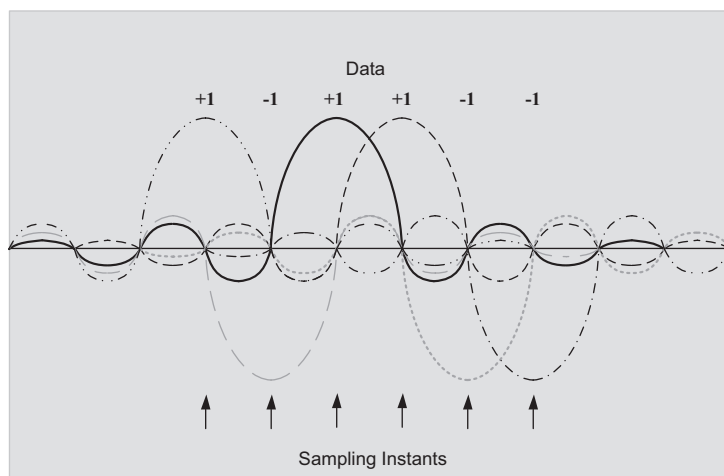


Chapter 2

Bandwidth-Efficient Modulation with Frequency Division Multiple Access (FDMA)



2.1 Introduction

In a wide variety of communication systems, modulation as a fundamental technique plays a very important role in data transmission through air within a specified spectral bandwidth. A modulation signal, which is usually represented by a low-frequency baseband signal and is commonly referred to as an information-bearing signal, varies or modulates one of three parameters: *amplitude*, *phase*, and *frequency* of the radio frequency (RF) carrier signal such that the baseband signal is carried by a varied parameter of the carrier signal through atmosphere propagation to the destination. Why does the information-bearing signal need to modulate a high-frequency carrier signal for this transmission? This is necessary because the

size of the antenna used to radiate the signal to free space depends on the wavelength λ of the transmitted signal. The wavelength λ is equal to c/f , where c is the speed of light and equals 3×10^8 m/s, and f is the frequency of the transmitted signal. For cellular communication systems, antennas are typically $\lambda/4$ in size [1]. If a baseband signal with a frequency of 15 kHz were to be transmitted through an antenna without modulating a carrier signal, the size of the antenna would be $\lambda/4 = 5,000$ m. However, it is only 8 cm if a carrier signal with a frequency of 900 MHz is modulated by such a baseband signal. For this reason, a high-frequency signal usually called a carrier signal is needed for all wireless communication systems to carry the modulation baseband signal. Thus, modulation is a necessary process in all wireless communication systems. Through this book, we mainly consider either the phase modulation scheme or a combination of an amplitude and phase modulation scheme such as M -QAM for its simple implementation and robust performance. In the phase modulation scheme, the information-bearing baseband signal is used to change the phase of a sinusoidal carrier signal whenever the polarity of the baseband signals changes. The phase change of the carrier signal, indicating either 1 or 0, is carried by the carrier signal.

After the carrier signal is modulated by the baseband signal, the RF-modulated signal needs to access the desired frequency channel for the transmission so that the receiver can reliably detect the received signal and extract the original information bits from it. Several different techniques allow access to the RF channel:

- Frequency division multiple access (FDMA)
- Time division multiple access (TDMA)
- Code division multiple access (CDMA)
- Time and frequency division multiple access (TDMA/FDMA)

FDMA is the earliest multiple-access technique mentioned above and is widely used in both satellite/earth station communications and first-generation cellular systems. In this technique a user is assigned a pair of frequencies for sending or receiving a call. One frequency is used for downlink (base station to mobile in cellular systems or satellite to earth station in satellite systems) and one for uplink (mobile to base station or earth station to satellite). This process is called “frequency division multiplexing.” Even though the user may not be talking, a pair of frequencies cannot be reassigned as long as a call is in place. Second-generation cellular systems, such as the global system for mobile communications (GSM), use an FDMA/TDMA technique for voice and data transmissions. The available spectrum band is divided into frequency slots, each with a 200-kHz sub-band. In addition, each frequency slot is also divided into time slots. Each user is assigned a pair of frequencies for uplink and downlink and a time slot during a frame. Therefore, the FDMA technique still plays an important role in modern digital communication systems.

In the early stages (during the late 1980s) satellite digital communication systems adopted an SCPC/FDMA (single channel per carrier) technique, where a total of 800 data/voice channels occupied a 36-MHz transponder bandwidth of the

satellite. In this system each user can transmit and receive either data at a rate of 64 kbps with a Quadrature Phase Shift Keying (QPSK) format or voice at a rate of 32 kbps with a Binary Phase Shift Keying (BPSK) format. The channel spacing is $36 \text{ MHz}/800 = 45 \text{ kHz}$. In the 2G GSM systems with a Gaussian Filtered Minimum Shift Keying (GMSK) modulation format, every symbol represents one bit, which means that symbol rate and bit rate are equal. In the 2.5G GSM Evolution Enhanced Data for Global Evolution (EDGE) systems with an 8PSK modulation format, every symbol represents three consecutive bits, which indicates that bit rate is three times the symbol rate. With the same symbol rate of 270.833 kbps for GMSK modulation, a bit rate up to 812.5 kbps for the EDGE system with an 8PSK modulation can be achieved within the same transmission bandwidth (200 kHz) as GSM systems.

FDMA systems have some common features: the relative low data transmission rate, narrow channel spacing, and restrictive transmission bandwidth. These features dictate that spectrally efficient modulation schemes are the best approach to achieving bandwidth-efficient transmission without significantly causing interference in adjacent channels. Meanwhile, the modulated signals in these applications prefer to have constant envelope or small envelope fluctuation in order to achieve energy efficiency when the power amplifiers operate in the saturation region or close to it. The energy-efficient operation of the power amplifier can extend the usage time of the direct current battery.

2.2 Definition of Energy and Spectral Efficiency

Energy- and spectrum-efficient modulation and transmission techniques are the first two high-priority requirements among three categories of most wireless communication systems: spectrum efficiency, energy efficiency, and cost efficiency. These three categories in the list order above are top priorities for most electrical system designers in designing or choosing a wireless communication system.

Recognizing the fact that a power amplifier (PA) is one of the most critical component and consumes most part of the total energy at the transmitter, we take some approaches to improving the energy efficiency of the PA from the perspective of the *signal design* without significantly causing the degradation of the spectral efficiency. With great demands for high-data-rate applications in the next wireless generations, such as 5G cellular networks and IEEE 802.11ax WLAN, the importance of the energy efficiency has been greatly recognized over the past several years. More and more research and development efforts in industry and academia focus on how to enhance and improve the energy efficiency of future wireless communication networks from perspectives of PA design, signal design and network design.

2.2.1 Bandwidth or Spectrum Efficiency

The bandwidth of a channel is the frequency range over which a modulated signal is transmitted and then reliably detected in the receiver. Since frequency spectrum resource is limited, it has to be utilized efficiently. The term *spectral efficiency* is used to describe the rate of maximum information being transmitted over a given bandwidth in a specific communication system. Hence, spectral efficiency may also be called “bandwidth efficiency.”

In general, spectral efficiency refers to the information rate that can be transmitted over a given bandwidth in a specific communication system to achieve reliable performance. Spectral efficiency is viewed as bits per second per hertz (bits/s/Hz) and defined as

$$\eta_s = \frac{R}{B_w} \text{ (bit/s/Hz)} \quad (2.1)$$

where R is the information rate or transmission rate in bit/s, and B_w is the *passband* (or double-side) transmission bandwidth in Hz. For a certain information rate, the spectral efficiency can be maximized by minimizing the transmission bandwidth. The minimum transmission bandwidth for intersymbol interference (ISI) free is determined by the Nyquist bandwidth (B_N) criterion, which states that the theoretical minimum bandwidth B_N of an ideal and linear phase brick-wall channel lowpass filter used for impulse transmission at a transmission rate of f_s symbols per second without ISI is equal to the Nyquist frequency f_N , or $B_N = f_N = f_s/2$. For a passband transmission system, the minimum *passband* bandwidth of B_w for ISI free is twice the Nyquist bandwidth, or $B_w = 2B_N = 2f_N$. If the transmission bandwidth of B_w is less than twice the Nyquist bandwidth B_N , the responses to these impulses at the output of the channel lowpass filter have ISI at the optimal sampling instants.

If rectangular pulses rather than impulses are used in the transmission channel, an inverse SINC function, or $x/\sin(x)$, shaped amplitude equalizer should be added before the ideal brick-wall channel filter so that the rectangular pulses can be transferred to the impulses before the ideal brick-wall channel filter.

Spectrum efficiency or bandwidth efficiency in (2.1) describes how efficiently the allocated bandwidth is utilized to accommodate the higher data transmission rate. For a given information rate R , the symbol rate f_s is associated with the information rate R or bit rate f_b through a modulation format such as M -order QAM. Therefore, the relationship between f_b and f_s is determined by the modulation format.

For a signaling alphabet with M alternative symbols, each symbol contains $N = \log_2 M$ bits, or N -bit represents the number of $M = 2^N$ symbols. The relationship between the bit rate f_b and symbol rate f_s is expressed as

$$f_b = Nf_s \quad (2.2)$$

or

$$f_s = \frac{f_b}{N} = \frac{f_b}{\log_2 M} \quad (2.3)$$

In the case of M -QAM, M represents M alternative symbols in (2.3). The theoretical spectral efficiency for the passband BPSK transmission, where the bit rate is equal to the symbol rate ($f_b = f_s$), is given by

$$\eta_s = \frac{f_b}{B_w} = \frac{f_s}{2f_N} = \frac{f_s}{2(f_s/2)} = 1 \text{ bit/s/Hz} \quad (2.4)$$

here the transmission bandwidth B_w for the passband signal is twice the Nyquist frequency f_N , or $B_w = 2f_N$. For a passband 16-QAM transmission, where M is equal to 16, the spectral efficiency is calculated by

$$\eta_s = \frac{f_b}{B_w} = \frac{\log_2 16 \times f_s}{2(f_s/2)} = 4 \text{ bit/s/Hz} \quad (2.5)$$

Table 2.1 illustrates the theoretical bandwidth efficiency limits for some main modulation formats. It should be noted that these figures cannot actually be achieved in practical radios since the ideal brick-wall channel filter is required, which is impossible to practically design.

In practice, a raised cosine filter with a roll-off factor of α is widely used to approximate the ideal brick-wall filter. The double-side bandwidth of the raised cosine filter is given by

$$B_w = 2(1 + \alpha)f_N, \quad 0 \leq \alpha \leq 1 \quad (2.6)$$

In (2.6), for $\alpha = 0$, the minimum double-side bandwidth is equal to twice the Nyquist frequency, or $B_w = 2f_N$, whereas for $\alpha = 1$, the maximum double-side bandwidth is four times the Nyquist frequency, or $B_w = 4f_N$. Theoretically, beyond the bandwidth expressed in (2.6) the attenuation has an infinite value. In practice, the raised cosine filter with a finite attenuation value can be realized, depending on

Table 2.1 Theoretically spectral efficiencies of main modulation formats

Modulation scheme	Theoretical bandwidth efficiency (bits/s/Hz)
BPSK	1
QPSK	2
8PSK	3
16-QAM	4
32-QAM	5
64-QAM	6
128-QAM	7
256-QAM	8

the allowed amount of adjacent channel interference. In the case of the channel raised cosine filter, spectral efficiency can be calculated using the bandwidth of the raised cosine filter in (2.6).

