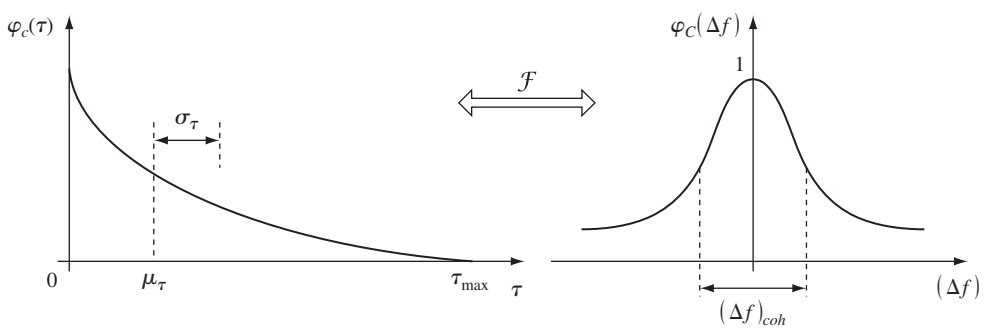

Figure 1.11 RMS delay spread and coherence bandwidth depicted on power delay profile and its autocorrelation function

In practice, the rms delay spread ΔF_D gives the bandwidth over which $S_C(\nu)$ is not close to zero.

To quantify how fast the radio channel changes with time, the notion of coherence time is introduced, which represents the time over which the time correlation function $\varphi_C(\Delta t)$ is not close to zero:

$$T_C : \varphi_C(\Delta t) \approx 0 \quad \text{for } \Delta t \geq T_C \tag{1.101}$$

More specifically, coherence time T_C indicates the time interval during which the channel impulse response does not change significantly. Coherence time T_C and rms Doppler spread are connected with an inverse relation of the form:

$$T_C \approx \frac{k}{\sigma_\nu} \tag{1.102}$$

Figure 1.12 shows $\varphi_C(\Delta t)$ and $S_C(\nu)$ and graphically depicts parameters ΔF_D and T_C.

1.4.3.4 Characterization of Small-scale Fading

Based on the above parameters of selectivity in frequency and time, small-scale fading can be classified in four categories. Two of them concern frequency selectivity and the other two are associated with selectivity in time.

Frequency Selectivity

By comparing the bandwidth W_{BP} of the bandpass signal to the coherence bandwidth of the channel $(\Delta f)_{\text{coh}}$, we classify fading in flat fading and frequency selective fading, given by the following criteria:

$$W_{\text{BP}} \ll (\Delta f)_{\text{coh}} \Rightarrow \text{Flat Fading} \tag{1.103}$$

$$W_{\text{BP}} \gg (\Delta f)_{\text{coh}} \Rightarrow \text{Frequency Selective Fading} \tag{1.104}$$

In the case of flat fading the relation between the input and output (transmitted and received) baseband signals is very simple and is given by:

$$r_{\text{L}}(t) = C(0;t)\int_{-\infty}^{+\infty} S_{\text{L}}(f)\mathrm{e}^{j2\pi ft}\mathrm{d}f = C(0;t)s_{\text{L}}(t) \tag{1.105}$$

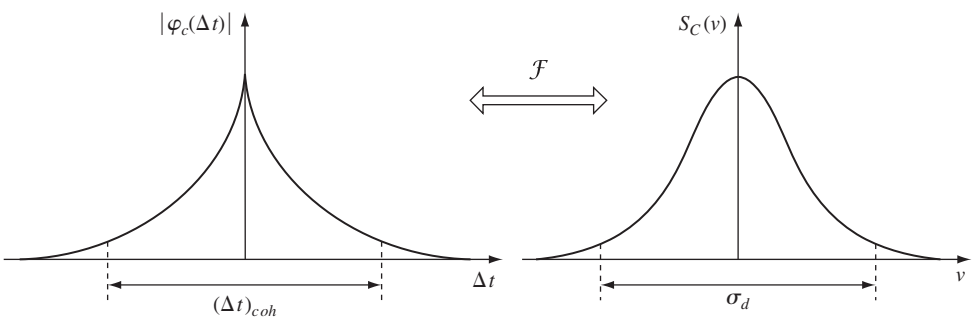

Figure 1.12 Coherence time and Doppler spread estimated from channel autocorrelation function and Doppler power spectrum

where

$$C(0;t) = \sum_n a_n(t) \mathrm{e}^{-j2\pi f_C \tau_n(t)} \tag{1.106}$$

In addition, in the time domain, the two conditions are expressed in terms of the rms delay spread and the signalling interval T ($T = 1/W_{\mathrm{BP}}$):

$$T \gg \sigma_\tau \Rightarrow \text{Flat Fading} \tag{1.107a}$$

$$T \ll \sigma_\tau \Rightarrow \text{Frequency Selective Fading} \tag{1.107b}$$

It is important to note that the above relations indicate that in a frequency-selective channel, due to the relation $T_S \ll \sigma_\tau$, the channel introduces ISI.

Finally, we must note that the characterization of fading as frequency selective or not depends only on the bandwidth of the transmitted signal, compared with the channel frequency response.

Time Selectivity

In this case, the symbol interval T is compared with coherence time T_C. If the channel impulse response does not change significantly within the symbol interval, we have slow fading. In the opposite case, fading is characterized as fast fading. These two are expressed as:

$$T \ll (\Delta t)_{\mathrm{coh}} \Rightarrow \text{Slow Fading} \tag{1.108a}$$

$$T \gg (\Delta t)_{\mathrm{coh}} \Rightarrow \text{Fast Fading} \tag{1.108b}$$

Furthermore, in terms of Doppler spread we have:

$$W_{\mathrm{BP}} \gg \sigma_\nu \Rightarrow \text{Slow Fading} \tag{1.109a}$$

$$W_{\mathrm{BP}} \ll \sigma_\nu \Rightarrow \text{Fast Fading} \tag{1.109b}$$

As with frequency selectivity, characterizing the channel as slow fading or fast fading mostly depends on the transmitted signal bandwidth. However, in this case, since the channel statistics can change due to change in Doppler frequency, which depends on the change in Doppler, the channel can be transformed from slow fading to fast fading and vice versa.

Figure 1.13 depicts graphically the radio channel categorization in terms of frequency selectivity and time selectivity as discussed above.

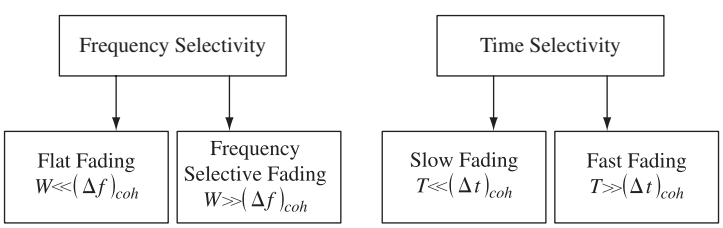

Figure 1.13 Categorization of channel properties with respect to selectivity in frequency and time

1.4.3.5 The Tapped Delay-line Channel Model

For a baseband transmitted signal of bandwidth W the received signal is given by:

$$r_L(t) = \sum_{n=1}^{L_{taps}} c_n(t) s_l\left(t - \frac{n-1}{W}\right) \tag{1.110}$$

It can be shown that [Proakis02] the coefficients are given by:

$$c_n(t) = \frac{1}{W} c\left(\frac{n-1}{W}; t\right) \tag{1.111a}$$

The number of taps is given by:

$$L_{taps} = \lceil \tau_{\max} W \rceil \text{ where} \lceil x \rceil = N_x, N_x \leq x < N_x + 1 \tag{1.111b}$$

Finally, it should be noted that the impulse response $c(\tau; t)$ as a function of the tap values is:

$$c(\tau; t) = \sum_{n=1}^{L_{taps}} c_n(t) \delta\left(t - \frac{n-1}{W}\right) \tag{1.112}$$

Coefficients $\{c(\tau; t)\}$ are uncorrelated and follow a complex Gaussian distribution. When there is no line-of-sight (N-LOS), the magnitude of $\{c(\tau; t)\}$ follows a Rayleigh distribution with uniform distribution in phase.

1.5 Nonlinearity and Noise in Radio Frequency Circuits and Systems

An RF system is mainly plagued by two shortcomings produced by respective nonidealities. These are nonlinearity and noise. These two factors greatly affect its performance and define the region of useful operation of the RF front-end.

1.5.1 Nonlinearity

Let $x(t)$ represent the input of a system which by means of an operator $L\{\cdot\}$ produces output $y(t)$:

$$y(t) = L\{x(t)\} \tag{1.113}$$

This system is linear if the superposition principle is satisfied at its output. The linear combination of two input signals $x_1(t)$ and $x_2(t)$ produces the linear combination of the corresponding outputs:

$$\begin{aligned} x(t) &= C_1 x_1(t) + C_2 x_2(t) \\ y(t) &= C_1 L\{x_1(t)\} + C_2 L\{x_2(t)\} \end{aligned} \tag{1.114}$$

Furthermore, a system is memoryless if at a time instant t its output depends only on the current value of the input signal $x(t)$ and not on values of it in the past:

$$y(t) = Cx(t) \tag{1.115}$$

In this subsection we present the effects of nonlinear memoryless systems that can be described by the general transfer function:

$$y(t) = \sum_{n=0}^{N} \alpha_n x^n(t) \tag{1.116}$$

The above equation specifies a system with nonlinearities of order N.

1.5.1.1 Output Saturation, Harmonics and Desensitization

We assume the input of a memoryless nonlinear system of order $N = 3$ is a simple sinusoid $x(t) = A\cos\omega t$. The output is given by:

$$y(t) = \sum_{n=1}^{3} a_n (A\cos\omega t)^n = \frac{1}{2}a_2A^2 + \left(a_1A + \frac{3}{4}a_3A^3\right)\cos(\omega t) + \frac{1}{2}a_2A^2\cos(2\omega t) + \frac{1}{4}A^3\cos(3\omega t) \tag{1.117}$$

If we had an ideal linear system the output would be:

$$y(t) = a_1x(t) \tag{1.118}$$

Coefficient a_1 constitutes the small-signal gain of the ideal system.

Hence, the above equation differs from the ideal in that it contains terms with frequencies 2ω and 3ω. These terms constitute harmonics of the useful input. There exist harmonics of all orders up to the order of the nonlinearity (in this case second and third).

In addition, from the above equation we see that the coefficient of the useful term $\cos(\omega t)$ contains one more term $(3a_3A^3/4)$ compared with the ideal. Assuming that the nature of nonlinearity does not change with the input signal, a_3 will have a constant value. Hence, the overall, nonideal gain is now a nonlinear function of the input amplitude A:

$$G(A) = A\left[a_1 + \frac{3}{4}a_3A^2\right] \tag{1.119}$$

Because, in most cases, $a_3 < 0$, when A starts increasing beyond a point, $G(A)$ stops increasing linearly and starts saturating. Figure 1.14 shows this variation. A usual measure to determine nonlinearity implicitly is the '1-dB compression point', which is defined as the input amplitude at which the output differs by 1 dB with respect to the ideally expected output amplitude. Using Equation (1.117) we get:

$$20\log\left|\frac{a_1 + \frac{3}{4}a_3A_{1\,\mathrm{dB}}^2}{a_1}\right| = -1\,\mathrm{dB} \quad \text{or} \quad A_{1\,\mathrm{dB}} = \sqrt{0.145\left|\frac{a_1}{a_3}\right|} \tag{1.120}$$

Desensitization concerns the impact of a high power interfering signal. To quantify the effect we assume that the input $x(t)$ of the system consists of a useful signal with amplitude A_1 at frequency f_1 and a strong interfering signal with amplitude A_2 at frequency f_2.

$$x(t) = A_1\cos(\omega_1 t) + A_2\cos(\omega_2 t) \tag{1.121}$$

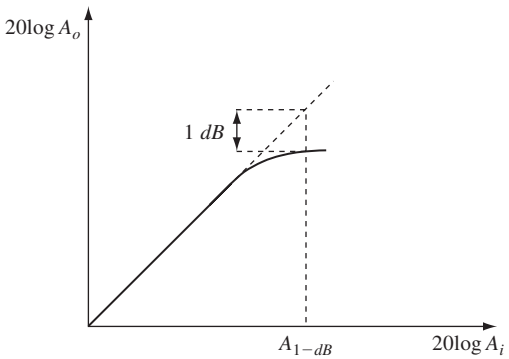

Figure 1.14 Determination of 1 dB compression point in nonlinear transfer function

Taking into account that $A_2 \gg A_1$, a good approximation of the output $y(t)$ is given as [Razavi98]:

$$y(t) = A_1\left[a_1 + \frac{3}{2}a_3 A_2^2\right]\cos\omega_1 t + \cdots \tag{1.122}$$

It is easy to realize that if $a_3 < 0$ the amplitude factor $[a_1 + (3a_3A_2^2)/2]$, which represents the gain at the output of the system, will keep reducing until it gets close to zero. In this case, the receiver is desensitized by the strong interferer and the weaker useful signal is 'blocked'.

1.5.1.2 Distortion due to Intermodulation Products

Intermodulation (IM) products are created by the mixing operation applied on two signals appearing at the input of the receiver at different frequencies. The mixing operation is due to nonlinearities and produces signals falling within the demodulation bandwidth of the receiver. Assuming again that the input is given by two signals of comparable amplitudes as in Equation (1.121), third-order nonlinearity will give an output:

$$y(t) = \sum_{n=1}^{3} a_n(A_1\cos\omega_1 t + A_2\cos\omega_2 t)^n \tag{1.123}$$

Expanding Equation (1.123) and using trigonometric identities will give intermodulation products and harmonics the number and strength of which is depicted in Table 1.1.