It has been shown so far that the minimally occupied bandwidth would be made equal to the symbol rate if a raised cosine filter with $\alpha = 0$ were implemented as the channel filter. From a system-design point of view, the occupied bandwidth is usually larger than the bandwidth of the channel filter given in (2.6) because the guard band should be included in the occupied bandwidth. The width of the guard band depends on the allowed amount of adjacent channel interference required by the specification of the system.

The TDMA version of the North American Digital Cellular (NADC) system adopts $\pi/4$ -DQPSK modulation with a root-raised cosine filter response with a roll-off factor of $\alpha = 0.35$. This system provides a 48.6-kbits/s data rate over a 30-kHz channel bandwidth. In this example, we can calculate spectral efficiency using three different bandwidths:

- Theoretical minimum bandwidth
- Actual filter bandwidth
- System channel bandwidth

Theoretical minimum bandwidth: Spectral efficiency with minimum bandwidth for the passband signal is calculated by setting $\alpha = 0$ in (2.6):

$$\eta_s = \frac{f_b}{B_w} = \frac{f_b}{2f_N} = \frac{f_b}{2(f_s/2)} = \frac{2f_s}{f_s} = 2 \text{ bit/s/Hz} \quad (2.7)$$

Actual filter bandwidth: Spectral efficiency with actual filter bandwidth is calculated by setting $\alpha = 0.35$ in (2.6):

$$\eta_s = \frac{f_b}{B_w} = \frac{f_b}{2(1 + 0.35)f_N} = \frac{f_b}{1.35f_s} = \frac{2f_s}{1.35f_s} = 1.48 \text{ bit/s/Hz} \quad (2.8)$$

System channel bandwidth: Spectral efficiency with system channel bandwidth is obtained by substituting the channel bandwidth of 30 kHz with $B_w = 30 \text{ kHz}$ into (2.4):

$$\eta_s = \frac{f_b}{B_w} = \frac{48.6 \text{ kbit/s}}{30 \text{ kHz}} = 1.62 \text{ bit/s/Hz} \quad (2.9)$$

It can be seen that the spectral efficiency of the system channel bandwidth is 1.62 bit/s/Hz in (2.9), where the channel bandwidth of 30 kHz used in the calculation is greater than the minimum bandwidth of 24.3 kHz in (2.7), but less than the actual filter bandwidth of 32.8 kHz in (2.8). Hence, the bandwidth efficiency with the system channel bandwidth is larger than that with the actual filter bandwidth because the system bandwidth in the former has a certain finite attenuation that meets the system requirement while the actual filter bandwidth in the latter is supposed to have an infinite attenuation calculated in (2.6).

2.2.2 Energy Efficiency

In highly integrated wireless systems, such as wireless system on chip (SoC) devices, the RF power amplifier is the subsystem that consumes most DC power in the whole transmission system. For example, it may consume up to more than 70% of DC power in the transmit path of certain RF mobile transceivers [2]. For this reason, the “energy efficiency” of the RF power amplifier, which is often described as “power efficiency” in the literature (and this is incorrect [3]), is directly proportional to and mainly represents the efficiency of the overall system, especially in the transmission path. In addition to power consumption, it is now apparent that energy consumption is an important metric for transmitter circuits. Energy consumption more accurately predicts the battery life, especially when a portable system operates with a wide range of the output power. Hence, energy efficiency tends to be a better metric of the performance than power efficiency in terms of the battery life. Actually, energy consumption depends on the power consumption and time spent in the power consumption duration.

Usually, in the published literature, the power efficiency of the power amplifier refers to how efficiently the input power to the power amplifier, including the input AC and DC powers, is converted to the output AC power to a load or an antenna, regardless of time. Energy efficiency of the power amplifier, however, is identical to its power efficiency as long as all powers, including the input AC power to the PA, power supply DC power, dissipated power as heat, and the output AC power from the PA are measured equivalently in time [3]. Therefore, we use the term *energy efficiency* instead of *power efficiency* in this book even though definitions are the same [3].

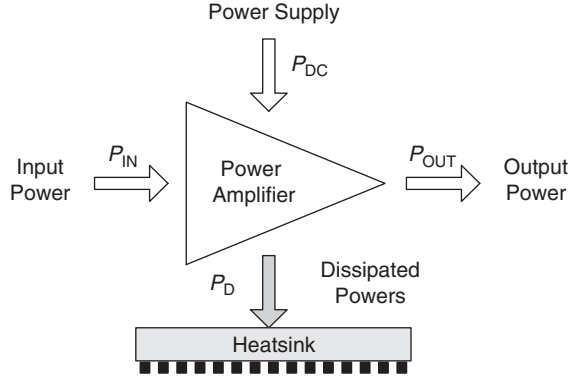
All power amplifiers can be represented by a four-port network: an input DC port, a RF input signal port, a RF output signal port, and a ground. The measured input DC power P_{DC} includes the power associated with all the bias lines of the power amplifier. It is assumed here that the PA is perfectly matched and has infinite reverse isolation. Therefore, the measured power P_{IN} at the input RF port corresponds to the input signal power at the fundamental frequency only. The measured power P_{OUT} at the output RF port corresponds to the output power at the fundamental frequency and all the spurious frequencies, which are generated by the PA itself.

A power amplifier is evaluated for efficiency using a conservation of power calculation based on the flows identified in Fig. 2.1:

$$P_{IN} + P_{DC} = P_{OUT} + P_D \quad (2.10)$$

The energy efficiency of the PA is a measure of its ability to convert the DC power of the power supply into the RF signal power delivered to the load. There are two definitions of power amplifier efficiency. One is basic PA efficiency and the other is power-added efficiency.

Fig. 2.1 Power flows in a power amplifier. Redrawn from [3]



Basic PA Efficiency: The basic PA efficiency is a ratio of the RF output power to the DC power and is derived from (2.10) as

$$\eta_B = \frac{P_{OUT}}{P_{DC}} = 1 - \frac{P_D - P_{IN}}{P_{DC}} \quad (2.11)$$

More precisely, PA efficiency is also called the overall efficiency [4]. If the power amplifier has relatively high power gain, the direct contribution of the RF input signal power to the RF output signal power is insignificant, and therefore it can be neglected. The basic efficiency can be approximated by

$$\eta_B \approx 1 - \frac{P_D}{P_{DC}} \quad (2.12)$$

Power-Added Efficiency: When the gain of the power amplifier is not significantly high, the RF input power needs to be subtracted from the RF output power in the efficiency expression, and then the power efficiency is referred to as the power-added efficiency (PAE):

$$\eta_{PAE} = \frac{P_{OUT} - P_{IN}}{P_{DC}} = 1 - \frac{P_D}{P_{DC}} \quad (2.13)$$

If the PA has a relatively large power gain, then, $\eta_B \approx \eta_{PAE}$, as expressed in (2.12) and (2.13). The power-added efficiency can be interpreted as the efficiency of the network to convert the input DC power into the amount of the RF net output power. PAE definition in (2.13) is widely used as a useful metric for evaluating the efficiency of the RF power amplifier. PAE becomes zero when the power gain of G_{PA} is unit or $P_{OUT} = G_{PA} \times P_{IN} = P_{IN}$. This means the power amplifier does not convert any DC power to the RF output power.

It can be seen from (2.11) and (2.13) that the PA efficiency increases as DC power of P_{DC} is reduced. The DC power P_{DC} is given by

Table 2.2 Classical modes of power amplifier operation

Classical mode	Conduction angle (°)	Operation range
A	360	Linear
AB	180–360	Either linear or nonlinear ^a
B	180	Nonlinear
C	0–180	Nonlinear

^aClass-AB is not a complete linear amplifier; a RF signal with non-constant envelope will be distorted significantly at its peak power level

$$P_{DC} = V_{DC}I_{DC} \quad (2.14)$$

The DC current I_{DC} can decrease monotonically as the conduction angle of the power amplifier is reduced [5]. The conduction angle of the power amplifier determines the classical modes of the power amplifier operation as listed in Table 2.2. In 3G and 4G cellular and 802.11 WLAN communication systems, power amplifiers usually operate in a class AB mode to achieve high efficiency by reducing the DC current I_{DC} and to avoid severe nonlinear distortion on the RF output signal as well. In general, DC supply voltage of V_{DC} is fixed during the power amplifier operation except the envelope tracking technique based power amplifier [6, 7], where the supply voltage dynamically and synchronously tracks or follows the envelope of the RF input signal. Furthermore, even in the class AB mode, the efficiency can be further increased due to the reduction of DC current I_{DC} as the output back-off of the power amplifier from its P1dB compression point is reduced [8]. Therefore, in terms of achieving high efficiency, power amplifiers are preferred to operating in the small back-off condition as long as the transmitted power spectral density (PSD) and error vector magnitude (EVM) meet the standard specifications with enough margins.

The expressions in (2.11) and (2.13) also demonstrate that the consideration of power amplifier efficiency is the same as the consideration of either amplifier output power or amplifier power dissipation. The amount of the total input power to the power amplifier in Fig. 2.1, either DC power or RF input power that is not converted into the RF output power, is dissipated as heat. Higher power output corresponds to higher energy efficiency. From (2.10), lower power dissipation leads to higher power output, which in turn results in higher power amplifier efficiency. From a design point of view, the design objective of maximum energy efficiency is identical to the design objective of either minimum power dissipation P_D or DC power P_{DC} . Therefore, the energy efficiency of the power amplifier has become a challenging requirement for most PA designers. Cell phone handset PAs have to operate efficiently to conserve battery power and base station PAs are also need to be efficient as possible due to cooling limitations [5].

The energy consumption of a system is highly related to the power consumption through a time interval T . There are many different definitions of the energy consumption in the literature. Considering that the energy consumption covered in this book only focus on the transmitter, especially for the PA, we define the

energy consumption of the PA at the transmitter as shown in Fig. 2.1 during the time interval T as

$$E = T(P_{\text{DC}} + P_{\text{D}}) [\text{Joule}] \quad (2.15)$$

where P_{DC} is the DC consumption power of the PA and P_{D} represents the static power dissipated in the PA as heat. Substituting $P_{\text{DC}} = P_{\text{OUT}}/\eta_{\text{B}}$ in (2.11) into (2.15), the energy consumption in (2.15) can be rewritten as:

$$E = T(\mu P_{\text{OUT}} + P_{\text{D}}) [\text{Joule}] \quad (2.16)$$

where $\mu = 1/\eta_{\text{B}}$ is a factor, with η_{B} the efficiency of the transmit power amplifier. In the case that the PA has a relatively large power gain, we have shown the relationship between η_{B} and η_{PAE} , or $\eta = \eta_{\text{B}} \approx \eta_{\text{PAE}}$.

If P_{D} includes the power dissipated in all other circuit blocks of the transmitter and receiver, the energy consumption expression in (2.16) represents the energy consumption of a system, which is the same as one used in [9, 10]. From the PA standpoint of view, expressions of (2.15) and (2.16) precisely represent the energy consumption of the PA.

As the energy consumption of the PA has been defined, the energy efficiency of the PA needs to be defined next. The energy efficiency of the PA is the ratio of the benefit obtained after sustaining the energy cost and the amount of energy consumption in (2.15) and (2.16). The benefit is usually related to the amount of data reliably transmitted in the time interval T . Several performance functions have been used in the literature to evaluate this quantity, such as *system capacity*, and *system throughput* in [9, 10]. Throughput is the amount of data reliably transmitted over a communication channel in the certain time period T . The throughput metric is measured in bits/s and also depends on the signal-to-noise ratio (SNR) and the transmission channel condition. Compared to capacity, throughput is more practically used to evaluate the system performance.

Energy Efficiency: With a general function $f(\text{SNR})$ that represents throughput as the system benefit, the energy efficiency of the PA is defined as

$$\text{EE} = \frac{T \times f(\text{SNR})}{T \times (\mu P_{\text{OUT}} + P_{\text{D}})} = \frac{f(\text{SNR})}{\mu P_{\text{OUT}} + P_{\text{D}}} [\text{bits/Joule}] \quad (2.17)$$

EE can be improved by either increasing the numerator or decreasing the denominator in (2.17). In this book, however, we focus on increasing the energy efficiency by decreasing the denominator through the increase of the efficiency η of the PA. It is clear that EE can be improved by decreasing the factor $\mu = 1/\eta$ through the increase of the efficiency η of the PA, which in turn requests either to increase the output power P_{OUT} of the PA or to decrease the DC power P_{DC} . Even though EE can be improved by increasing the throughput in (2.17), the increase of the throughput takes the entire system involved, including a transmitter and receiver,

multi-antennas, a channel condition and so on. Therefore, energy efficiency improvements through increasing the throughput are very complicated and are beyond the scope of this book.

2.3 Fundamentals of Modulation

The main objective of the modulation process is to shift the spectrum of the information-bearing baseband signal to a high-frequency band suitable for transmission. The band center of the modulated signal is located at the carrier frequency and its bandwidth is twice the bandwidth of the baseband signal.

2.3.1 The Convolution Property

One of the most important properties of the Fourier transform in linear time-invariant (LTI) systems is the convolution operation. With the convolution operation, the relation between the input $x(t)$ and output $y(t)$ of a continuous-time LTI system with impulse response $h(t)$ is given by

$$y(t) = x(t) * h(t) = \int_{-\infty}^{+\infty} x(\tau)h(t - \tau)d\tau \quad (2.18)$$

The relationship between the Fourier transform and inverse Fourier transform for the output signal $y(t)$ is expressed as

$$y(t) = F^{-1}\{Y(\omega)\} = \frac{1}{2\pi} \int_{-\infty}^{+\infty} Y(\omega)e^{j\omega t}d\omega \quad (2.19)$$

$$Y(\omega) = F\{y(t)\} = \int_{-\infty}^{+\infty} y(t)e^{-j\omega t}dt \quad (2.20)$$

Then, (2.20) can be written as

$$\begin{aligned} Y(\omega) &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{+\infty} x(\tau)h(t - \tau)d\tau \right] e^{-j\omega t}dt \\ &= \int_{-\infty}^{+\infty} x(\tau) \left[\int_{-\infty}^{+\infty} h(t - \tau)e^{-j\omega t}dt \right] d\tau \\ &= \int_{-\infty}^{+\infty} x(\tau)e^{-j\omega\tau}H(\omega)d\tau \\ &= H(\omega) \int_{-\infty}^{+\infty} x(\tau)e^{-j\omega\tau}d\tau = H(\omega)X(\omega) \end{aligned} \quad (2.21)$$

In the derivation of (2.21), $H(\omega)$ and $X(\omega)$ are the Fourier transform of the system impulse response $h(t)$ and the Fourier transform of the input signal $x(t)$. Thus, the convolution of two signals in the time domain corresponds to the multiplication of their Fourier transforms in the frequency domain.

2.3.2 Modulation Property

By using duality, a property of the Fourier transform between the time and frequency domains, we can obtain the relationship in the frequency domain from the multiplication of two signals in the time domain. If the multiplication of one signal $x(t)$ by another $s(t)$ in the time domain is expressed as

$$y(t) = x(t) \times s(t) \quad (2.22)$$

Then, Fourier transform expression of (2.22) is given by

$$Y(\omega) = \frac{1}{2\pi} [X(\omega) * S(\omega)] \quad (2.23)$$

Such multiplication of one signal by another in the time domain is also called the *modulation process*, or *modulation*, where the signal with the lower frequency is referred to as the modulating signal while the other one with the higher frequency is called the carrier signal. The property connected by (2.22) and (2.23) is called the modulation property.

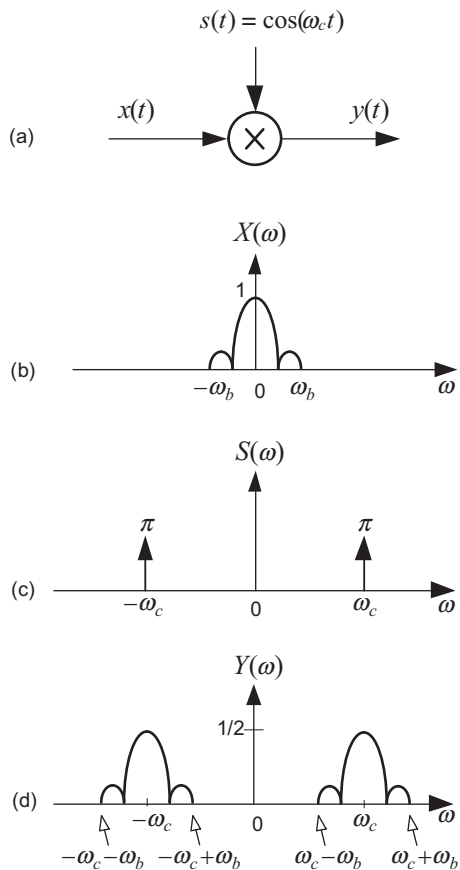
In a practical modulation process, a cosine waveform is usually used as a carrier signal. Let $x(t)$ be a modulation signal whose Fourier transform $X(\omega)$ is limited to the frequency of ω_b and $s(t) = \cos(\omega_c t)$ a carrier signal whose Fourier transform is expressed as $S(\omega) = \pi\delta(\omega - \omega_c) + \pi\delta(\omega + \omega_c)$. Then, from (2.23) we have Fourier transform of the signal $y(t)$ given by

$$Y(\omega) = \frac{1}{2\pi} X(\omega) * S(\omega) = \frac{1}{2} [X(\omega - \omega_c) + X(\omega + \omega_c)] \quad (2.24)$$

It is clear from (2.24) that the spectrum of the baseband signal $x(t)$ is completely shifted to higher frequencies centered at the carrier frequencies of $\pm\omega_c$ after the modulation process illustrated in Fig. 2.2. This frequency-shift process does not cause any distortion if the condition $\omega_c > \omega_b$ is met. In practice, it should be more than twice the frequency of ω_b (or $2\omega_b$). So the original signal $x(t)$ can completely recovered by multiplying the modulated signal $y(t)$ with a referenced carrier signal, which will be discussed later.

The carrier signal generally used in practice is a cosine signal, even though it is not a necessary restriction, such as a square wave signal. In the case of a cosine carrier with the frequency of ω_c , the modulated signal can be expressed as

Fig. 2.2 Spectrum shift after amplitude modulation with a sinusoidal signal:
 (a) amplitude modulation with a sinusoidal signal,
 (b) spectrum of modulating signal $x(t)$, (c) spectrum of sinusoidal signal $s(t) = \cos(\omega_c t)$, and
 (d) spectrum of modulated signal $y(t)$



$$y(t) = A(t) \cos [\omega_c t + \phi(t)] = A(t) \cos [\theta(t)] \quad (2.25)$$

where $\theta(t)$ is the instantaneous phase and $A(t)$ is the instantaneous amplitude. The instantaneous frequency $\omega(t)$ is defined as

$$\omega(t) = \frac{d\theta(t)}{dt} = \omega_c + \frac{d\phi(t)}{dt} \quad (2.26)$$

The functions $\phi(t)$ and $d\phi(t)/dt$ are called the *phase deviation* and *frequency deviation*, respectively.

If the amplitude $A(t)$ is only proportional to the modulating signal $x(t)$, the expression of $y(t)$ in (2.25) is referred to as *amplitude modulation*, which is expressed as

$$y(t) = k_a x(t) \cos [\omega_c t + \phi_c] \quad (2.27)$$

where k_a is the amplitude deviation constant in volt and ϕ_c is the constant phase.

If the phase deviation $\phi(t)$ is proportional to the modulating signal $x(t)$ only, or $\phi(t) = k_p x(t)$, the expression $y(t)$ in (2.25) is referred to as *phase modulation*, which is defined as

$$y(t) = A_c \cos [\omega_c t + k_p x(t)] \quad (2.28)$$

where k_p is the phase deviation constant in radian and A_c is the amplitude constant. Similarly the frequency deviation $d\phi(t)/dt$ is only proportional to the modulating signal $x(t)$, or $d\phi(t)/dt = k_f x(t)$, the expression of (2.25) is referred to as *frequency modulation*, which is defined as

$$y(t) = A_c \cos \left[\omega_c t + k_f \int_{-\infty}^t x(\tau) d\tau \right] \quad (2.29)$$

where k_f is the frequency deviation constant in hertz.

For the amplitude, phase and frequency modulations expressed in (2.27) to (2.29), the instantaneous frequencies are ω_c , $\omega_c + k_p dx(t)/dt$, and $\omega_c + k_f x(t)$, respectively. Since phase modulation and frequency modulation differ only in an integration operation, they are also called *angle modulation*. For detailed descriptions regarding fundamental modulations, the interested reader can reference the Ziemer and Tranter [11].

If not only the phase deviation $\phi(t)$ but also the amplitude $A(t)$ are proportional to the modulating signal $x(t)$, the modulated signal $y(t)$ is referred to as combination of both phase and amplitude modulations, such as a Quadrature Amplitude Modulation (QAM) modulation format. Due to its robust bit error rate (BER) performance and relatively simple architecture, the combination of phase and amplitude modulations shall be mostly used through this book.