The most important intermodulation products are, depending on the receiver architecture, the third order and second order. Most wireless communication systems have considerable numbers of user channels (10–100) close to each other. Neighbouring channels are located at distances equal to the channel bandwidth of the system. Hence, in such a system, two neighbouring channels, ch1 and ch2, with strong amplitudes, can create serious problems to the third successive channel ch3 because the third-order products will fall into its useful band. This is depicted in Figure 1.15. If the receiver is tuned to ch3, the neighbouring channels ch1 and ch2 will be treated as interferers and they will deteriorate the SINR (signal-to-interference and-noise ratio).

Table 1.1 Intermodulation products and harmonics and their respective amplitudes

Frequency	Component amplitude
ω_1	$a_1A_1 + \frac{3}{4}a_3A_1^3 + \frac{3}{2}a_3A_1A_2^2$
ω_2	$a_1A_2 + \frac{3}{4}a_3A_2^3 + \frac{3}{2}a_3A_2A_1^2$
$\omega_1 \pm \omega_2$	$a_2A_1A_2$
$2\omega_1 \pm \omega_2$	$\frac{3a_3A_1^2A_2}{4}$
$2\omega_2 \pm \omega_1$	$\frac{3a_3A_1A_2^2}{4}$

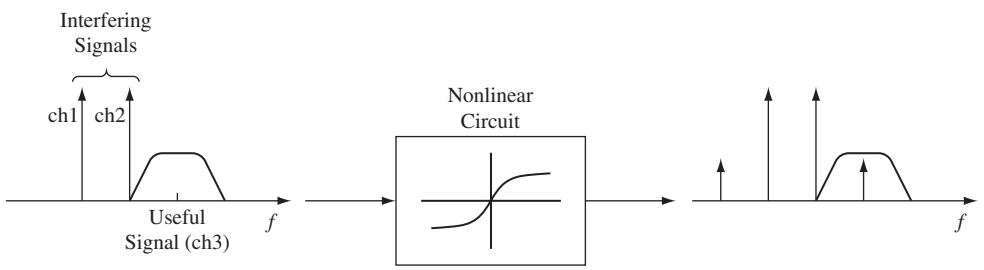

Figure 1.15 Useful signal and two interfering signals passing through a nonlinear circuit

To characterize third-order IM products, the parameter 'third-order intercept point' is used, also known as IP3. To evaluate IP3 we assume that the input $U_{\text{IM}}(t)$ of the nonlinear device creating IM distortion is the summation of two equal amplitude sinusoids tuned at ω_1 and ω_2:

$$U_{\text{IM}}(t) = A\cos\omega_1 t + A\cos\omega_2 t \tag{1.124}$$

Using Table 1.1, the corresponding output, taking into account only the terms at frequencies ω_1, ω_2, $2\omega_1 - \omega_2$ and $2\omega_2 - \omega_1$, is given as:

$$U_{\text{IM-O}}(t) = \left[a_1 + \frac{9}{4}a_3A^2\right]U_{\text{IM}}(t) + \frac{3}{4}a_3A^3[\cos(2\omega_1 - \omega_2)t + \cos(2\omega_2 - \omega_1)t] \tag{1.125}$$

The input IP3 is defined as the input amplitude for which the components at ω_1, ω_2 and $2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$ of the output signal, have equal amplitudes:

$$\left[a_1 + \frac{9}{4}a_3A_{\text{IP3}}^2\right]A_{\text{IP3}} = \frac{3}{4}a_3A_{\text{IP3}}^3 \tag{1.126}$$

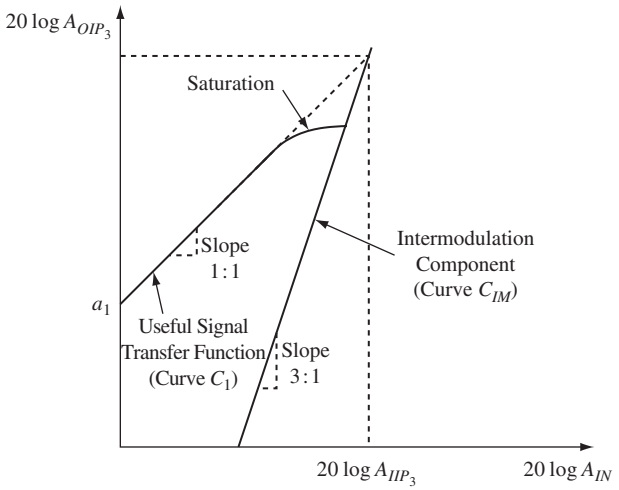

Figure 1.16 Intersection of ideal transfer function and third-order IM component

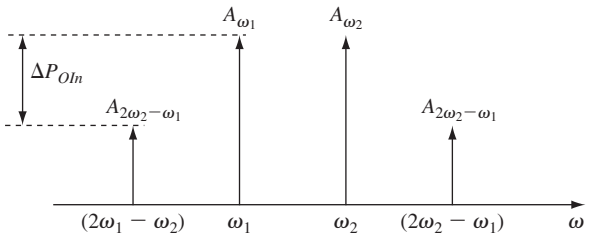

Figure 1.17 Output spectrum of nonlinear circuit when input consists of two tones at frequencies ω_1 and ω_2

Assuming $a_1 \gg \frac{9}{4}a_3A_{\text{IP3}}^2$, the input and output IP3 are given as:

$$A_{\text{IP3}} = \sqrt{\frac{4}{3}\left|\frac{a_1}{a_3}\right|}, \quad A_{\text{OIP3}} = a_1 A_{\text{IP3}} \tag{1.127}$$

However, as we mentioned before, when the concept of saturation was presented, the above assumption for α_1 is not valid for high input amplitude and the $U_O(t)$ exhibits saturation.

Figure 1.16 shows geometrically the relations between the useful transfer function curve C_1 and the IM curve C_{IM} plotted in logarithmic scale. The above equations indicate that the slope of the latter is three times the slope of C_1 (3 dB/dB vs. 1 dB/dB). Furthermore, due to saturation the intercept point cannot be measured experimentally. However, the measurement can be done indirectly by applying the two tones at the input and noting the difference $\Delta P_{\text{O-IM}}$ at the output of the components at frequencies ω_1, ω_2 and $2\omega_1 - \omega_2$, $2\omega_2 - \omega_1$.

Figure 1.17 shows this relation graphically. Elementary calculations using the above formulas (1.125) and (1.126) show that this relation is [Razavi98]:

$$20\log A_{\text{IP3}} = \frac{1}{2}20\log\left(\frac{A_{\omega_1}}{\text{A}_{2\omega_1-\omega_2}}\right) + 20\log A_{\text{in}} \tag{1.128}$$

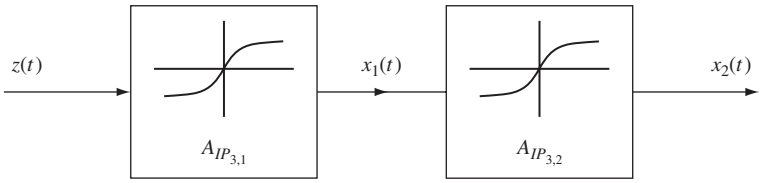

Figure 1.18 Cascading of two nonlinear stages with given IP_3

1.5.1.3 The Third-order Intercept Point of Cascaded Stages

Assume we have two successive stages with order of nonlinearity equal to 3 at the receiver, as illustrated in Figure 1.18. The input $z(t)$ of the first stage produces an output $x_1(t)$ which is used as the input of the next stage giving $x_2(t)$ at its output. The signals $x_1(t)$ and $x_2(t)$ are given as:

$$x_1(t) = \sum_{n=1}^{3} a_n z^n(t) \tag{1.129}$$

$$x_2(t) = \sum_{n=1}^{3} b_n x_1^n(t) \tag{1.130}$$

It is easy to express $x_2(t)$ as a function of $z(t)$ by eliminating $x_1(t)$:

$$x_2(t) = \sum_{n=1}^{6} K_n z^n(t) \tag{1.131}$$

Calculations show that K_1 and K_3 are given by [Razavi98]:

$$K_1 = a_1 b_1, \quad K_3 = a_3 b_1 + 2a_1 a_2 b_2 + a_1^3 b_3 \tag{1.132}$$

A_{IP3} is given by dividing K_1 by K_3 to get:

$$A_{\text{IP3}} = \sqrt{\frac{4}{3}\left|\frac{a_1 b_1}{a_3 b_1 + 2a_1 a_2 b_2 + a_1^3 b_3}\right|} \tag{1.133}$$

and consequently:

$$\frac{1}{A_{\text{IP3}}^2} = \frac{1}{A_{\text{IP3,1}}^2} + \frac{3a_2 b_2}{2b_1} + \frac{a_1^2}{A_{\text{IP3,2}}^2} \tag{1.134}$$

The terms $A_{\text{IP3},i}^2, i = 1, 2$ give the input IP3 of the two successive stages 1 and 2 respectively.

Finally, considering low values for second-order coefficients a_2, b_2 we get:

$$\frac{1}{A_{\text{IP3}}^2} \approx \frac{1}{A_{\text{IP3,1}}^2} + \frac{a_1^2}{A_{\text{IP3,2}}^2} \tag{1.135}$$

When there are more than two stages we have [Razavi98]:

$$\frac{1}{A_{\text{IP3}}^2} \approx \frac{1}{A_{\text{IP3,1}}^2} + \frac{a_1^2}{A_{\text{IP3,2}}^2} + \frac{a_1^2 b_1^2}{A_{\text{IP3,3}}^2} + \cdots \tag{1.136}$$

1.5.2 Noise

As mentioned in the beginning of this section, apart from nonlinearities noise is another factor that has considerable impact on the circuit and system performance. The aim of this subsection is to identify the noise sources in electronic elements (active and passive) and, based on these, to examine the parameters and techniques that will give the noise behaviour of complete circuits and systems.

1.5.2.1 The Noise Model of Bipolar Transistors

The fluctuation of the collector current due to the crossing of holes and electrons through the PN base–collector junction of a bipolar transistor, creates shot noise on the collector current and its spectral density is given by:

$$\overline{i_C^2} = 2qI_C\Delta f \tag{1.137}$$

A similar noise component (shot noise) is created at the base of the transistor originating from recombination at the base and carrier-crossing through the base–emitter junction. Apart from that, it was shown experimentally that two other sources of noise (flicker and burst) contribute to the overall noise at the base of the transistor. Flicker noise has a $1/f$ dependence on frequency whereas burst noise has $1/[1+(f/f_C)^2]$ frequency dependence. Hence, the base noise spectral density for a bandwidth Δf is given as:

$$\overline{i_{\text{B}}^2} = 2qI_{\text{B}}\Delta f + K_{\text{FN}}\frac{I_{\text{B}}^{\text{f}}}{f}\Delta f + K_{\text{BN}}\frac{I_{\text{B}}^{\text{b}}}{1+(f/f_C)^2}\Delta f \tag{1.138}$$

where $K_{\text{FN}}, K_{\text{BN}}$ are multiplication constants and f, b are exponents associated with the base current I_{B} for flicker and burst noise, respectively.

Finally, there is thermal noise associated to physical resistance at the three terminals, base, collector and emitter, of the transistor:

$$\overline{i_{rb}^2} = \frac{4kT}{r_b}\Delta f, \quad \overline{i_{rc}^2} = \frac{4kT}{r_c}\Delta f, \quad \overline{i_{re}^2} = \frac{4kT}{r_e}\Delta f \tag{1.139}$$

where, k is the Boltzman constant and T is the absolute temperature (usually 290 K). We must note that $\overline{i_{rb}^2}$ is the most important because the other current densities have very low values.

Taking into account the above presentation of noise sources, Figure 1.19 depicts the noise model of bipolar transistors. Note that r_π, r_μ and r_o represent resistances in the noise model without contributing any thermal noise.

1.5.2.2 The Noise Model of MOS Transistors

The major contribution of noise in FETs comes from the channel that exhibits thermal noise due to its resistive behaviour [Gray01]. This is due to modulation of the channel by V_{GS}.

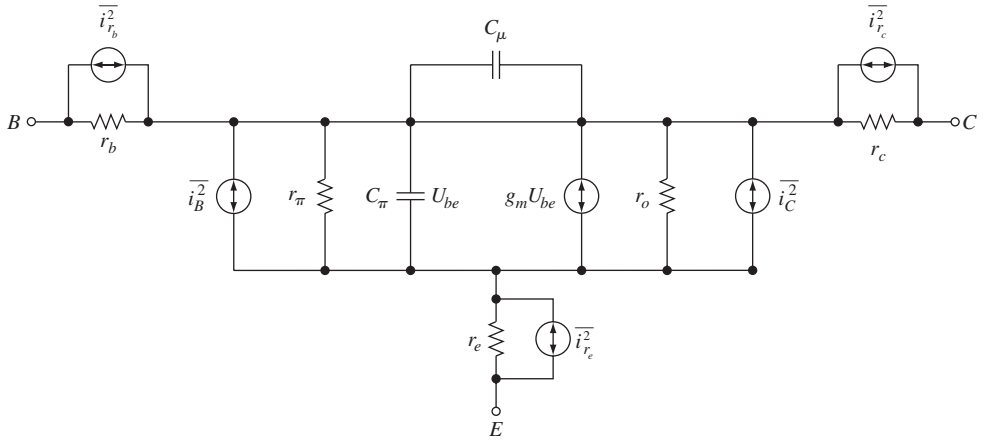

Figure 1.19 Noise model of bipolar transistor

In addition, experiments have shown that a flicker noise term can be included in the drain current spectral density $\overline{i_d^2}$:

$$\frac{\overline{i_d^2}}{\Delta f} = 4kT\gamma g_{do} + K_C \frac{I_d^m}{f}, \quad m = 0.5 - 2 \tag{1.140}$$

where g_{do} is the drain conductance of the device under zero-bias conditions and I_d is the drain current. γ is a factor taking values depending on the bias and the type of the device:

$$\left.\begin{array}{ll} \dfrac{2}{3} \le \gamma \le 1 & \text{for long channel devices} \\ 2 \le \gamma \le 3 & \text{for short channel devices (L} \le 0.7\,\mu\text{m)} \end{array}\right\} \tag{1.141}$$

Furthermore, a significant noise contributor in MOS devices is the induced gate noise current which is due to fluctuations of the channel charge originating from the drain noise current $4kT\gamma g_{do}$ [Shaeffer97]. This is given by:

$$\frac{\overline{i_g^2}}{\Delta f} = 4kT\delta g_g, \quad g_g = \frac{\omega^2 C_{gs}^2}{5 g_{do}} \tag{1.142}$$

and δ is a coefficient associated with gate noise.