2.4 Digital Baseband Modulation

In modern communications, information is usually transmitted in the form of a bit stream in order to achieve high-quality transmission performance. The bit stream, however, must be transferred into continuous-time waveforms that are suitable for transmission over a communications channel. This transformation is called mapping, in which one of finite energy waveforms $\{s_m(t), m = 1, 2, \dots, M\}$ is selected to present one of $M = 2^k$ possible normalized symbols or bits $\{A_m = (2m - 1 - M), m = 1, 2, \dots, M\}$ at a time for transmission over the channel.

2.4.1 2-Level Pulse Amplitude Modulation (2-PAM) and Binary Phase Shift Keying (BPSK)

In the case of $M=2$ symbols, a two-level pulse amplitude modulation (PAM) is called the binary *Pulse Amplitude Modulation* (PAM) waveforms, whose waveforms are chosen as

$$s_m(t) = A_m g(t), \quad m = 1, 2 \quad (2.30)$$

where the symbol amplitudes A_m have the values of $A_m = \pm 1$, $m = 1, 2$ and the waveform $g(t)$ is a shaping pulse. Usually the special waveforms (or $A_1 = 1$ for $m = 1$, $A_2 = -1$ for $m = 2$) are chosen as

$$s_1(t) = -s_2(t), \quad 0 \leq t \leq T_b \quad (2.31)$$

where the bit duration $T_b = 1/f_b$.

In digital PAM, also called *Amplitude Shift Keying* (ASK), with the case of $M=2$, the modulated signal is represented as

$$\begin{aligned} y(t) &= \sum_{n=-\infty}^{\infty} A_m g(t - nT_b) \cos(2\pi f_c t) \\ &= s_m(t) \cos(2\pi f_c t), \quad m = 1, 2 \quad 0 \leq t \leq T_b \end{aligned} \quad (2.32)$$

where f_c is the carrier frequency.

In digital phase modulation, called *Phase Shift Keying* (PSK), with the case of $M=2$, the modulated signal is expressed as

$$y(t) = |s_m(t)| \cos[2\pi f_c t + \pi(m-1)], \quad m = 1, 2 \quad 0 \leq t \leq T_b \quad (2.33)$$

We see from (2.32) and (2.33) that in the case of $M=2$ digital PAM signals are identical to digital PSK. Furthermore, we note that the PAM and PSK modulated signals are dependent on the combination of the baseband amplitude and carrier phase. In practice, amplitude shapes of the phase-modulated signals play a very important role in determining the bandwidth of the RF-transmitted signals. An illustration of these two modulation types is shown in Fig. 2.3, where two kinds of pulse shapes or a squared waveform and one-half cycle of a sinusoid are used. In the former, the modulation process only changes the phase of the carrier signal. In the latter, it changes not only the phase of the carrier signal, but also its amplitude. Later, we shall see that the bandwidth efficiency of the modulated signal can significantly benefit from the property of such a smooth pulse shape. Figure 2.4 shows a block diagram of a 2-level (or $M=2$) PAM or bi-phase PSK (BPSK) transmitter in the radio frequency (RF) or the intermediate frequency (IF) domain.

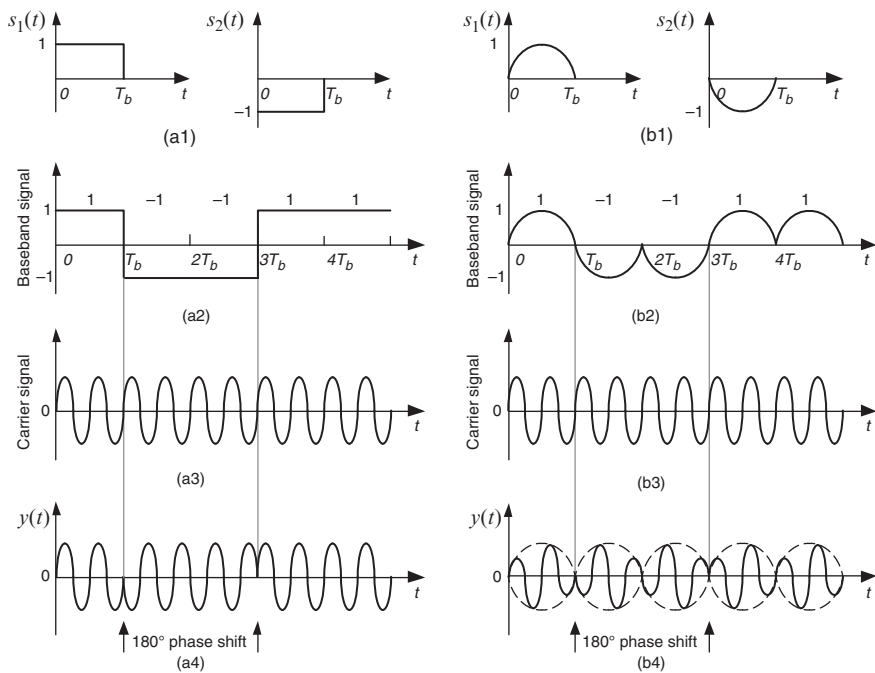


Fig. 2.3 Signal waveforms of PAM and PSK for $M = 2$: (a1, b1) pulse signals, (a2, b2) baseband signals, (a3, b3) carrier signals, and (a4, b4) modulated signals

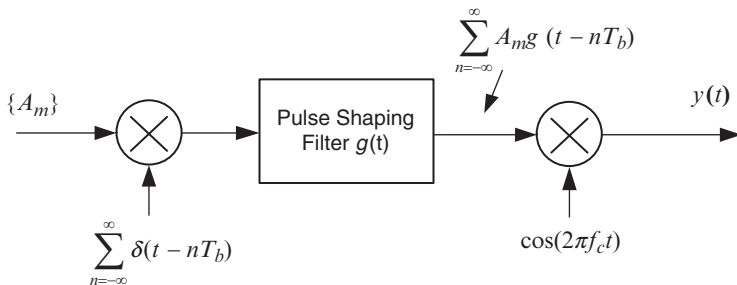


Fig. 2.4 Block diagram of a 2-level PAM or BPSK modulator

2.4.2 Quadrature Amplitude Modulation (QAM) and Quadrature Phase Shift Keying (QPSK)

In order to increase the bandwidth efficiency of 2-level PAM and BPSK signals, we can use two independent baseband signal streams to modulate a pair orthogonal carrier signals. The simplest quadrature modulation format is 4-ary quadrature

PAM, which is more often called 4-ary quadrature amplitude modulation (4QAM), or quadrature PSK (QPSK). The bandwidth efficiency of QPSK is twice the bandwidth efficiency of BPSK because the information transmitted over the same bandwidth is doubled. In practice, QPSK type modulation is widely used in many standards due to its robust BER performance compared with high-level QAM formats.

In general, QPSK signal can be expressed as

$$\begin{aligned}
 y(t) &= \sum_{n=-\infty}^{\infty} A_{mi} g(t - nT_s) \cos(2\pi f_c t) - \sum_{n=-\infty}^{\infty} A_{mq} g(t - nT_s) \sin(2\pi f_c t) \\
 &= s_{mi}(t) \cos(2\pi f_c t) - s_{mq}(t) \sin(2\pi f_c t) \\
 &= M_m(t) \cos[2\pi f_c t + \theta_m(t)], \quad m = 1, 2 \quad 0 \leq t \leq T_s
 \end{aligned} \tag{2.34}$$

where the modulus and the phase are calculated by

$$\begin{aligned}
 M_m(t) &= \sqrt{s_{mi}^2(t) + s_{mq}^2(t)} \\
 \theta_m(t) &= \tan^{-1}[s_{mq}(t)/s_{mi}(t)]
 \end{aligned} \tag{2.35}$$

In (2.34), A_{mi} and A_{mq} , which are independent from each other, are shown the symbol amplitudes of the in-phase (I) and quadrature (Q) branches, and $s_{mi}(t)$ and $s_{mq}(t)$ are the baseband waveforms of the I and Q branches. Here T_s is the symbol duration and is equal to twice the bit duration, or $T_s = 2T_b$.

A QPSK modulator consists of two BPSK modulators plus a serial-to-parallel converter and combiner as shown in Fig. 2.5. In a QPSK modulator, the input bit stream must be converted into the symbol stream on the I and Q branches. Such a

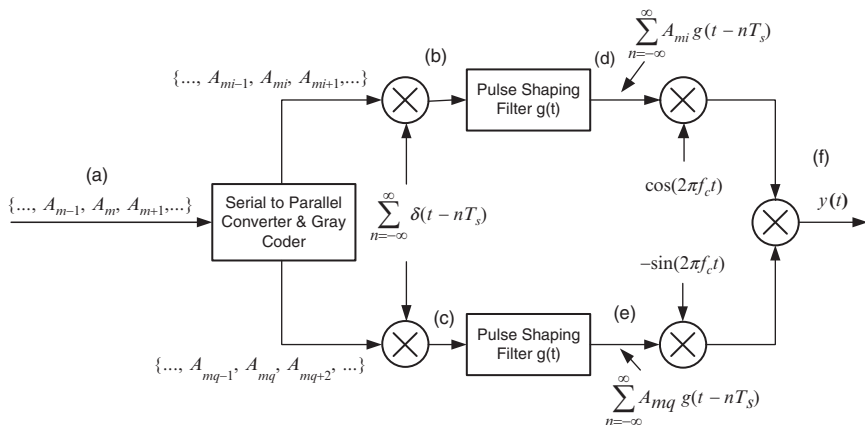


Fig. 2.5 Block diagram of QPSK modulator

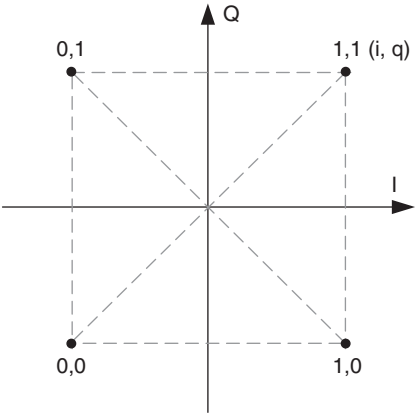
conversion can be realized through either a serial-to-parallel converter or a mapping circuit, in which two consecutive bits are converted into a pair of symbols on the I and Q branches. In the conversion, the bit interval T_b before the conversion becomes the symbol interval T_s after the conversion, which is $T_s = 2T_b$. This means that the bit rate is twice the symbol rate for QPSK. Thus, the theoretical spectral efficiency of 2 bit/s/Hz can be achieved with QPSK modulation in some applications, where the 1-bit/s/Hz theoretical spectral efficiency of BPSK modulation is insufficient to provide the available bandwidth efficiency. In QPSK modulation, every pair of symbols on the I and Q branches determines the carrier phase state. The four possible symbols result in the four different carrier phase states. Usually, the mapping format from the four possible symbols to the four phase states is performed in accordance with the *Gary code* as listed in Table 2.3, in which the signs of each pair of adjacent symbols after coding differ by only one to the corresponding symbols as illustrated in Fig. 2.6. The advantage of the Gary code is to ensure that a single symbol error corresponds to a single bit error after the parallel-to-serial converter at the receiver [12].