Because induced gate noise is related to the drain noise as seen from the above expressions, the two noise factors are partially correlated:

$$\frac{\overline{i_g^2}}{\Delta f} = 4kT\delta g_g(1 - |c|^2) + 4kT\delta g_g |c|^2 \tag{1.143}$$

$$|c| = \left| \frac{\overline{i_g i_d^*}}{\sqrt{\overline{i_g^2} \cdot \overline{i_d^2}}} \right| \tag{1.144}$$

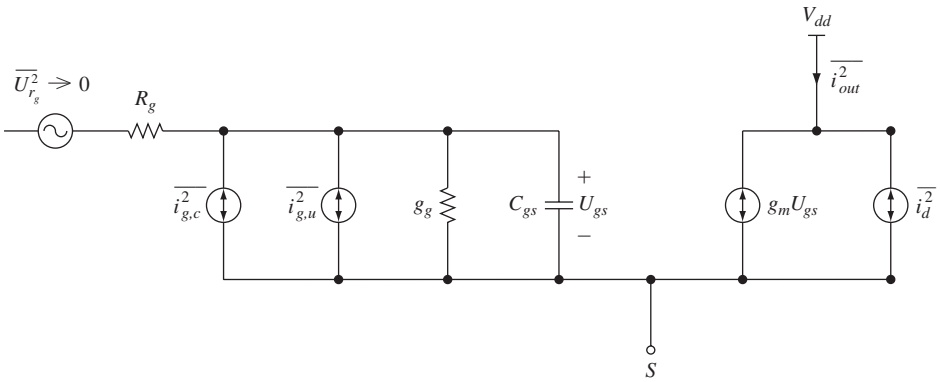

Figure 1.20 MOS transistor noise model

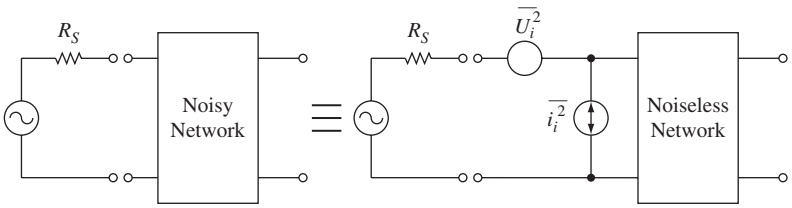

Figure 1.21 Transformation of noisy into a noiseless network with noise sources at the input

Another noise factor, similar to that of bipolar transistors, is shot noise current at the gate:

$$\frac{\overline{i_{gs}^2}}{\Delta f} = 2qI_{\mathrm{G}} \tag{1.145}$$

This is due to gate leakage current and it has been experimentally found that its contribution to the gate noise current is significant at sub-GHz frequencies [Scholten03]. However, the gate noise current is dominated by the induced gate noise as in Equation (1.142) at frequencies above approximately 1 GHz.

Finally, there is a noise voltage due to the distributed gate resistance R_g. Taking into account all of the above, a reasonably accurate equivalent noise model of a MOS device for $f \geq 1$ GHz is as in Figure 1.20. Note that the shot noise component is not included for the reasons presented above and $\overline{U_{Rg}^2}$ is negligible.

1.5.2.3 Noise Performance of Two-port Networks and Noise Figure

According to the two-port noisy network theorem, any two-port noisy network is equivalent to a noiseless network with two noise sources, a voltage and a current noise source, connected at its input, as shown in Figure 1.21. When correlation of the two noise sources $\overline{U_{\mathrm{in}}^2}$ and $\overline{i_{\mathrm{in}}^2}$ is taken into account, the model is valid for any kind of source impedance.

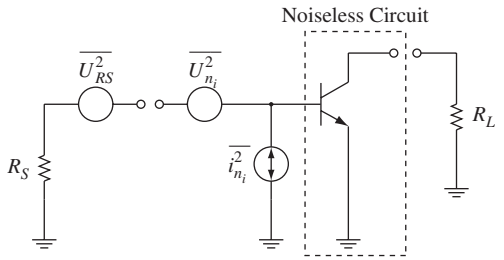

Figure 1.22 Equivalent two-port noise model of bipolar transistor

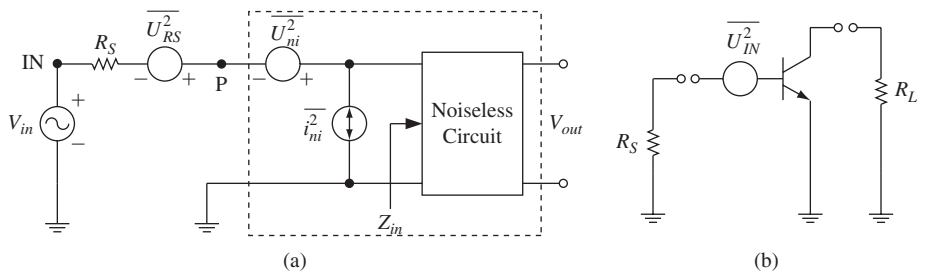

Figure 1.23 (a) Two-port network including all noise sources for NF calculation, (b) representation of overall noise (including noise from Rs) with $\overline{U_{\text{IN}}^2}$

The above theorem can be applied to a bipolar transistor to produce the equivalent two-port noise model, as shown in Figure 1.22, with input noise voltage and current given by the following expressions [Meyer94]:

$$\overline{U_{\text{ni}}^2} = 4kT[r_b + 1/(2g_m)]\Delta f, \quad \overline{i_{\text{ni}}^2} = 2q\left[I_B + \frac{I_C}{|\beta(j\omega)|^2}\right]\Delta f \tag{1.146}$$

where flicker noise has been neglected in the calculation of noise current.

The SNR is defined as the ratio between signal power and noise power at any point along the receiver chain. The noise figure (NF) or noise factor (F) of a circuit (or system) along the receiver is defined as the amount by which SNR deteriorates when passing through it. Quantitatively F is expressed as:

$$F = \frac{\text{SNR}_{\text{in}}}{\text{SNR}_{\text{out}}} \tag{1.147}$$

The noise figure is given by $NF = 10\log(F)$.

Figure 1.23(a) shows a two-port network including all noise sources, which is used to find a suitable expression for the noise factor. The source resistance R_S produces noise equal to:

$$\overline{V_S^2} = 4kTR_S \tag{1.148}$$

We assume that the voltage gain is A_1 from point IN to point P and A_2 from point P to the output. The input SNR is measured at point P and is equal to the ratio of signal power produced

by an input signal of amplitude V_{in} to the noise due to the source resistance R_S. Consequently, the expressions for SNR_{in} and SNR_{out} are as follows [Razavi98]:

$$\text{SNR}_{\text{in}} = \frac{A_1^2 V_{\text{in}}^2}{A_1^2 \overline{V_{Rs}^2}} \tag{1.149}$$

$$\text{SNR}_{\text{out}} = \frac{V_{\text{in}}^2}{\left[\overline{V_{Rs}^2} + \overline{(U_{\text{ni}} + I_{\text{ni}} R_S)^2}\right] A_1^2 A_2^2} \tag{1.150}$$

Further calculations produce the following expression for F, useful for measurement and simulation:

$$\begin{aligned} F &= \frac{\left[\overline{V_{Rs}^2} + \overline{(U_{\text{ni}} + I_{\text{ni}} R_S)^2}\right]}{(A_1 A_2)^2} \frac{1}{\overline{V_{Rs}^2}} \\ &= \left(\frac{\text{Overall output noise power}}{\text{Voltage gain from IN to OUT}}\right) \Big/ (\text{Noise due to } R_S) = \frac{\overline{V_{n,o}^2}}{(A_1 A_2)^2} \frac{1}{4kTR_S} \end{aligned} \tag{1.151}$$

As an example of how to use the above considerations to determine the noise figure we use a simple common emitter amplifier followed by subsequent stages, as shown in Figure 1.23(b). Elimination of $(A_1 A_2)^2$ in the first expression of Equation (1.151) shows that the noise factor can also be calculated by dividing the overall equivalent noise at the input $\overline{V_{n,\text{In}}^2} = \overline{V_{Rs}^2} + \overline{(U_{\text{ni}} + I_{\text{ni}} R_S)^2}$ by the noise due to R_S. Assuming $U_{\text{ni}}, I_{\text{ni}}$ are statistically independent, $\overline{V_{n,In}^2}$ becomes:

$$\overline{V_{n,\text{In}}^2} = \overline{V_{Rs}^2} + \overline{U_{\text{ni}}^2} + \overline{I_{\text{ni}}^2} R_S^2 \tag{1.152}$$

Using the expressions for $\overline{U_{\text{ni}}^2}, \overline{I_{\text{ni}}^2}$ from Equation (1.146), we can finally obtain:

$$\begin{aligned} F &= \frac{\overline{V_{n,\text{In}}^2}}{\overline{V_{Rs}^2}} = \frac{4kTR_S + 4kT[r_b + 1/(2g_m)] + 2q[I_B + I_C/|\beta(j\omega)|^2]R_S}{4kTR_S} \\ &= 1 + \frac{r_b}{R_S} + \frac{1}{2g_m R_S} + \frac{g_m R_S}{2\beta(0)} + \frac{g_m R_S}{2|\beta(\text{j}\omega)|^2} \end{aligned} \tag{1.153}$$

$\beta(0)$ is the value of β at DC. It must be noted that the load resistance appearing at the collector due to subsequent stages was not taken into account in this calculation.

1.5.2.4 Noise Figure of N Cascaded Stages

Let N stages at the input of the receiver after the antenna are connected in cascade, as shown in Figure 1.24. Let A_i denote the power gain of the ith stage assuming conjugate matching at both input and output of the corresponding stage. By replacing in the cascaded structure each stage by a noiseless two-port network with noise voltage and current connected at its input, as

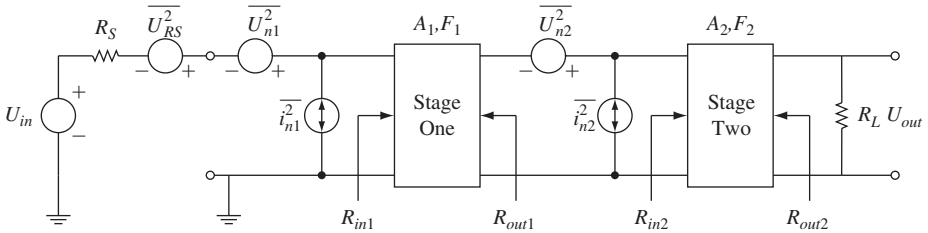

Figure 1.24 Noise model of two receiver stages connected in cascade

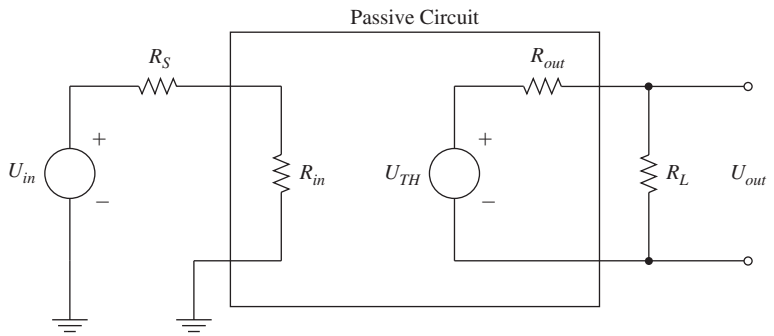

Figure 1.25 Noise model of passive circuit

shown in Figure 1.21, it can be shown after mathematical manipulation that the overall noise factor is given by [Razavi98]:

$$F_T = 1 + (F_1 - 1) + \frac{F_2 - 1}{A_1} + \frac{F_3 - 1}{A_1 A_2} + \cdots + \frac{F_n - 1}{A_1 A_2 \ldots A_{N-1}} \tag{1.154}$$

This is the Friis noise formula and computes the overall noise factor for N cascaded stages along the receiver chain. Noise factors at each stage are determined by taking into account the impedance appearing at the input of the corresponding stage.

1.5.2.5 Noise Figure of Passive Circuits

A passive circuit is connected to a voltage input V_{in}, with R_{S} and R_{L} being the source and load resistances. R_{in} and R_{out} represent the input and output resistances respectively, as shown in Figure 1.25. First, we determine the output noise voltage and the overall voltage gain and subsequently we use Equation (1.151) to calculate the noise figure. This is done by employing the Thevenin equivalent of the output circuit. It is then easy to show that noise figure is equal to the losses of the passive circuit [Razavi98]:

$$F = \text{Losses} = (4kTR_{\text{out}}) \left(\frac{V_{\text{in}}}{V_{\text{Th}}} \right)^2 \left(\frac{1}{4kTR_S} \right) \tag{1.155}$$

1.6 Sensitivity and Dynamic Range in Radio Receivers

1.6.1 Sensitivity and Dynamic Range

When the receiver is matched in impedance at its input, the noise power delivered to the receiver per unit bandwidth is equal to kT [Razavi98]. Consequently, for a receiver with information bandwidth B, the minimum detectable signal (MDS) represents the lowest possible power of the signal that just exceeds the noise threshold and is given by:

$$N_{\text{TH}} = N_{\text{MDS}} = kTB \tag{1.156a}$$

For a specific application, the sensitivity of the receiver is defined as the minimum received power for which we can have satisfactory detection. This means that the SNR (or equivalently E_b/N_0), should have the minimum value to guarantee a required bit error rate (BER). For example, a GSM receiver needs approximately 9 dB in SNR in order to achieve satisfactory error rate. In addition, the noise figure of the receiver must be taken into account because, by its nature, it is the principle factor of deterioration of the SNR at the input of the receiver.

Consequently, the sensitivity is given in decibels as follows:

$$N_S = 10\log(kT) + 10\log B + 10\log(E_b/N_o) + NF \tag{1.156b}$$

Another factor that can cause deterioration is the implementation losses L_{imp} associated with implementation of the RF receiver and the modem. For example, synchronization subsystems can contribute an SNR loss in the order of 0.5–2 dB due to remaining jitter of the carrier and time synchronizers. Implementation losses can worsen the receiver sensitivity by a few decibels.

On the other hand, there is a strongest allowable signal that the receiver is able to handle. Above that level, distortion becomes dominant and receiver performance deteriorates rapidly. The difference between the strongest allowable signal and the noise floor is defined as the dynamic range (DR) of the receiver. The highest allowable signal at the input is set as the input power at which the third-order IM products become equal to the system noise floor [Razavi98], [Parssinen01]. Figure 1.26 shows the fundamental and IP3 curves along with the noise floor. Point A is the point representing the strongest allowable signal and consequently DR is equal to the length AB. From the similar triangles it can be easily shown that DR is [Razavi98], [Parssinen01]:

$$DR = \frac{2}{3}(IIP_3 - N_{\text{TH}} - NF) \tag{1.157}$$

For example, a GSM-900 system has a bandwidth of 200 kHz and requires an SNR of 9 dB. If in addition the receiver has noise figure and IP3 equal to NF = 8 dB and IIP3 = −10 dBm, the DR is 68.7 dB.

1.6.2 Link Budget and its Effect on the Receiver Design

Link budget refers to the analysis and calculation of all factors in the transceiver system in order to ensure that the SNR at the input of the digital demodulator is adequate for achieving satisfactory receiver performance according to the requirements of the application. In doing that, transmitted power and antenna gain must be taken into account in the transmitter. The

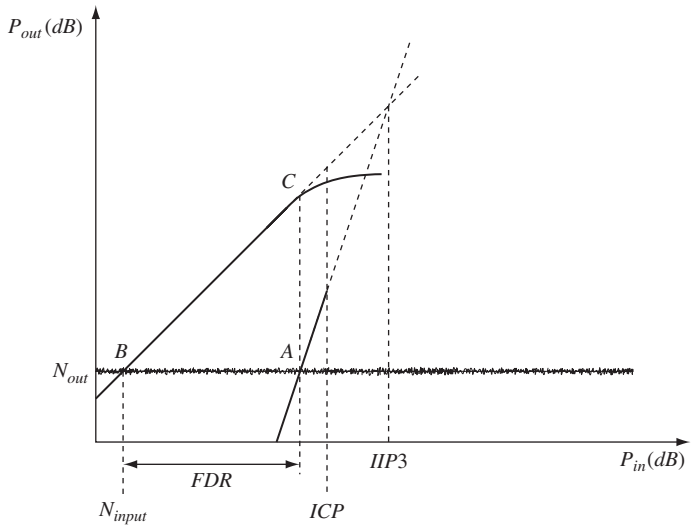

Figure 1.26 Relation between noise floor, IP3 and dynamic range

radio channel losses must be also determined. At the receiver side, sensitivity, noise, distortion and implementation losses have to be carefully determined.

The required sensitivity is a figure determined by taking into account the transmitted power, the gain of the transmitter and receiver antennas, as well as the average expected propagation losses between the transmitter and the receiver for the particular system application:

$$N_{\text{req}} = f(EIRP, G_r, L_{\text{prop}}) \tag{1.158}$$

The overall system design and network planning sets the transmitted power levels, the antenna gains and the expected average propagation losses. An additional parameter is the fade margin associated with signal attenuation due to multipath fading. Fade margin is usually a factor of 2–4 dB and can be accommodated within the average propagation loses L_{prop} factor. Taking all these into account at the right-hand side of Equation (1.158), the required sensitivity N_{req} is determined and dictated by the system specification requirements.

Consequently, Equation (1.156) is used to determine the maximum allowable noise figure NF by setting: $N_{\text{req}} > N_S$. As an example we mention GSM900, requiring a reference sensitivity of -104 dBm for a 9 dB E_b/N_0. Taking into account that the information bandwidth is 200 kHz and assuming zero implementation losses, Equation (1.156b) gives a maximum allowable NF of 8 dB for the overall receiver.

Table 1.2 illustrates, in the form of an example for multiband UWB [Aiello03] system, the procedure of evaluating the basic quantities in the link budget. On the other hand, regarding linearity, the system requirements usually refer to a maximum level of a neighbouring interfering signal under which the useful signal can still be demodulated with satisfactory SNR. This maximum interferer power lets us calculate the input IP3 of the overall receiver. Hence here we do the inverse compared with the calculation of DR above. We use the required specification to calculate the receiver IIP3. For example, in GSM we should have satisfactory demodulation while two interfering signals of power -43 dBm are present, located 0.8 and 1.6 MHz

Table 1.2 Link Budget calculation for multiband UWB

Bit rate (R_b)	112 Mb/s
Transmitted power	−8.3 dBm
Tx, Rx antenna gains G_T, G_R	0
Path loss at 1 m	44.5 dB
Path loss at 10 m	20 dB
Rx power (at 10 m) $= P_r = P_t + G_T + G_R - L_1 - L_2$	−72.8 dBm
Rx noise figure at antenna terminal	7 dB
Noise power per bit $[N_{Th} = -174 + 10\log(R_b) + NF]$	−86.5 dBm
Minimum E_b/N_0	3.6 dB
Implementation losses (IL)	3 dB
Code rate	0.5
Raw bit rate	224 Mb/s
Link margin at 10 m	7.1 dB

away from the useful carrier. Input IP3 can be calculated by using $IIP_3 = P_{\text{INT}} - IM_3/2$. In our example, $P_{\text{INT}} = -43$ dBm and $IM_3 = -104 - 4 - (-43) = -65$ dB. Consequently we get $IIP_3 = -10.5$ dB.

From the above considerations it is easy to realize that the link budget involves a balancing procedure resulting in specific requirements for noise and linearity for the overall receiver. The overall noise and linearity requirements can then be translated into noise and linearity specifications for each circuit in the receiver chain (LNA, filters, mixers, etc.), taking into account the Friis formula for noise figure calculation of the cascaded receiver stages and the corresponding formula for overall IP3 calculation presented in Section 1.5.

1.7 Phase-locked Loops

1.7.1 Introduction

The phase-locked loop is one of the most frequently used subsystems in communications for detection, synchronization and frequency generation. We briefly present below the basic principles and operation of linear and hybrid PLLs. Linear PLLs are the ones using a linear analogue multiplier as a phase detector (PD), whereas 'hybrid' refers to the PLLs that use both digital and analogue components. Sometimes they are also called 'mixed-signal' PLLs [Best03].

Linear and hybrid PLLs represent the great majority of PLLs used in most applications today. All-Digital PLLs (ADPLL) are mostly used in purely digital subsystems where jitter requirements are not very stringent.

1.7.2 Basic Operation of Linear Phase-locked Loops

Figure 1.27 illustrates the basic structure of a linear PLL. Let us assume that its input and output are expressed as follows:

$$\begin{aligned} y_{\text{i}}(t) &= A \cdot \sin[\omega_{\text{i}} t + \varphi_{\text{i}}(t)] \\ y_{\text{o}}(t) &= B \cdot \cos[\omega_{\text{o}} t + \varphi_{\text{o}}(t)] \end{aligned} \tag{1.159}$$

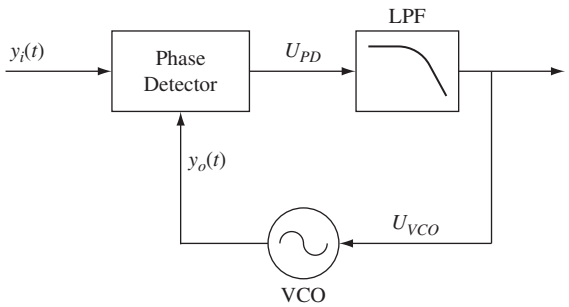

Figure 1.27 Phase-locked loop block diagram

In more detail, it is a feedback system in which the incoming signal $y_{\mathrm{i}}(t)$ and the output signal of the VCO $y_{\mathrm{o}}(t)$ are used as the inputs for a phase detector, which compares the phase quantities $[\omega_{\mathrm{i}}t + \varphi_{\mathrm{i}}(t)]$ and $[\omega_{\mathrm{o}}t + \varphi_{\mathrm{o}}(t)]$ of the two signals. The PD generates an output signal U_{PD} which is proportional to the phase difference of y_{i} and y_{o}. The lowpass filter following the PD passes only the DC component $\overline{U}_{\mathrm{PD}}$. Subsequently, the voltage-controlled oscillator (VCO) is commanded by $\overline{U}_{\mathrm{PD}}$. We use U_{VCO} to represent the tuning voltage at the input of the VCO, which changes its output frequency and phase such that its frequency becomes equal to the frequency of the input signal $y_{\mathrm{i}}(t)$, whereas its phase locks to the phase of $y_{\mathrm{i}}(t)$. The above equations give the state of the system while the feedback path is still open. By closing it, $y_{\mathrm{o}}(t)$ changes its phase and frequency until it locks to the input. At that time the output signal is given by:

$$y_{\mathrm{o}}(t) = B \cdot \cos\left[\omega_{\mathrm{i}}t + \psi_{\mathrm{o}}\right] \tag{1.160}$$

Before the system reaches its steady-state condition (final locking), the output of the phase detector can be represented by the low frequency component of the product of y_{i} and y_{o}:

$$\overline{U}_{\mathrm{PD}} = K_{\mathrm{PD}} \cos\left[(\omega_{\mathrm{i}} - \omega_{\mathrm{o}}(t)) + \varphi_{\mathrm{i}} - \varphi_{\mathrm{o}}(t)\right] \tag{1.161}$$

where K_{PD} is the sensitivity constant of the phase detector and φ_{o} is a function of time representing the dynamic change of the output phase and frequency of the VCO during the locking procedure.

Figure 1.28 shows the procedure of locking when a step Δf occurs at the frequency of the input signal y_{i} at $t = t_0$. The frequency difference is translated into a phase difference $\theta_e = \Delta f \cdot t$. Hence, in terms of phase, in the beginning the phase error keeps increasing. However, application of this DC voltage at the input of the VCO forces it to increase its output frequency which has as a consequence the continuous reduction of Δf and, therefore, of the corresponding phase error. Hence, as depicted in the figure, after an initial increase of the phase error and $\overline{U}_{\mathrm{PD}}$, at t_1 and t_2, the resulting increase of the VCO frequency will decrease the phase error (see time instants t_3, t_4) until the system becomes locked.

The equation of operation of the VCO is given in terms of angular frequency as:

$$\omega_{\mathrm{inst}} = \omega_{\mathrm{o}} + K_{\mathrm{VCO}}U_{\mathrm{VCO}} = \frac{\mathrm{d}}{\mathrm{d}t}[\omega_{\mathrm{o}}(t)t + \varphi_{\mathrm{o}}(t)], \quad \frac{\mathrm{d}\varphi_{\mathrm{o}}(t)}{\mathrm{d}t} = K_{\mathrm{VCO}}U_{\mathrm{VCO}} \tag{1.162}$$

where K_{VCO} is the sensitivity constant of the VCO in Hz/V or rad/sec/V.

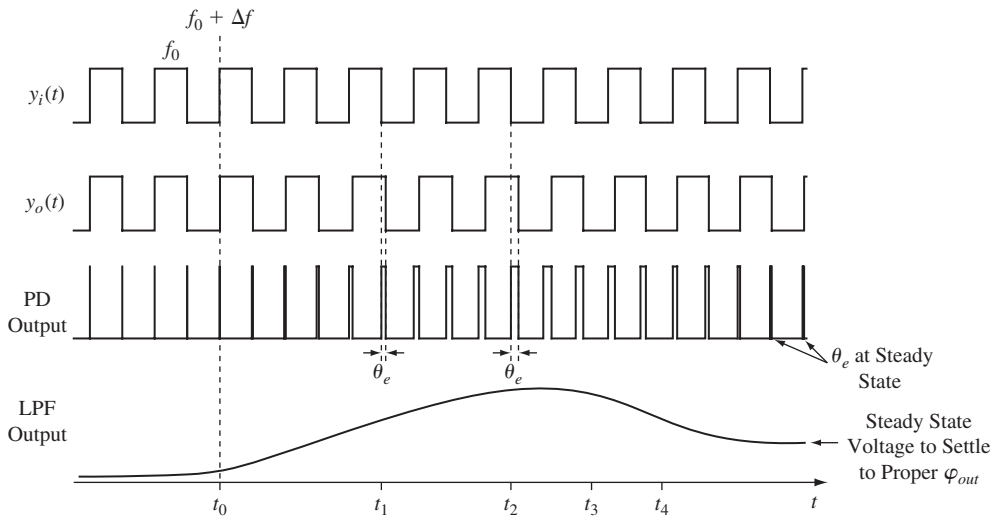

Figure 1.28 PLL's basic behaviour in the time domain when a frequency step of the input signal occurs

If $f(t)$ and $F(j\omega)$ are the impulse response and transfer function of the LPF respectively, the output of the LPF and input of the VCO is given by:

$$U_{\mathrm{VCO}}(t) = U_{\mathrm{PD}}(t) * f(t) \tag{1.163}$$

We assume, for purposes of simplicity, that the frequency of the two signals y_{i} and y_{o} is the same. By combining the above equations, we obtain:

$$\frac{\mathrm{d}\varphi_0(t)}{\mathrm{d}t} = K_{\mathrm{PD}} K_{\mathrm{VCO}} \cdot \left\{ \sin\left[\varphi_i(t) - \varphi_0(t)\right] * f(t) \right\} \tag{1.164}$$

1.7.3 The Loop Filter

The loop filter, apart from eliminating possible high-frequency components at the output of the phase detector, also affects the stability of the feedback system. In most cases, three types of filters are used.