All the waveforms corresponding to points from (a) to (f) in Fig. 2.5 are plotted in Fig. 2.7, where the rectangular pulse for $g(t)$ is assumed. The phase transition of the modulated carrier signal $y(t)$ in Fig. 2.7f is calculated using (2.35) based on the waveforms in Fig. 2.7d, e. The phase transition of the modulated carrier signal $y(t)$ can be also obtained from its constellation diagram in Fig. 2.8.

Table 2.3 Two-bit binary to Gary code

Decimal	Binary	Gary code	Gary code as decimal
0	00	00	0
1	01	01	1
2	10	11	3
3	11	10	2

Fig. 2.6 Constellation of QPSK baseband signals



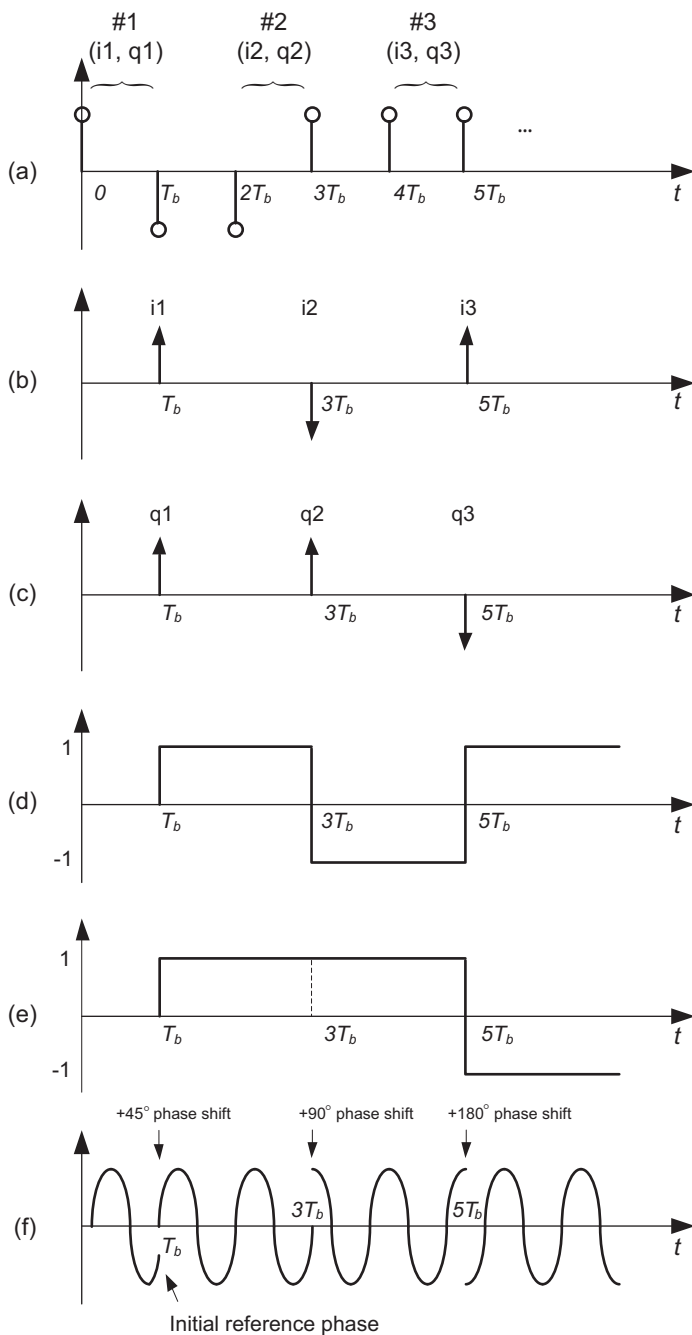
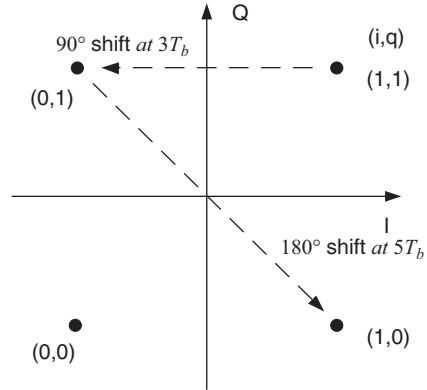


Fig. 2.7 Signal waveforms for QPSK in Fig. 2.4: (a) information bits, (b) symbol impulses of the I branch, (c) symbol impulses of the Q branch, (d) baseband waveform of the I branch, (e) baseband waveform of the Q branch, (f) phase-modulated waveform

Fig. 2.8 Phase transitions on constellation of QPSK for Fig. 2.7f



2.4.3 Power Spectral Density of Baseband Signals

Most digital communication systems are highly band-limited because the usable spectrum resources are severely congested. As a result, system architects must consider the bandwidth-efficient modulation technique as the highest priority in the assigned channel. The bandwidth efficiency of a modulated signal is usually characterized by its power spectral density (PSD), which is the distribution of power as a function of frequency. Therefore the designers can choose the modulation format based on its PSD characteristic to meet the requirements of the channel bandwidth.

For a random process, $X(t)$, its PSD $\Psi(f)$ and autocorrelation function $R(\tau)$ are related through the Fourier transform, or the Fourier transform

$$\Psi(f) = F[R(\tau)] = \int_{-\infty}^{\infty} R(\tau) e^{-j2\pi f\tau} d\tau \quad (2.36)$$

and the inverse Fourier transform

$$R(\tau) = F^{-1}[\Psi(f)] = \int_{-\infty}^{\infty} \Psi(f) e^{j2\pi f\tau} df \quad (2.37)$$

If τ is equal to zero, (2.37) becomes

$$R(0) = \int_{-\infty}^{\infty} \Psi(f) df \quad (2.38)$$

In above expression, $R(0)$ represents the average power of the random process, which is equal to the area under $\Psi(f)$.

For a stationary process, the autocorrelation function of $X(t)$ is independent of time and defined as

$$R(\tau) = R(t_1 - t_2) = E[X(t_1)X(t_2)] = E[X(t)X(t + \tau)] \quad (2.39)$$

where t_1 and t_2 are two different time instants and $E[\bullet]$ presents the ensemble average.

Most bandwidth efficient modulation techniques use a pair of orthogonal carrier signals $\cos(2\pi f_c t)$ and $\sin(2\pi f_c t)$ to carry two independent baseband signals $x_I(t)$ and $x_Q(t)$ on the I and Q branches, which are equivalent to $s_{mi}(t)$ and $s_{mq}(t)$ in (2.34), respectively. This modulation scheme is called *quadrature modulation*. The expression for such a quadrature-modulated signal is given as

$$s(t) = x_I(t) \cos(2\pi f_c t) - x_Q(t) \sin(2\pi f_c t) \quad (2.40)$$

The baseband signals on the I and Q branches are represented in general form:

$$x_I(t) = \sum_{n=-\infty}^{\infty} a_n g(t - nT_s), \quad x_Q(t) = \sum_{n=-\infty}^{\infty} b_n g(t - nT_s) \quad (2.41)$$

where the information symbols a_n and b_n are independent with equiprobable values, T_s is the symbol interval, and $g(t)$ is the spectrum-shaping pulse.

Another representation of the signal in (2.40) is

$$\begin{aligned} s(t) &= \text{Re}\{[x_I(t) + jx_Q(t)]e^{j2\pi f_c t}\} \\ &= \text{Re}\{x_L(t)e^{j2\pi f_c t}\} \end{aligned} \quad (2.42)$$

where $\text{Re}\{\}$ stands for the real part of the complex signal in the bracket and $x_L(t)$ is the *equivalent lowpass signal* of the modulated signal $s(t)$.

To derive the expression for PSD, we assume that $s(t)$ is a stationary processing, and the baseband signals of $x_I(t)$ and $x_Q(t)$ have zero mean values. The autocorrelation function of $s(t)$ in (2.42) is

$$R_s(\tau) = \text{Re}\{R_{x_L}(\tau)e^{j2\pi f_c \tau}\} \quad (2.43)$$

The PSD of the modulated signal in (2.40) is the Fourier transform of $R_s(\tau)$, or

$$\begin{aligned} \Psi_s(f) &= F[R_s(\tau)] = \int_{-\infty}^{\infty} \{\text{Re}[R_{x_L}(\tau)e^{j2\pi f_c \tau}]\} e^{-j2\pi f \tau} d\tau \\ &= \int_{-\infty}^{\infty} \left\{ \frac{1}{2} [R_{x_L}(\tau)e^{j2\pi f_c \tau} + R_{x_L}^*(\tau)e^{-j2\pi f_c \tau}] \right\} e^{-j2\pi f \tau} d\tau \\ &= \frac{1}{2} [\Psi_{x_L}(f - f_c) + \Psi_{x_L}(-f - f_c)] \end{aligned} \quad (2.44)$$

where $\Psi_{x_L}(f)$ is the PSD of the equivalent lowpass process $x_L(t)$. In (2.44), we used the property $R_{x_L}^*(\tau) = R_{x_L}(-\tau)$ and the quality property of the Fourier transform $e^{j2\pi f_c t} x(t) \leftrightarrow X(f - f_c)$.

Similar to (2.37), the PSD of the modulated signal expressed in (2.44) is the shifted PSD of the equivalent lowpass signal $x_L(t)$ to the center frequency $\pm f_c$ except the constant factor $1/2$. It becomes useful to determine the PSD of the modulated signal through the PSD of the equivalent lowpass signal.

The derivations of the PSD of the equivalent lowpass signal $x_L(t)$ are complicated and fairly long and beyond the scope of this book. Here we just present its expression

$$\Psi_{x_L}(f) = \sigma^2 f_s |G(f)|^2 + \mu^2 f_s^2 |G(0)|^2 \delta(f) + 2\mu^2 f_s^2 \sum |G(mf_s)|^2 \delta(f - mf_s) \quad (2.45)$$

where σ^2 and μ are the variance and mean of the sequence of information symbols $\{a_n\}$ or $\{b_n\}$, respectively. $G(f)$ is the Fourier transform of the shaping pulse $g(t)$ and $f_s = 1/T_s$ is the symbol rate.

The expression (2.45) consists of three terms to emphasize the three different types of spectral components. The first term is the continuous spectral component, and its shape is completely dependent on the squared module of the Fourier transfer of the shaping pulse $g(t)$. The second term is the DC component. The third term consists of discrete harmonics, each spaced f_s apart in frequency.

If the information binary symbols $\{a_n\}$ or $\{b_n\}$ are assumed to be independent random variables with equal probability for the values of ± 1 , the information symbols have zero mean and unit variance. In this case, all discrete spectral harmonics and DC components vanish and only the continuous component is left in (2.45). This condition is valid for most digital modulation techniques. Thus, the designer can simply choose a proper shaping pulse $g(t)$ to achieve a bandwidth-efficient transmission system. In this condition, (2.45) can be simplified to

$$\Psi(f) = \Psi_{x_L}(f) = f_s |G(f)|^2 \quad (2.46)$$

Here, we drop off the subscript x_L for the purpose of simplicity.