(1) The simple RC filter (which we call S-RC) with transfer function

$$F(j\omega) = \frac{1}{1 + j\omega\tau_1}, \quad \tau_1 = RC \tag{1.165}$$

Figure 1.29(a) shows the circuit and the magnitude of the transfer function.

(2) When the capacitor in parallel is replaced by a combination of a capacitor and a resistor we have what we call the lowpass with phase lead filter (LP-PL). Its transfer function is given by:

$$F(j\omega) = \frac{1 + j\omega\tau_2}{1 + j\omega\tau_1}, \quad \tau_2 = R_2 C, \ \tau_1 = (R_1 + R_2)C \tag{1.166}$$

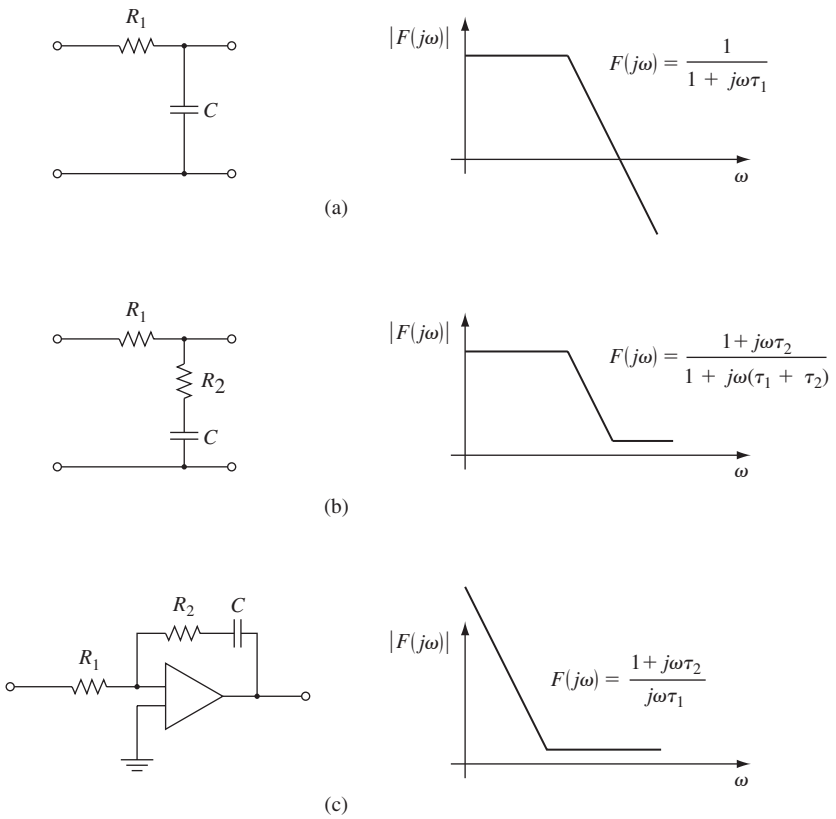

Figure 1.29 Circuits and magnitude of transfer functions of PLL filters: (a) simple RC filter, (b) lowpass filter with phase lead, (c) active filter

Figure 1.29(b) depicts the circuit and transfer function for this type of filter.

(3) The third choice is an active filter, the circuit and transfer function of which are illustrated in Figure 1.29(c). By analysing the circuit we can easily obtain the corresponding equation for the transfer function:

$$F(j\omega) = -G\frac{1+j\omega\tau_2}{1+j\omega\tau_1}, \quad \tau_2 = R_2C,\ \tau_1 = (R_1 + GR_1 + R_2)C \tag{1.167}$$

If the gain of the filter G is high, the above transfer function can be approximated by:

$$F(j\omega) \approx -G\frac{1+j\omega\tau_2}{j\omega\tau_1} \tag{1.168}$$

The denominator corresponds to the integrator function and consequently we call this filter integrator with phase lead filter (I-PL).

1.7.4 Equations and Dynamic Behaviour of the Linearized PLL

We use Equation (1.164) to determine the linear equation for PLL and its transfer function for all three types of filters presented above. For this purpose, we consider that phase difference $(\varphi_i(t) - \varphi_0(t))$ is small and therefore $\sin[\varphi_i(t) - \varphi_0(t)] \approx \varphi_i(t) - \varphi_0(t)$. Hence, the resulting equation becomes:

$$\frac{\mathrm{d}\varphi_0(t)}{\mathrm{d}t} = K_{\mathrm{PD}} K_{\mathrm{VCO}} \cdot \{[\varphi_i(t) - \varphi_0(t)] * f(t)\} \tag{1.169}$$

We assume that the PLL remains locked. By taking Fourier transforms, we obtain the PLL transfer function and error transfer function in terms of the overall open loop gain K after elementary manipulations [Blanchard78]:

$$H(j\omega) = \frac{\Phi_{\mathrm{o}}(j\omega)}{\Phi_{\mathrm{i}}(j\omega)} = \frac{KF(j\omega)}{j\omega + KF(j\omega)}, \quad H_e(j\omega) = \frac{j\omega}{j\omega + KF(j\omega)} \tag{1.170}$$

By inserting in the above equations the expressions for $F(j\omega)$ presented in the previous section, we can determine expressions for the PLL transfer function in terms of K and the time constants of the filters.

In the case where there is no filter in the loop $[F(j\omega) = 1]$, the PLL transfer function is given by:

$$H(j\omega) = \frac{K}{j\omega + K} \tag{1.171}$$

Similarly, for the PLL with filter $F(j\omega) = (1 + j\omega\tau_2)/j\omega\tau_1$, the transfer function becomes:

$$H(j\omega) = \frac{K + j\omega K\tau_2}{K - \omega^2\tau_1 + j\omega K\tau_2} \tag{1.172}$$

Replacing the $j\omega$ operator by the Laplace operator s, we have:

$$H(s) = \frac{(K\tau_2)s + K}{\tau_1 s^2 + (K\tau_2)s + K} \tag{1.173}$$

In similar fashion, for the other two lowpass filters (S-RC and LP-PL) we obtain:

$$H(s) = \frac{K}{\tau_1 s^2 + s + K} \tag{1.174}$$

$$H(s) = \frac{(K\tau_2)s + K}{\tau_1 s^2 + (K\tau_2 + 1)s + K} \tag{1.175}$$

Taking into account the feedback systems theory we observe that, by elementary manipulations, the denominator in all cases can be expressed in the form $s^2 + 2\zeta\omega_n s + \omega_n^2$. After straightforward calculations we obtain the expressions presented in Table 1.3 for the three different filters.

By using Fourier transform properties in the time domain, we can easily derive the differential equations for first-order PLLs (no filter) and second-order PLLs (using one of the previously presented LPFs). For these cases the corresponding equations are [Blanchard78]:

$$\frac{\mathrm{d}\varphi_{\mathrm{o}}(t)}{\mathrm{d}t} + K\varphi_{\mathrm{o}}(t) = K\varphi_{\mathrm{i}}(t) \tag{1.176}$$

Table 1.3 Transfer function and loop parameters for second order PLL with different filters

Filter type	$H(s)$	$\omega_n^2, 2\zeta\omega_n$
$1/(1+j\omega\tau_1)$	$\dfrac{\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\omega_n^2=\dfrac{K}{\tau_1}, \quad 2\zeta\omega_n=\dfrac{1}{\tau_1}$
$\dfrac{(1+j\omega\tau_2)}{j\omega\tau_1}$	$\dfrac{2\zeta\omega_n s+\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\omega_n^2=\dfrac{K}{\tau_1}, \quad 2\zeta\omega_n=\dfrac{K\tau_2}{\tau_1}$
$\dfrac{(1+j\omega\tau_2)}{(1+j\omega\tau_1)}$	$\dfrac{(2\zeta\omega_n-\omega_n^2/K)s+\omega_n^2}{s^2+2\zeta\omega_n s+\omega_n^2}$	$\omega_n^2=\dfrac{K}{\tau_1}, \quad 2\zeta\omega_n=\dfrac{1+K\tau_2}{\tau_1}$

$$\tau_1\frac{\mathrm{d}^2\varphi_\mathrm{o}(t)}{\mathrm{d}t^2}+K\tau_2\frac{\mathrm{d}\varphi_\mathrm{o}(t)}{\mathrm{d}t}+K\varphi_\mathrm{o}(t)=K\tau_2\frac{\mathrm{d}\varphi_\mathrm{i}(t)}{\mathrm{d}t}+K\varphi_\mathrm{i}(t) \tag{1.177}$$

Observing that the first equation is a first-order differential equation (DE) and the second is a second-order DE, we define the loop without filter as a first-order PLL, whereas the loop with one of the LPFs presented above is called a second-order PLL.

It is of great interest to determine the dynamic behaviour of the PLL under various kinds of excitation such as an abrupt change in phase or frequency of the input signal. The behaviour of the system to such changes finds application to phase and frequency demodulation and carrier recovery techniques, as we shall see in Section 1.7.8. By using standard Laplace transform techniques, the resulting time domain function can be obtained for each kind of excitation.

As an example, let us assume that we have a unit step change at the phase of the input signal $y_\mathrm{i}(t)$ which in time and Laplace domains is given by:

$$\varphi_i(t)=\Delta\varphi u(t), \quad \Phi_i(s)=\frac{\Delta\varphi}{s} \tag{1.178}$$

where $u(t)$ represents the unit step function.

Using the expression for the error function $H_e(s)$ from Equation (1.170) we can obtain expressions for $\Phi_e(s)$ and subsequently for $\varphi_e(t)$. For example, for the integrator with phase-lead filter we get:

$$\Phi_e(s)=\frac{\tau_1 s\Delta\varphi}{\tau_1 s^2+K\tau_2 s+K}=\frac{s\Delta\varphi}{s^2+2\zeta\omega_n s+\omega_n^2} \tag{1.179}$$

By splitting the denominator into a product of the form $(s+\alpha)\,(s+\beta)$ and taking the inverse Laplace transform, we obtain the following expression for $\varphi_e(t)$:

$$\varphi_e(t)=\begin{cases}\Delta\varphi\exp\left[-\zeta\omega_n t\right]\left\{\cosh\left[\omega_n\sqrt{\zeta^2-1}\,t\right]-\dfrac{\zeta}{\sqrt{\zeta^2-1}}\sinh\left[\omega_n\sqrt{\zeta^2-1}\,t\right]\right\}, & \zeta>1\\ \Delta\varphi\exp\left[-\omega_n t\right](1-\omega_n t), & \zeta=1\\ \Delta\varphi\exp\left[-\zeta\omega_n t\right]\left\{\cos\left[\omega_n\sqrt{1-\zeta^2}\,t\right]-\dfrac{\zeta}{\sqrt{1-\zeta^2}}\sin\left[\omega_n\sqrt{1-\zeta^2}t\right]\right\}, & \zeta<1\end{cases} \tag{1.180}$$

Figure 1.30 shows the normalized response of the phase error $\varphi_e/\Delta\varphi$ as a function of normalized (with respect to natural frequency ω_n) time. These equations and corresponding

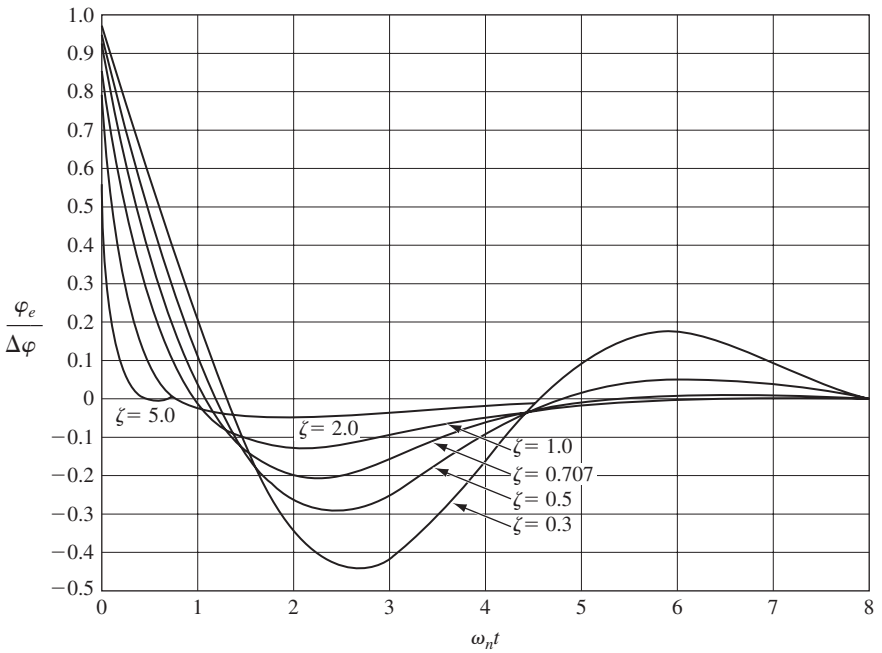

Figure 1.30 Resulting phase error as a response to a phase step $\Delta\varphi$ (from 'Phaselock Techniques', 2nd edn, F. Gardner, copyright © 1979 by John Wiley & Sons)

diagrams are very useful to determine the time needed for the system to settle to the new phase or frequency.

Sometimes it is of interest to know only the steady-state value of the phase or frequency in order to realize whether the system will finally lock or not. In that case, it is only necessary to use the final value theorem of Laplace transform:

$$\lim_{t\to\infty}[\varphi_e(t)] = \lim_{s\to 0}[s\Phi_e(s)] \tag{1.181}$$

For example, if a step change $\Delta\omega \cdot u(t)$ of the input frequency is applied at the input of the PLL, the steady-state error for a second-order loop is given by:

$$\lim_{t\to\infty}[\varphi_e(t)] = \lim_{s\to 0}\left[s\frac{\Delta\omega/s^2}{1+\frac{K}{s(1+s\tau_1)}}\right] = \frac{\Delta\omega}{K} \tag{1.182}$$

At this point it is helpful to define lock range, hold range and lock time.

Lock range is defined as the frequency range within which the PLL will lock, while initially unlocked, within one cycle. Hold range represents the frequency range within which the PLL will remain locked while it is already in lock.

Straightforward analysis shows that the hold range $\Delta\omega_{\mathrm{HR}}$ is the maximum value that $\Delta\omega$ takes in the following equation [Gardner05]:

$$\lim_{t\to\infty}[\sin\varphi_e(t)] = \frac{\Delta\omega}{K_{\mathrm{PD}}K_{\mathrm{VCO}}F(0)} \tag{1.183}$$

Since the maximum value of $\sin \varphi_e(t)$ is equal to 1 (for $\varphi_e = \pi/2$), the hold range is given by

$$\Delta\omega_{\mathrm{HR}} = K_{\mathrm{PD}} K_{\mathrm{VCO}} F(0) \tag{1.184}$$

The typical lowpass with phase lead filter with low frequency gain $F(0) = K_F$ will have a hold range of $K_{\mathrm{PD}} K_{\mathrm{VCO}} K_F$ whereas the perfect integrator filter I-PL theoretically exhibits infinite hold range. However, its hold range is limited by the tuning frequency range of the VCO.

On the other hand, in order to determine the lock range, we initially assume that the PLL is unlocked and the frequency applied at its input is $\omega_{\mathrm{i}S} = \omega_{\mathrm{o}} + \Delta\omega$. Assuming that the phase detector is a simple multiplier and that the high frequency term is filtered by the LPF, the signal at the output of the LPF (input of the VCO) is:

$$U_F = U_{\mathrm{VCO}} \approx K_{\mathrm{PD}} |F(j\Delta\omega)| \sin(\Delta\omega t + \varphi_{\mathrm{i}}(t)) \tag{1.185}$$

This represents a slow sinusoid modulating the VCO. Thus, the VCO output frequency from Equation (1.162) is given by:

$$\omega_{\mathrm{i}} = \omega_{\mathrm{o}} + K_{\mathrm{VCO}} K_{\mathrm{PD}} |F(j\Delta\omega)| \sin(\Delta\omega t) \tag{1.186}$$

The above equation indicates that the VCO output frequency increases and decreases in a sinusoidal fashion with upper and lower limits:

$$\omega_{\mathrm{i}} = \omega_{\mathrm{o}} \pm K_{\mathrm{VCO}} K_{\mathrm{PD}} |F(j\Delta\omega)| \tag{1.187}$$

In order for locking to take place within one period, the upper limit $\omega_{\mathrm{i\,max}}$ of the VCO frequency ω_{i}, should exceed the input frequency $\omega_{\mathrm{i}S}$ and consequently:

$$K_{\mathrm{VCO}} K_{\mathrm{PD}} |F(j\Delta\omega)| \geq \Delta\omega \tag{1.188a}$$

The upper limit of $\Delta\omega$ in Equation (1.188a) gives the lock range $\Delta\omega_{\mathrm{L}}$:

$$K_{\mathrm{VCO}} K_{\mathrm{PD}} |F(j\Delta\omega_{\mathrm{L}})| = \Delta\omega_{\mathrm{L}} \tag{1.188b}$$

To solve this nonlinear equation, approximate expressions for $F(j\Delta\omega_{\mathrm{L}})$ must be used. These give approximately the same value for all types of filters [Best03]:

$$\Delta\omega_{\mathrm{L}} \approx 2\zeta\omega_n \tag{1.189}$$

To find the lock-in time the transient response of the second-order PLL must be determined. It is shown that the transient response is confined to the steady-state value within a small percentage of it, in approximately one signal period. Hence, the lock in time can be expressed as:

$$T_{\mathrm{L}} = \left(\omega_n / 2\pi\right)^{-1} \tag{1.190}$$

1.7.5 Stability of Phase-locked Loops

The requirement for stability is that the phase of the open-loop transfer function $G(j\omega)$ at the gain crossover frequency ω_{CO} (at which $|G(j\omega)| = 1$) is higher than $-180°$:

$$\angle G(j\omega_{\mathrm{CO}}) \geq -180° \tag{1.191}$$

Taking into account that the open-loop gain is expressed as:

$$G(s) = \frac{K_{\mathrm{PD}} K_{\mathrm{VCO}} F(s)}{s} \tag{1.192}$$

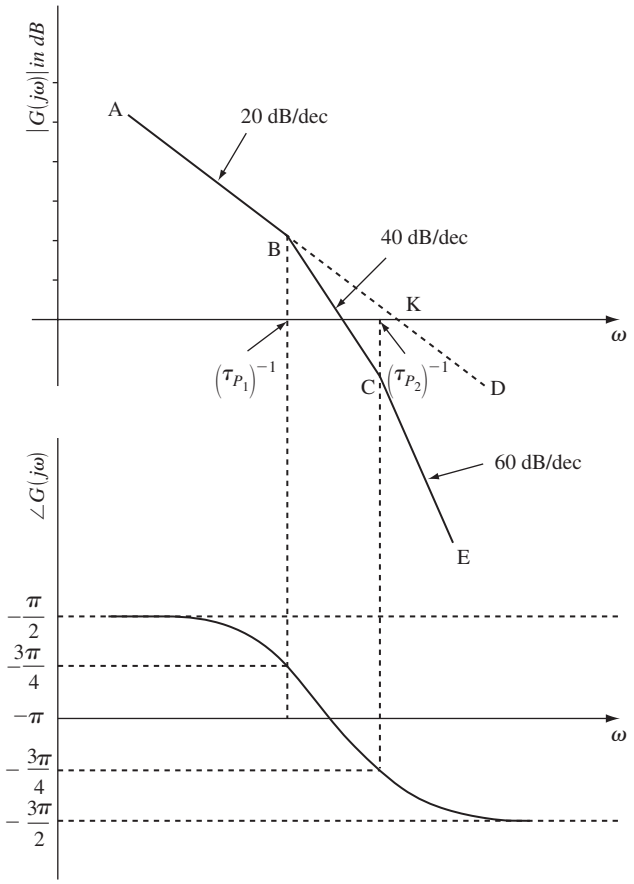

Figure 1.31 Magnitude and phase Bode plots of G(s) for loop with no filter (ideal and lowpass behaviour of VCO and PD)

we distinguish the following cases:

(1) *No filter.* In that case, the Bode diagrams of the magnitude and phase for $G(s)$ are shown in Figure 1.31, indicated by the ABD straight line and a constant $-\pi/2$ phase. It is obvious that in this case the system is stable. However, the filter is necessary after the PD for proper operation. Even if such a filter is absent, the phase detector and VCO exhibit lowpass behaviour at the respective outputs. This behaviour can be modelled as an LPF at the output of the respective components with poles at $1/\tau_{P1}$ and $1/\tau_{P2}$, where τ_{P1} and τ_{P2} represent the time constants of the PD and VCO respectively. Hence, the final open-loop transfer function will become:

$$G(j\omega) = \frac{K_{\mathrm{PD}} K_{\mathrm{VCO}} K_F}{j\omega(1 + j\omega\tau_{P1})(1 + j\omega\tau_{P2})} \tag{1.193}$$

Figure 1.31 shows hypothetical but possible Bode plots for such a system. It is easy to realize that depending on the values of τ_{P1}, τ_{P2} and K, the system can become unstable.

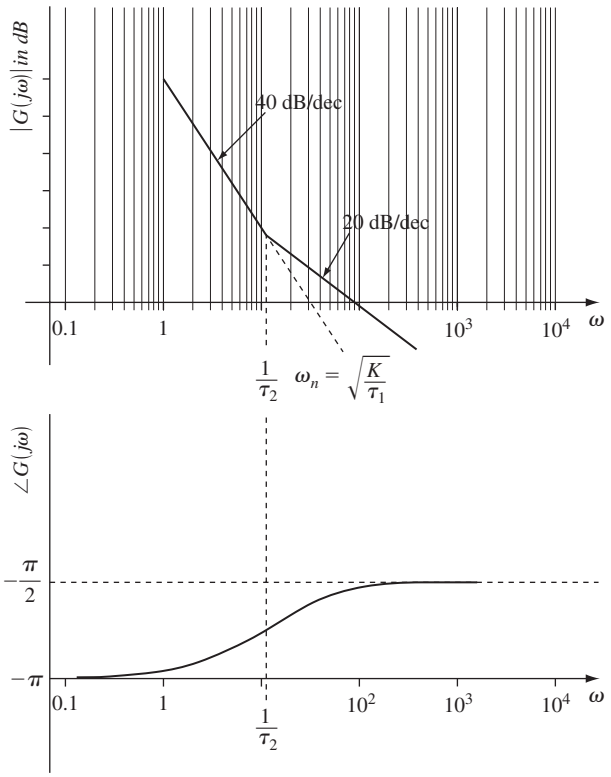

Figure 1.32 Bode diagram with perfect integrator with phase lead filter ($\omega_n > 1/\tau_2$)

(2) *Perfect integrator with phase lead filter.* Figure 1.32 shows the Bode diagrams for the case that $\omega_n > 1/\tau_2$, for which stability can be very satisfactory. The condition for stable operation is: $\omega_n \tau_2 > 2\zeta > 1$. Hence, high values of ζ guarantee good stability.

(3) *Filter with LPF and phase-lead correction.* This filter has an open-loop transfer function given by:

$$G(j\omega) = \frac{K(1 + j\omega\tau_2)}{j\omega(1 + j\omega\tau_1)} \tag{1.194}$$

Because of its two time constants, this filter permits to independently choose values for τ_1 (and subsequently ω_n) and τ_2 and ζ, which is a function of τ_2. Hence, no matter how low the value for ω_n is chosen, increasing τ_2 provides better stability [Blanchard78].

1.7.6 Phase Detectors

The phase detector is one of the principle elements of the PLL as it serves as the component closing the feedback loop through its second input coming from the VCO. The most popular phase detectors are the analogue multiplier and the digital PDs based on flip flops and logic circuitry. The analogue multiplier performs direct multiplication of the two input signals y_i and y_o resulting in a useful term of the form $\sin[(\omega_i - \omega_o(t)) + \varphi_i - \varphi_o(t)]$ as given previously

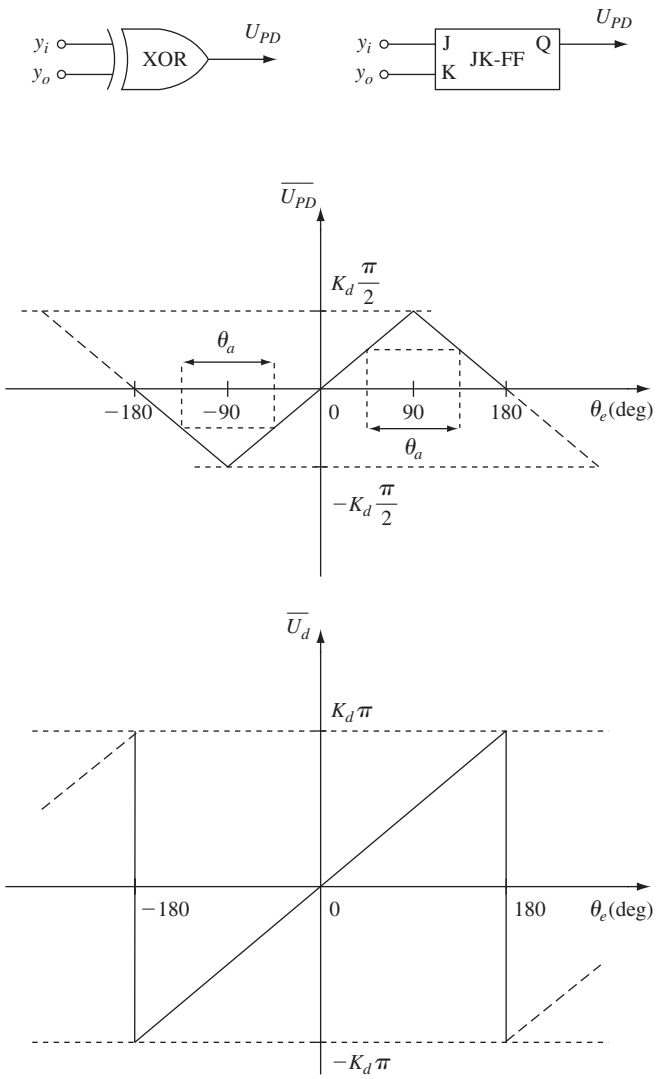

Figure 1.33 Exclusive-OR and JK flip-flop phase detectors with their characteristics

in Equation (1.161). Disregarding the frequency difference, the output signal is a nonlinear function of the phase difference $\sin[\varphi_{\mathrm{i}} - \varphi_{\mathrm{o}}(t)]$, resulting in a major disadvantage. However, when the input signal y_{i} contains considerable noise, this PD exhibits superior performance compared with the digital detectors. On the other hand, owing to the advances of digital technology, digital PDs have dominated most applications excluding those involving very high-speed analogue signals where the analogue detector seems irreplaceable [Gardner05]. The most popular digital PD are the exclusive-OR PD, the Edge-triggered JK-flip flop (ET-JK) detector and the phase-frequency detector (PFD).