It is clearly seen from (2.46) that the PSD of the equivalent baseband signal $s_L(t)$ is proportional to the squared module of $|G(f)|^2$ of the Fourier transform of the equivalent baseband signal. Meanwhile, the PSD of the modulated signal, which is expressed in (2.44), is the shifted PSD of the equivalent lowpass signal $x_L(t)$ to the center frequency $\pm f_c$ except the constant factor $1/2$. Therefore finding a spectrum-shaping pulse with the property of the narrow main lobe and small side-lobes in the frequency domain plays an important role in achieving high bandwidth efficiency for the spectrum-shaping types of modulation techniques.

2.4.4 Non-Overlapped Pulse Waveform Modulation

From (2.46), we know that the PSD shape of the modulated signal depends only on the shape of the Fourier transform of the shaping pulse $g(t)$. The fundamental type of shaping pulse that is widely used in digital modulation techniques is a non-overlapped shaping pulse. This non-overlapped shaping pulse only lasts one

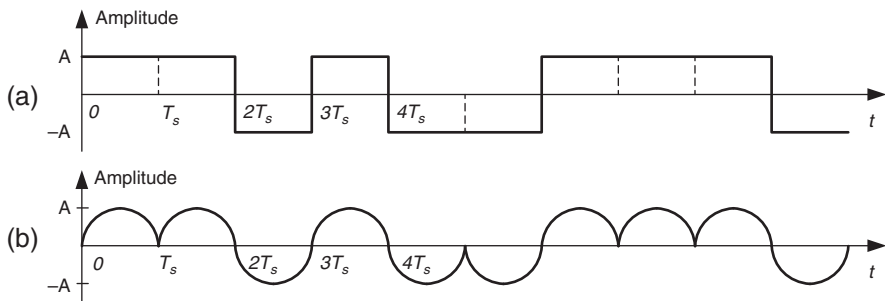


Fig. 2.9 Two different types of baseband waveforms: (a) signals consisting of non-return to zero (NRZ) and (b) signals consisting of one-half cycle of a sinusoid

symbol interval T_s in the time domain. For a BPSK signal, the symbol interval T_s equals the bit interval T_b . Two widely used non-overlapped pulses are the rectangular and the one-half cycle of a sinusoid. The former is used for the unfiltered BPSK/QPSK/OQPSK modulations, while the latter is used for MSK modulation and is also adopted in the ZIGBEE standard [13].

To see the PSD related with Fourier transform of the pulse $g(t)$, first we consider the rectangular pulse whose expression is given by

$$g(t) = \begin{cases} A, & 0 \leq t \leq T_s \\ 0, & \text{elsewhere} \end{cases} \quad (2.47)$$

This pulse is used to weight the sequence of random variables, each having zero mean and unit variance, as illustrated in Fig. 2.9a. The Fourier transform of the $g(t)$ is

$$\begin{aligned} G(f) &= \int_{-\infty}^{\infty} g(t) e^{-j2\pi f t} dt = \int_0^{T_s} A e^{-j2\pi f t} dt \\ &= A T_s \frac{\sin(\pi f T_s)}{\pi f T_s} e^{-j\pi f T_s} \end{aligned} \quad (2.48)$$

Hence

$$|G(f)|^2 = A^2 T_s^2 \left(\frac{\sin(\pi f T_s)}{\pi f T_s} \right)^2 \quad (2.49)$$

Since the mean and variance of the random information sequences are zero and unit, respectively, we can use (2.46) to calculate the PSD of the QPSK signal:

$$\Psi_{\text{QPSK}}(f) = A^2 T_s \left(\frac{\sin(\pi f T_s)}{\pi f T_s} \right)^2 \quad (2.50)$$

Next, we consider the one-half cycle sinusoid pulse expressed as

$$g(t) = \begin{cases} A \sin(\pi t/T_s), & 0 \leq t \leq T_s \\ 0, & \text{elsewhere} \end{cases} \quad (2.51)$$

The random waveform weighted with such a pulse is illustrated in Fig. 2.9b. Similar to the derivation above, the Fourier transform of the $g(t)$ is

$$\begin{aligned} G(f) &= \int_0^{T_s} A \sin(\pi t/T_s) e^{-j2\pi f t} dt \\ &= \frac{2AT_s}{\pi} \frac{\cos(\pi f T_s)}{1 - 4f^2 T_s^2} e^{-j\pi f T_s} \end{aligned} \quad (2.52)$$

From (2.46) the power spectral density is

$$\Psi_{\text{MSK}}(f) = \frac{4A^2 T_s}{\pi^2} \left(\frac{\cos(\pi f T_s)}{1 - 4f^2 T_s^2} \right)^2 \quad (2.53)$$

It is seen in (2.50) and (2.53) that the main lobe of the power spectral density (PSD) of QPSK/OQPSK is located at $f = 1/T_s$, while the main lobe of MSK is located at $f = 1.5/T_s$. This means that the main lobe of MSK is 50% wider than that for QPSK/OQPSK. On the other hand, the side lobes in QPSK/OQPSK fall off at a rate of f^{-2} , while the side lobes in MSK drop at a rate of f^{-4} , which is faster than QPSK/OQPSK.

It is also noted that the wider the pulse width T_s in the time domain is, the narrower the main lobe in the frequency domain is. Hence, a pulse period lasting more than T_s would increase the bandwidth efficiency. The baseband signals can be generated by overlapping pulse sequences, which will be introduced in the next section.

2.5 Overlapped Pulse-Shaping Modulation

To increase the bandwidth efficiency by means of extending the period of the pulse lasting, it is also preferable to be either free of intersymbol interference or have minimal ISI. Besides achieving bandwidth efficiency, it is also important for the modulated signals to achieve energy efficiency when passing through nonlinear power amplifiers. For most modulation techniques with bandwidth and energy efficiency, data in the Q channel is delayed by a half-symbol interval $T_s/2$ relative to data in the I channel in order to reduce the envelope fluctuation. Equation (2.41) can be rewritten as

$$\begin{aligned}
 x_I(t) &= \sum_{n=-\infty}^{\infty} a_n g(t - nT_s) \\
 x_Q(t) &= \sum_{n=-\infty}^{\infty} b_n g(t - nT_s - T_s/2)
 \end{aligned}
 \tag{2.54}$$

2.5.1 Overlapped Raised-Cosine Pulse-Shaping Modulation

A basic ideal for the overlapped raised-cosine modulation is to achieve ISI free at the decision instants in the time domain by overlapping two consecutive shaping pulses, in which each pulse lasts exactly twice symbol interval of $2T_s$ in the time domain and has a narrow main lobe and fast roll-off of sidelobes of the Fourier transform in the frequency domain. The Raised-cosine pulse with the interval of $2T_s$ is obtained by convolving a square pulse with a half-cycle of the sine waveform, each with interval of T_s .

The convolution of two pulse signals is to simply let one pulse signal pass through a filter $H(s)$ with an impulse response of $h(t)$, which is equal to another pulse signal, as shown in Fig. 2.10.

One early-published paper regarding the convolution pulse-shaping method is Quadrature Overlapped Raised-Cosine (QORC) modulation [14]. The input $x(t)$ to the filter is a square waveform and impulse response of the filter $h(t)$ is a half-cycle sine signal

$$x(t) = \begin{cases} 1, & 0 \leq t \leq T_s \\ 0, & \text{elsewhere} \end{cases}
 \tag{2.55}$$

$$h(t) = \begin{cases} \frac{\pi}{2T_s} \sin\left(\frac{\pi t}{T_s}\right), & 0 \leq t \leq T_s \\ 0, & \text{elsewhere} \end{cases}
 \tag{2.56}$$

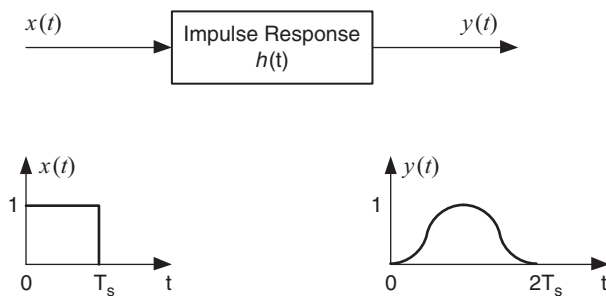


Fig. 2.10 Pulse-shaping filter

where amplitude $\pi/2T_s$ is used to normalize the filter output or raised-cosine pulse waveform. In fact, (2.55) is the pulse shape of QPSK signal while (2.56) is the pulse shape of MSK signal.

The output of the filter is the convolution of the input and impulse response

$$g(t) = x(t) * h(t) = \begin{cases} \frac{1}{2} \left(1 - \cos \frac{\pi t}{T_s} \right), & 0 \leq t \leq 2T_s \\ 0, & \text{elsewhere} \end{cases} \quad (2.57)$$

Thus, the duration of the raised-cosine pulse through the convolution operation becomes twice the duration T_s or $2T_s$.

Similar to the relationship between QPSK and Staggered QPSK (SQPSK), the symbol sequences on the Q branch for QORC can be delayed by a half-symbol interval offset $T_s/2$ relative to that on the I branch in order to reduce the envelope fluctuation of the modulated signal. This offset QORC is called the staggered QORC (SQORC) [14].

Figure 2.11 shows a block diagram of SQORC modulator, where a conceptual block diagram of raised-cosine (RC) pulse shaping can be implemented using either a filter as shown in Fig. 2.10 or a circuit illustrated in Fig. 2.12. In order to overlap raised-cosine pulses, each RC pulse-shaping needs one serial-to-parallel converter and two delay T_s devices.