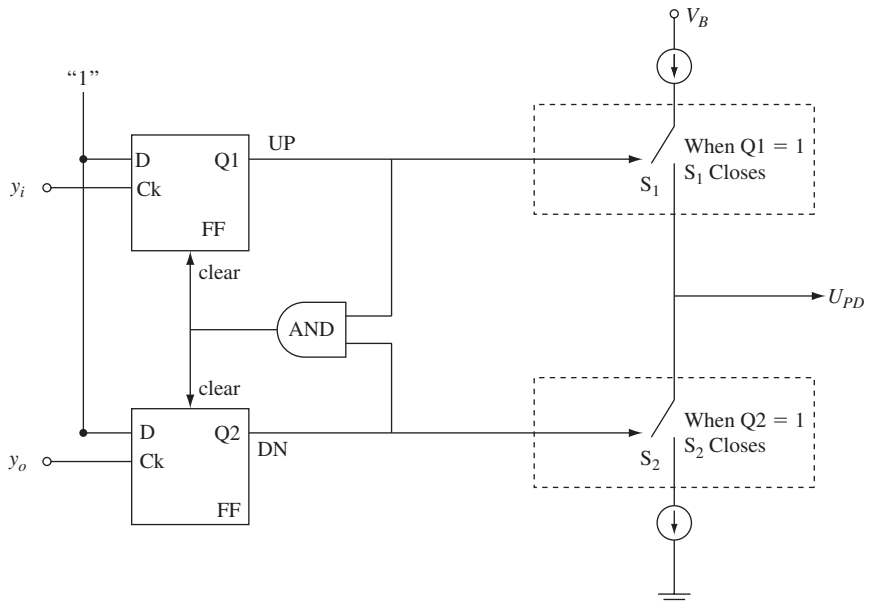

Figure 1.34 The phase–frequency detector

1.7.6.1 The Exclusive-OR and the Edge-triggered JK FF Phase Detectors

Figure 1.33 shows both detectors with inputs and output y_i, y_o and U_{PD} respectively. The exclusive-OR phase detector gives a high-level (logic '1') output when the two input signals have different levels and a logic '0' when they both have high or low levels. Based on that, the exclusive-OR characteristic is also given in Figure 1.33.

However, when one of the signals has a duty cycle different from 50%, the characteristic curve exhibits flattening. This happens because the detector is insensitive to phase changes during time periods so that the high level of one of the inputs is not crossed by the falling or the rising edge of the other.

The edge-triggered-JK flip-flop produces a high level at the output when a rising edge occurs at input y_i whereas it gives a low-level when a rising edge takes place at the second input y_o. Figure 1.33 also shows the JK PD characteristic exhibiting a linear range that is twice as high as that of the exclusive-OR detector. The edge-triggered-JK flip-flop does not have the same problem with that of exclusive-OR detector since the logic levels '1' and '0' at the output result from occurrences of rising edges and not from durations of logic states '1' and '0' at the two inputs.

1.7.6.2 The Phase-frequency Detector

Figure 1.34 depicts the PFD consisting of two D flip-flops with their clock inputs connected to the PD input signals y_i and y_o. Their Q-outputs represent the UP and DN PFD digital outputs the combination of which gives the current state of PFD. Three combinations of logic levels for UP and DN are possible, whereas the fourth (UP = DN = '1') is prohibited through the AND gate. Table 1.4 illustrates all possible states and transitions, which actually can be represented

Table 1.4 Phase-frequency detector states and transitions

Current state	Output signal U_{PD}	Next state for rising edge of y_i	Next state for rising edge of y_o
State = '−1': UP = '0', DN = '1'	$U_{PD} = 0$ Volts	'0'	'−1'
State = '0': UP = '0', DN = '0'	High impedance	'+1'	'−1'
State = '+1': UP = '1', DN = '0'	$U_{PD} = V_B$ Volts	'+1'	'0'

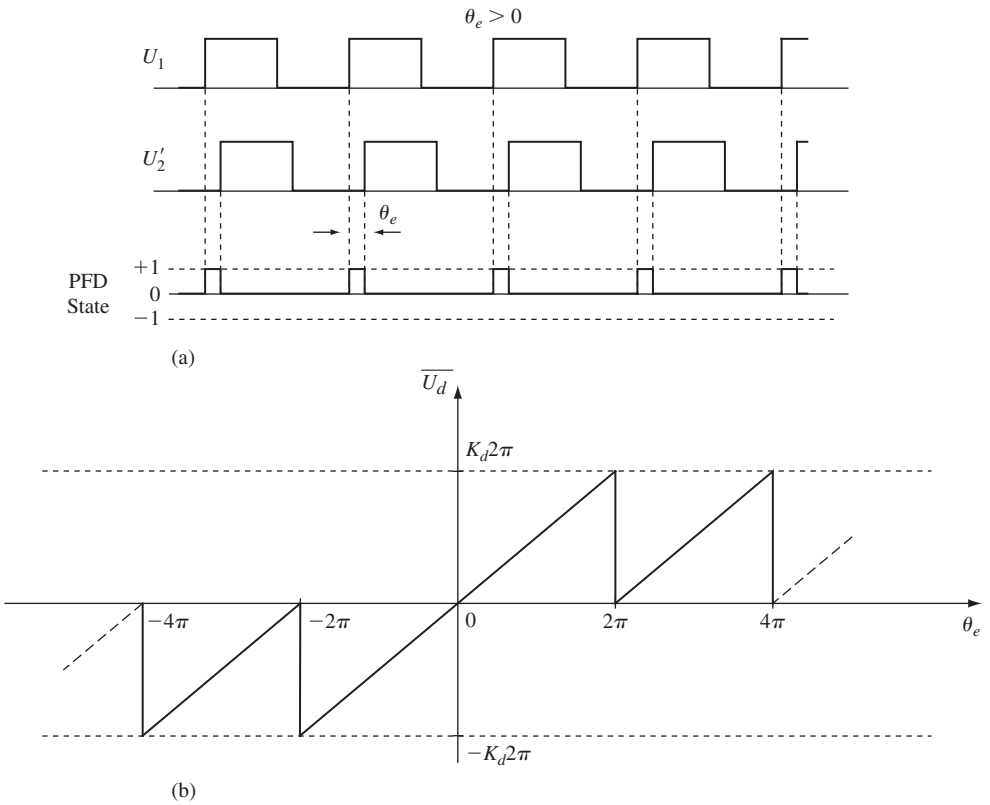

Figure 1.35 (a) Output states of the PFD with the first input leading the second, (b) characteristics of the PFD

graphically by a state diagram [Best03]. It is easy to realize that the three possible states correspond to three states for the output U_{PD} of the PD as shown at the table, taking into account that switches S_1 and S_2 are controlled by UP and DN respectively as shown in Figure 1.34.

Figure 1.35(a) shows the output states of the PFD when y_i leads in phase compared with y_o, whereas Figure 1.35(b) illustrates the characteristic of the PD for the average $\overline{U}_{PD}$. The value of $\overline{U}_{PD}$ depends on the average occupancy in time of a particular state. However, when the frequency of the two signals y_i and y_o is different, the average occupancy of a particular state will be higher than the other states. Consequently, $\overline{U}_{PD}$ changes as a function of frequency making the PFD sensitive in frequency variations, as well.

1.7.7 PLL Performance in the Presence of Noise

Let the input y_i at the phase detector be corrupted by additive noise:

$$y_\mathrm{i}(t) = V_S \sin(\omega_\mathrm{i} t + \varphi_\mathrm{i}) + n(t) \tag{1.195a}$$

The output of the VCO and second input of the PD is given by:

$$y_\mathrm{o}(t) = V_\mathrm{o} \cos(\omega_\mathrm{i} t + \varphi_\mathrm{o}) \tag{1.195b}$$

where θ_o can be initially treated as time invariant for reasons of convenience. It can be shown [Gardner79] that the output signal of the PD can be expressed as:

$$U_\mathrm{PD} = K[\sin(\varphi_\mathrm{i} - \varphi_\mathrm{o})] + n'(t) \tag{1.196}$$

where K is a constant representing the product $V_\mathrm{S} V_\mathrm{O} K_\mathrm{PD}$. The variance $\sigma^2_{n'}$ of the equivalent noise $n'(t)$ as a function of input SNR, $\mathrm{SNR_i}$, is given by:

$$\sigma^2_{n'} = \frac{\sigma^2_n}{V^2_S} = \frac{P_n}{2P_S} = \frac{1}{2SNR_i} \tag{1.197}$$

where P_S and P_n are the signal and noise power at the input of the PLL, respectively. The noise at the output of the PD $n'(t)$ could also be created by a phase disturbance $\sin\varphi_\mathrm{ni}(t) = n'(t)$. When the signal-to-noise ratio at the input is high, the disturbance can be linearized due to small $\varphi_\mathrm{ni}(t)$ [Gardner79]. In this case, it can be considered that the noise is created by an input phase variance $\overline{\varphi^2_\mathrm{ni}}$ which represents the jitter of the input signal due to phase noise.

Calculation of the spectral density of the noise $n'(t)$ gives [Gardner79]:

$$\Phi_{n'}(f) = \frac{2N_0}{V^2_S} \tag{1.198}$$

Consequently, since the PLL transfer function is $H(j\omega)$ the spectral density of the VCO output phase noise $\Phi_\mathrm{no}(f)$ and the corresponding variance are given by:

$$\Phi_\mathrm{no}(f) = \Phi_{n'}(f)|H(j\omega)|^2 \tag{1.199}$$

$$\overline{\varphi^2_\mathrm{no}} = \int_0^\infty \Phi_{n'}(f)|H(j2\pi f)|^2 \mathrm{d}f \tag{1.200}$$

When the noise density of $n(t)$ at the input is white in the band of interest (corresponding to the bandwidth of a bandpass filter in front of the PLL), then $\Phi_n(f) = N_0$ and the VCO output phase noise is:

$$\overline{\varphi^2_\mathrm{no}} = \frac{2N_0}{V^2_S} \int_0^\infty |H(j2\pi f)|^2 \mathrm{d}f \tag{1.201}$$

As noted in a previous section, the above integral represents the equivalent noise bandwidth:

$$B_\mathrm{L} = \int_0^\infty |H(j2\pi f)|^2 \mathrm{d}f \tag{1.202}$$

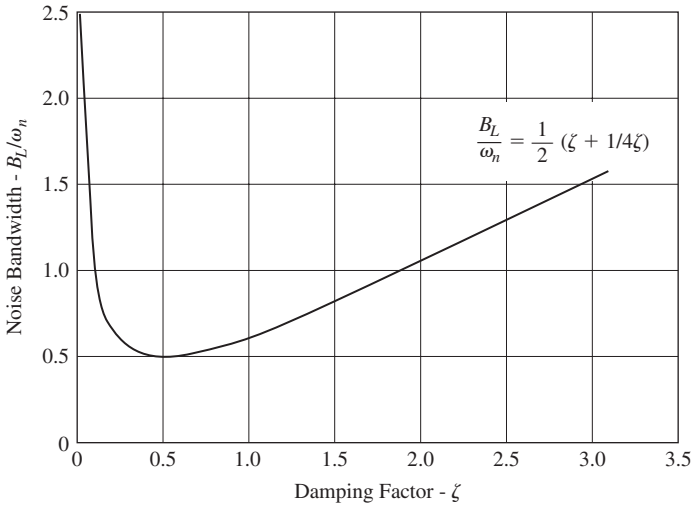

Figure 1.36 Normalized noise bandwidth of a second order PLL (from 'Phaselock Techniques', 2nd edn, F. Gardner, copyright © 1979 by John Wiley & Sons, Inc.)

$\overline{\varphi_{\text{no}}^2}$ represents the phase jitter at the output of the VCO and, like the noise at the input, it can be associated with the signal-to-noise ratio of the loop SNR_L:

$$\overline{\varphi_{\text{no}}^2} = \frac{1}{2\text{SNR}_\text{L}} \tag{1.203}$$

Consequently, the PLL improves the SNR of the signal at its input as follows:

$$\text{SNR}_\text{L} = \text{SNR}_\text{i} \cdot \frac{B_\text{i}}{2B_\text{L}} \tag{1.204}$$

where B_i is the bandwidth of a bandpass filter at the input of the PLL.

Figure 1.36 illustrates the noise bandwidth of a second-order PLL, normalized to the natural angular frequency ω_n. One can realize that there is an optimum value for ζ close to 0.5 resulting in the minimum noise bandwidth B_L.

1.7.8 Applications of Phase-locked Loops

Phase-locked loops have a wide variety of applications in radio communications ranging from frequency synthesizers to modulators and demodulators of analogue and digital signals. Furthermore, they constitute important building blocks in carrier recovery and synchronization in coherent receivers.