Figure 2.13 shows the baseband waveforms of the SQORC signal. Figure 2.13a illustrates the SQORC baseband signal performed by the overlapping method implemented in Fig. 2.12. Figure 2.13b shows the eye diagram. It can be seen that there are no ISI at the sampling decision instants and no jitter at the jitter instants. The modulation technique with such a property is also called *Intersymbol Interference and Jitter-Free (IJF) OQPSK (IJF-OQPSK)*, which will be introduced in the next section. Figure 2.13c shows the constellation of the SQORC signal, where the maximum envelope fluctuation is 3 dB. The envelope is always 1 when consecutive symbols have alternative polarity on both the I channel and Q channel,

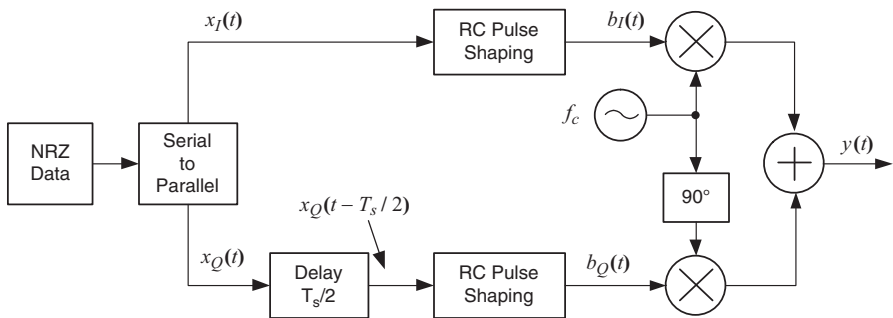


Fig. 2.11 Block diagram of QORC/SQORC modulator

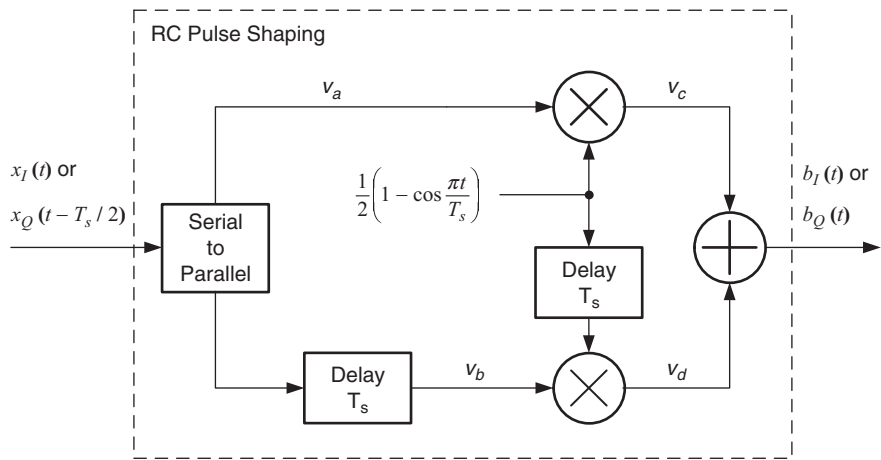


Fig. 2.12 Block diagram of RC pulse shaping

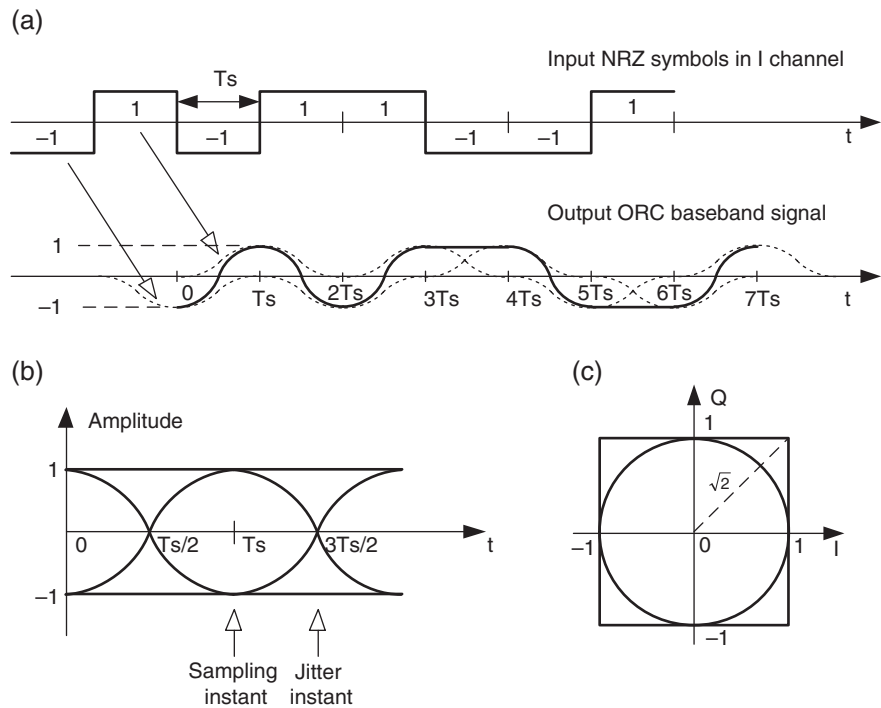


Fig. 2.13 Baseband waveform: (a) overlapped baseband signal, (b) eye diagram, and (c) constellation

while the maximum envelope amplitude occurs when consecutive symbols have the same polarity on both the I channel and Q channel. This results in the maximum envelope amplitude of $\sqrt{2}$, as shown in Fig. 2.13c. This amount of envelope fluctuation causes the regrowth of the PSD at the output of the power amplifier when it operates at the saturation or close to the saturation region.

Equation (2.57) in the time domain has the following relationship in the Fourier transform domain:

$$G(\omega) = X(\omega) \times H(\omega) \quad (2.58)$$

The square waveform $x(t)$ in (2.55) and impulse response $h(t)$ in (2.56) have the Fourier transform

$$X(\omega) = T_s \frac{\sin\left(\frac{\omega T_s}{2}\right)}{\frac{\omega T_s}{2}} e^{-j\omega T_s/2} \quad (2.59)$$

$$H(\omega) = \frac{\cos\left(\frac{\omega T_s}{2}\right)}{\left[1 - \left(\frac{\omega T_s}{\pi}\right)^2\right]} e^{-j\omega T_s/2} \quad (2.60)$$

Substituting (2.59) and (2.60) into (2.58), the Fourier transform of the filter output is

$$G(\omega) = \frac{\sin(\omega T_s)}{\omega \left[1 - \left(\frac{\omega T_s}{\pi}\right)^2\right]} e^{-j\omega T_s} \quad (2.61)$$

Hence

$$|G(f)|^2 = \frac{\sin^2(\omega T_s)}{\omega^2 \left[1 - \left(\frac{\omega T_s}{\pi}\right)^2\right]^2} \quad (2.62)$$

Since the mean and variance of the random information sequences are zero and unit, respectively, we can use (2.46) to calculate the PSD for SQORC signal as

$$\Psi_{\text{SQORC}}(f) = f_s |G(f)|^2 = T_s \frac{\sin^2(2\pi f T_s)}{(2\pi f T_s)^2 \left[1 - \left(\frac{2\pi f T_s}{\pi}\right)^2\right]^2} \quad (2.63)$$

The normalized PSD is

$$\frac{\Psi_{\text{SQORC}}(f)}{\Psi_{\text{SQORC}}(0)} = \frac{\sin^2(2\pi f T_s)}{(2\pi f T_s)^2 \left[1 - \left(\frac{2\pi f T_s}{\pi}\right)^2\right]^2} \quad (2.64)$$

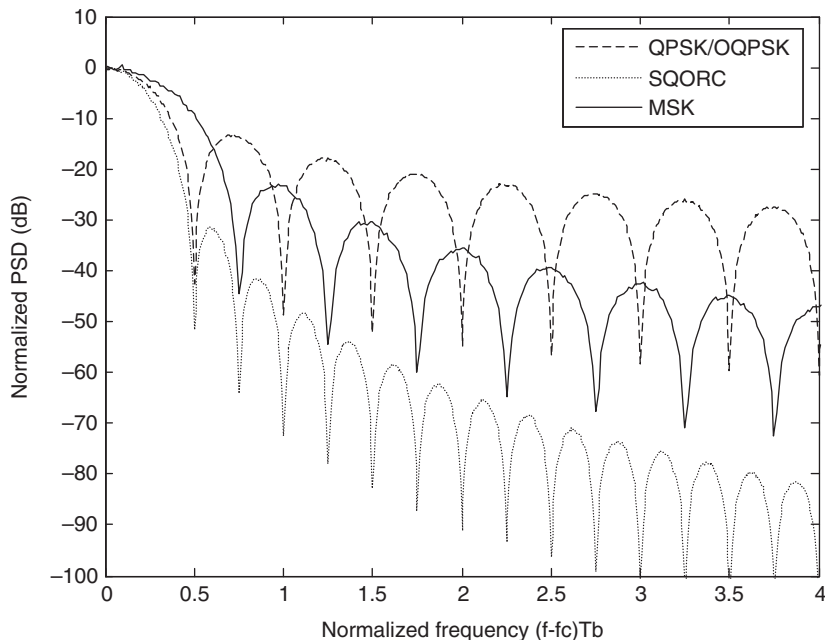


Fig. 2.14 Power spectral densities of OQPSK, MSK and SQORC in a linear channel

The power spectral density curves of the QPSK/OQPSK, MSK, and SQORC signals in a linear channel are illustrated in Fig. 2.14. The main lobe of the PSD of SQORC signal is located at $f = 1/T_s = 0.5/T_b$, which is the same as QPSK/OQPSK signals and the main lobe of MSK signal is located at $f = 1.5/T_s = 0.75/T_b$. The side lobes in SQORC fall off at a rate of f^{-6} , which is much faster compared with QPSK/OQPSK at a rate of f^{-2} and MSK at a rate of f^{-4} . In fact, the PSD of SQORC takes on the form of the product of the PSDs of QPSK/OQPSK and MSK because the Fourier transform of QORC/SQORC is the product of the Fourier transforms of QPSK/OQPSK and MSK. So the main lobe of SQORC retains the same width as the main lobe of QPSK/OQPSK and the side lobes of SQORC become narrower and fall off at a rate of f^{-6} , which is equal to the addition of their rates of f^{-2} and f^{-4} .

2.5.2 IJF-OQPSK Modulation

Intersymbol interference and jitter-free (IJF-OQPSK) [15] baseband signals are identical to SQORC baseband signals except for their implementation methods. The former uses a simple and precise logic switch circuit to generate its baseband waveforms instead of using overlapped raised-cosine pulse. This method resulted in

fast applications of IJF-OQPSK/SQORC in satellite earth stations in the late 1980s. After that, a look-up table (LUT) based IJF-OQPSK and SQORC was widely used in digital communication systems to achieve energy and spectrally efficient transmissions.

As we have seen, overlapping a current raised-cosine pulse with $2T_s$ and a previous raised-cosine pulse forms a current SQORC baseband signal during the symbol interval T_s . In turn, it is determined by a current and previous symbol data. Thus a total of four types of baseband signals in one symbol interval T_s depending on combinations of the input symbol data are plotted in Fig. 2.15.