The principle of frequency synthesizers is based on deriving an output frequency *fvco*, which in the most general case could be the linear combination of a number of reference frequencies. In the most usual case, *fvco* is an integer multiple of an input frequency f_{in} used as reference frequency. This is widely known as the integer-*N* frequency synthesizer. The synthesizer should be able to generate a wide range of frequencies necessary for down-conversion in the receiver or up-conversion in the transmitter. The number of frequencies and frequency

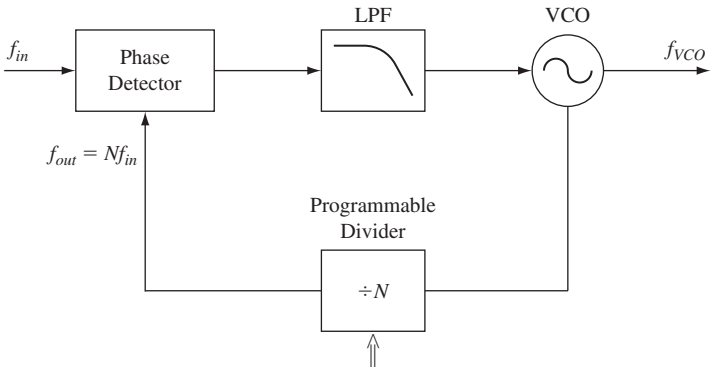

Figure 1.37 Typical integer-N frequency synthesizer

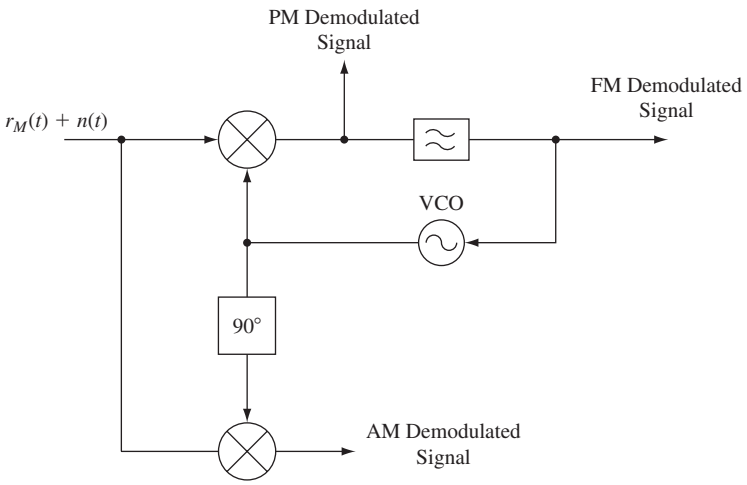

Figure 1.38 FM demodulator using PLL

resolution (minimum frequency step) of the synthesizer are dictated by the application. Phase noise generated by the synthesizer is an important issue as it has a considerable impact on the receiver performance (this will be examined later in Chapter 4). It can be shown that in a simple integer-N configuration the phase noise of the output signal f_o of the VCO follows the phase noise of the reference frequency (which is usually low) within the loop bandwidth. This rule aids in compromising frequency resolution for phase noise. Figure 1.37 illustrates a typical integer-N synthesizer. One can notice that it is a classical PLL with a programmable divider $(\cdot/N)$ inserted at the feedback loop at the output of the VCO.

When the input at the phase detector is an FM/FSK modulated signal (including noise) $r_M + n(t)$, the output of the LPF (and input of the VCO) produces the noisy modulated signal (Figure 1.38). This is because the VCO of the locked system follows the frequency variation of the input signal r_M. In order for the VCO to follow the frequency variation, its input is a time-varying voltage corresponding to the information signal. Improved demodulator performance is achieved by designing the loop to have an increased output SNR.

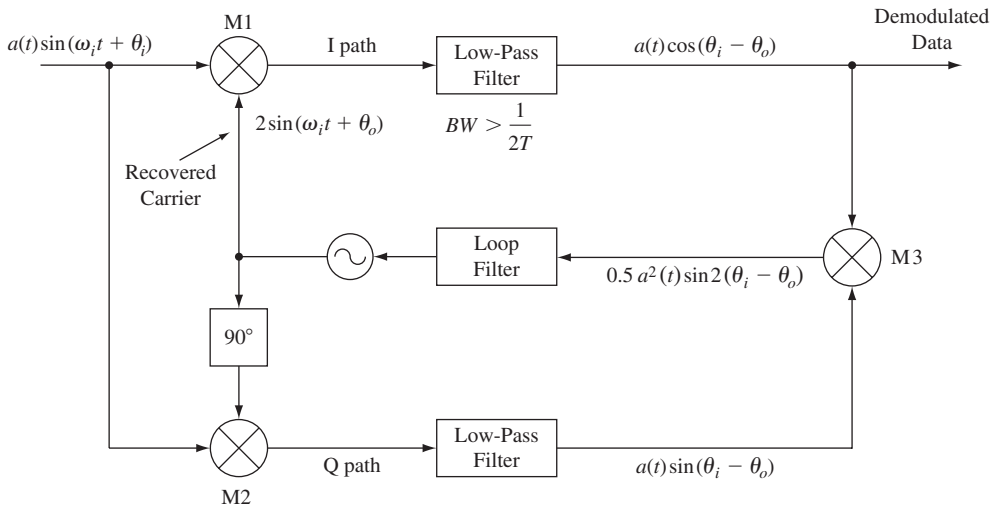

Figure 1.39 Costas loop for carrier recovery and demodulation

Because PLLs can implement maximum likelihood (ML) phase or frequency estimators of feedback nature [Proakis02], they find wide use in carrier and data synchronizers. For example, the Costas loop illustrated in Figure 1.39 is widely known and applicable in BPSK carrier recovery and demodulation. Indeed, by carefully looking at it, if a BPSK modulated signal is used at the input, we see that the output $0.5a^2(t)\sin 2(\theta_\mathrm{i} - \theta_\mathrm{o})$ of the multiplier M3 due to the factor of 2 within the phase argument has eliminated phase modulation and consequently the output of the VCO produces the recovered carrier in phase and frequency. In addition, by designing the I-path LPF as a matched filter with bandwidth equal to the information bandwidth, the output of the I-arm constitutes the BPSK data demodulator [Gardner05]. Other forms of the Costas loop can be applied for carrier recovery in M-PSK modulation. The Costas loop can be designed as a digital system for carrier and clock recovery [Mengali97] in modern communication systems. Commercial digital implementations [HSP50210] find wide use in digital modem design and development.

References

[Aiello03]: R. Aiello, 'Challenges for ultra-wideband (UWB) CMOS integration', IEEE Radio Frequency Integrated Circuits Symp., RFIC'2003, pp. 497–500.

[Best03]: R. Best, 'Phase-Locked Loops: Design, Simulation and Applications', 5th edn, McGraw-Hill, New York, 2003.

[Blanchard78]: A. Blanchard, 'Phase-locked Loops', John Wiley & Sons Inc., New York, 1978.

[Chugh05]: M Chugh, D. Bhatia, P. Balsara, 'Design and implementation of configuable W-CDMA rake receiver architectures on FPGA', IEEE Int. Parallel and Distributed Processing Symposium (IPDPS'05).

[Copani05]: T. Copani, S. Smerzi, G. Girlando, G. Palmisano, 'A 12-GHz silicon bipolar dual-conversion receiver for digital satellite applications', IEEE J. Solid St. Circuits, vol. 40, June 2005, pp. 1278–1287.

[Euro-COST 231-1991]: 'Urban Transmission Loss models for Mobile Radio in the 900 and 1800 MHz bands', September 1991.

[Fleury00]: B. Fleury, 'First- and second-order characterization of dispersion and space selectivity in the radio channel', IEEE Trans. Inform. Theory, vol. 46, September 2000, pp. 2027–2044.

[Gardner79]: F. Gardner, 'Phaselock Techniques', 2nd edn, John Wiley & Sons Inc., New York, 1979.
[Gardner05]: F. Gardner, 'Phaselock Techniques', 3rd edn, John Wiley & Sons Inc., New York, 2005.
[Geer06]: D. Geer, 'UWB standardization effort ends in controversy', Computer, July 2006, pp. 13–16.
[Goldsmith05]: A. Goldsmith, 'Wireless Communications', Cambridge University Press, Cambridge, 2005.
[Gray01]: P. Gray, P. Hurst, S. Lewis, R. Meyer, 'Analysis and Design of Analog Integrated Circuits', John Wiley & Sons Inc., New York, 2001.
[Hata80]: M. Hata, 'Empirical formula for propagation loss in land mobile radio services', IEEE Trans. Vehicular Technol., vol. VT-29, no. 3, August 1980, pp. 317–325.
[Hawwar06]: Y. Hawwar, E. Farag, S. Vanakayala, R. Pauls, X. Yang, S. Subramanian, P. Sadhanala, L. Yang, B. Wang, Z. Li, H. Chen, Z. Lu, D. Clark, T. Fosket, P. Mallela, M. Shelton, D. Laurens, T. Salaun, L. Gougeon, N. Aubourg, H. Morvan, N Le Henaff, G. Prat, F. Charles, C. Creach, Y. Calvez, P. Butel, '3G UMTS wireless system physical layer: baseband proccessing hardware implementation perspective', IEEE Commun. Mag., September 2006, pp. 52–58.
[HSP50210]: 'Digital Costas Loop', Data sheet, 2 July 2008, Intersil.
[Ibnkahla04]: M. Ibnkahla, Q. M. Rahman, A. I. Sulyman, H. A. Al-Asady, J. Yuan, A. Safwat, 'High-speed satellite mobile communications: technologies and challenges', IEEE Proc., vol. 92, February 2004, pp. 312–339.
[Jo04]: G-D Jo, K-S Kim, J-U Kim, 'Real-time processing of a software defined W-CDMA modem', IEEE Int. Conf. on Vehicular Technology (VTC04), pp. 1959–1962.
[Mengali97]: U. Mengali, A. D'Andrea, 'Synchronization Techniques for Digital Receivers', Plenum Press, London, 1997.
[Meyer94]: R. Meyer, W. Mack, 'A 1-GHz BiCMOS RF Front-End IC', IEEE J. Solid-St. Circuits, vol. 29, March 1994, pp. 350–355.
[Muhammad05]: K. Muhammad, 'Digital RF processing: towards low-cost reconfigurable radios', IEEE Commun. Mag., August 2005, pp. 105–113.
[Okumura68]: T. Okumura, E. Ohmori, K. Fukuda, 'Field strength and its variability in VHF and UHF land mobile service,' Rev. Elect. Commun. Lab., vol. 16, nos 9–10, September–October 1968, pp. 825–873.
[Parssinen01]: A. Parssinen, 'Direct Conversion Receivers in Wide-Band Systems', Kluwer Academic, Dordrecht, 2001.
[Pra98]: R. Prasad, 'Universal Wireless Personal Communications', Artech House, Norwood, MA, 1998.
[Proakis02]: J. Proakis, M. Salehi, 'Communication Systems Engineering', 2nd edn, Prentice Hall, Englewood Cliffs, NJ, 2002.
[Rappaport02] T. Rappaport, 'Wireless Communications: Principles and Practice', 2nd edn, Prentice Hall, Englewood Cliffs, NJ, 2001.
[Razavi97]: B. Razavi, 'Design considerations for direct conversion receivers', IEEE Trans. Circuits Systems II, vol. 44, June 1997, pp. 428–435.
[Razavi98]: B. Razavi, 'RF Microelectronics', Prentice Hall, Englewood Cliffs, NJ, 1998.
[Scholten03]: A. Scholten, L. Tiemeijer, R. van Langevelde, R. Havens, A. Zegers-van Duijnhoven, V. Venezia, 'Noise modeling for RF CMOS circuit simulation', IEEE Trans. Electron Devices, vol. 50, March 2003, pp. 618–632.
[Schreier02]: R. Schreier, J. Lloyd, L. Singer, D. Paterson, M. Timko, M. Hensley, G. Patterson, K. Behel, J. Zhou, 'A 10-300-MHz F-digitizing IC with 90–105-dB dynamic range and 15–333 kHz bandwidth', IEEE J. Solid St. Circuits, vol. 37, December 2002, pp. 1636–1644.
[Shaeffer97]: D. Shaeffer, T. Lee, 'A 1.5-V, 1.5-GHz CMOS low noise amplifier', IEEE J. Solid St. Circuits, vol. 32, May 1997, pp. 745–759.
[Tolson99]: N. Tolson, 'A novel receiver for GSM TDMA radio', IEEE Int. Proc. Vehicular Technlogy, 1999, pp. 1207–1211.
[Van Nee99]: R. van Nee, G. Awater, M. Morikura, H. Takanashi, M. Webster, K. Halford, 'New high-rate wireless LAN standards', IEEE Commun. Mag., December 1999, pp. 82–88.
[Wheeler07]: A. Wheeler, 'Commercial applications of wireless sensor networks using ZigBee', IEEE Commun. Mag., April 2007, pp. 71–77.
[Wu00]: Y.-C. Wu, T.-S. Ng, 'New implementation of a GMSK demodulator in linear-software radio receiver', IEEE Personal, Indoor Radio Commun. Conf. (PIMRC'00), pp. 1049–1053.
[Zervas01]: N. Zervas, M. Perakis, D. Soudris, E. Metaxakis, A. Tzimas, G. Kalivas, K. Goutis, 'low-power design of direct conversion baseband DECT receiver', IEEE Trans. Circuits Systems II, vol. 48, December 2001.
[Zhang05]: X. Zhang, G. Riley, 'Energy-aware on-demand scatternet formation and routing for Bluetooth-based wireless sensor networks', IEEE Commun. Mag., July 2005, pp. 126–133.