In Fig. 2.15, the shape of the current baseband signal during T_s is determined by the combination of the current and previous Non-Return-to-Zero (NRZ) data (symbols) $d_n d_{n-1}$ in either I channel or Q channel. An actual hardware implementation of IJF-OQPSK baseband signal is illustrated in Fig. 2.16. Each pair $d_n d_{n-1}$ of the current and previous NRZ symbols in the I and Q channels are used as addresses to turn on one of four switches in each symbol duration. The output baseband signal

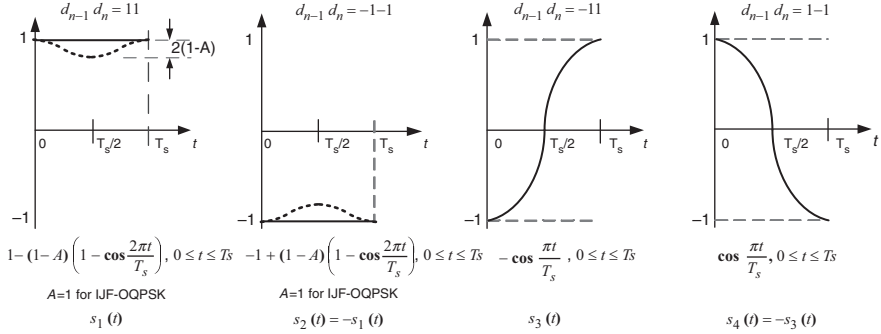
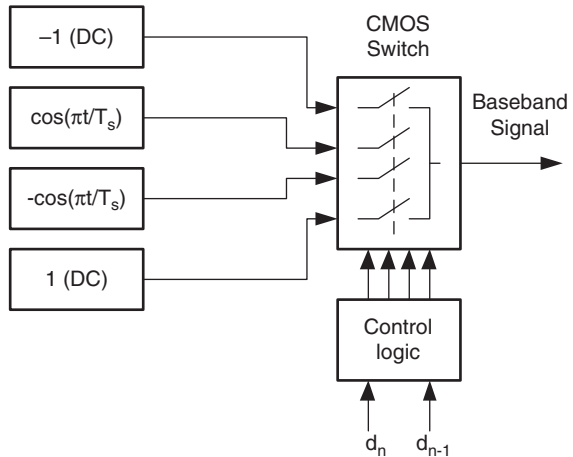


Fig. 2.15 Total four overlapped baseband segments in one symbol interval T_s , where the solid-line waveforms are for IJF-OQPSK and dot-line are for SQAM with $0.5 \leq A \leq 1.0$

Fig. 2.16 IJF-OQPSK switch based baseband signal generator. Redrawn from [12]



is sent to the I-Q modulator to modulate the carrier signal after passing through a smooth lowpass filter to remove high frequency harmonics due to switching operation. The baseband waveform, eye diagram and constellation of the IJF-OQPSK are identical to those for SQORC as shown in Fig. 2.13. The mathematic expression PSD for IJF-OQPSK is also the same as that for SQORC as expressed in (2.64) and plotted in Fig. 2.14.

2.5.3 Other Overlapped Pulse-Shaping Modulations

SQORC and IJF-OQPSK signals have a 3-dB envelope fluctuation, as shown in Fig. 2.13c. As a result, the spectral side lobes slightly spread up due to AM-AM and AM-PM conversions of the power amplifiers after they pass through nonlinear amplification channels. In order to reduce such a 3 dB envelope fluctuation and also keep the interval T_s of the pulse waveform unchanged, two modified overlapped pulse-shaping waveforms, called superposed quadrature amplitude modulation (SQAM) [16] and self-convolving minimum shift key (SCMSK) modulation [17], were proposed. In the former, the pulse waveform with the interval of $2T_s$ is generated by superposing two raised-cosine pulses, each having the interval of T_s and opposite polarity, to the raised-cosine pulse with the interval of $2T_s$ that is the same as a pulse-shaping waveform used in SQORC/IJF-OQPSK. In the latter, the pulse waveform with the interval of $2T_s$ is created by convolving a half-cycle of a sinusoidal pulse with the interval of T_s with itself. In the following, major concepts of these two modulation techniques are described simply. The interested reader can reference the Appendix C.3 for the detailed derivations.

SQAM: The SQAM was developed based on IJF-OQPSK modulation for the purpose of further reducing the maximum envelope fluctuation of the IJF-OQPSK by 3 dB. To form a SQAM pulse waveform, another two raised-cosine pulses, each having single symbol duration of T_s and adjustable amplitude parameter A , with negative polarity are superposed to the original raised-cosine pulse with double symbol interval of $2T_s$. The superposed pulse waveform is expressed by

$$s(t) = g(t) + d(t) \quad (2.65)$$

where

$$g(t) = \frac{1}{2} \left(1 + \cos \frac{\pi}{T_s}(t - T_s) \right) \quad (2.66)$$

$$d(t) = -\frac{1-A}{2} \left(1 - \cos \frac{2\pi t}{T_s} \right), \quad 0.5 \leq A \leq 1.0, \quad 0 \leq t \leq 2T_s \quad (2.67)$$

In (2.67), A is an adjustable amplitude parameter. Figure 2.17 illustrates the SQAM pulse-shaping process by adding two raised-cosine pulses $d(t)$ to one raised-cosine pulse $g(t)$. Figure 2.18 illustrates the comparison among SQAM pulse waveforms with different parameters of A , where $A=1$ corresponds to IJF-OQPSK signal.

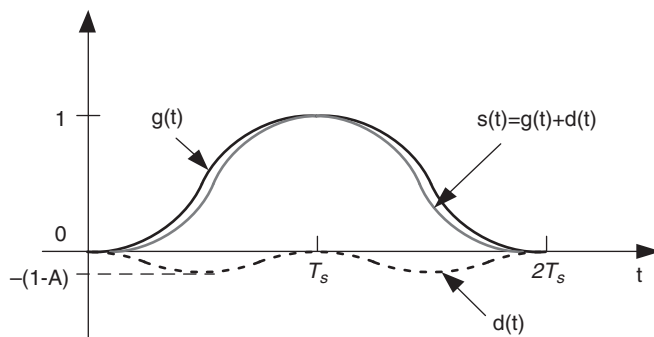


Fig. 2.17 SQAM pulse shaping by superposing two raised-cosine pulses with symbol interval of T_s to one with $2T_s$

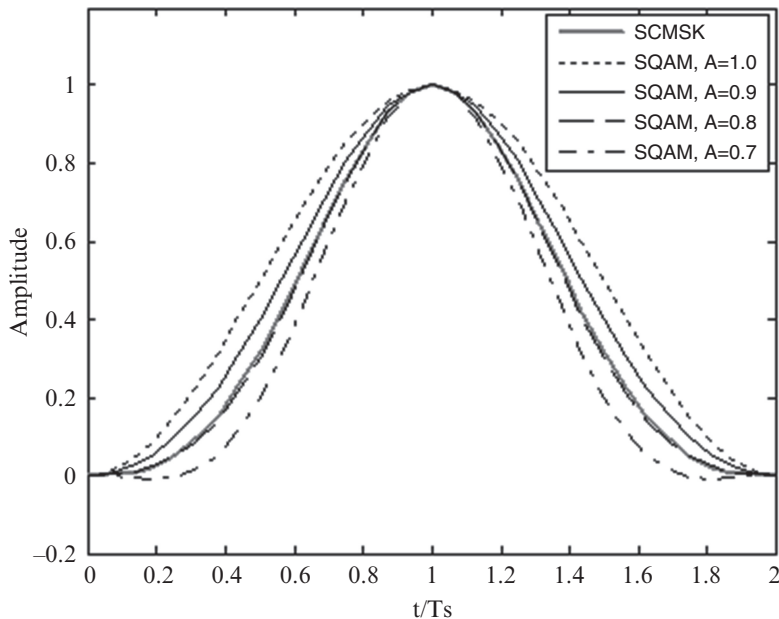
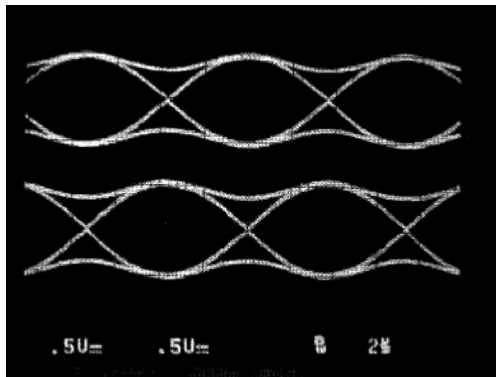


Fig. 2.18 Shaping pulses of SCMSK and SQAM in a twice symbol interval $2T_s$. Note that SQAM with $A=1$ is equal to IJF-OQPSK pulse

Fig. 2.19 Eye diagrams of SQAM baseband I–Q signals with $A = 0.8$ at the symbol rate of 135.416 kilosymbols per second (or the bit rate of 270.833 kilobits per second is divided by 2). Referenced from [18]



The SQAM modulator is the same as the IJF-OQPSK modulator except for the replacement of IJF-OQPSK waveform segments of $s_1(t)$ and $s_2(t)$ by SQAM waveform segments of $s_1(t)$ and $s_2(t)$, as shown in Fig. 2.15. To avoid the maximum envelope fluctuation of 3 dB that appears when two consecutive symbols have the same polarity in both the I and Q channels for IJF-OQPSK, the waveforms around the centers of the segments $s_1(t)$ and $s_2(t)$ for SQAM are reduced from 1 to $1 - 2 \times (1 - A)$ with $0.5 \leq A \leq 1.0$.

For example, the maximum envelope is significantly reduced from 3 to 0.7 dB when A changes from 1 to 0.7. Therefore, IJF-OQPSK is a special case of SQAM signal at $A = 1$. Figure 2.19 illustrates eye diagrams of SQAM-baseband signals with $A = 0.8$ at the symbol rate of 135.417 kilosymbols/s, which corresponds to the bit rate of 270.833 kbits/s for a GMSK signal in the 2G GSM system. A detailed description of the SQAM signal is given in Appendix C.

SCMSK: The idea of SCMSK pulse waveform generation was triggered by a SQORC pulse waveform generation introduced in (2.57), in which its pulse waveform is generated by convolving a rectangular pulse of QPSK in (2.55) with one half-cycle of the sinusoidal pulse of MSK in (2.56); and hence the PSD of the SQORC modulated signal takes the form of the production of the power spectral densities of QPSK and MSK. This fact reveals a way to find a new overlapped pulse whose PSD has fast roll-off sidelobes by convolving the pulse shaping waveforms of another two modulation signals, each having the symbol interval of T_s .

Similar to SQORC, a SCMSK pulse waveform is generated by convolving a half-cycle of sinusoidal pulse of MSK signal with itself, or by letting one half-cycle of the sinusoidal pulse pass through a filter with the impulse response of another half-cycle of the sinusoidal pulse. As a result, the sidelobes of SCMSK signal roll off as twice fast as the sidelobes of MSK signal, while the main lobe of SCMSK signal remains the same as the main lobe of MSK signal. The pulse waveform of SCMSK is generated as

$$g(t) = x(t) * h(t) \quad (2.68)$$

The input pulse to the filter and the impulse response of the filter are given by

$$x(t) = \begin{cases} \sin(\pi t/T_s), & 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (2.69)$$

$$h(t) = \begin{cases} \frac{2}{T_s} \sin(\pi t/T_s), & 0 \leq t \leq T_s \\ 0, & \text{otherwise} \end{cases} \quad (2.70)$$

The convolution of $x(t)$ and $h(t)$ is easily derived by substituting (2.69) and (2.70) into (2.68) and it may be expressed in the form

$$g(t) = \begin{cases} (1/\pi) \sin(\pi t/T_s) - (t/T_s) \cos(\pi t/T_s), & 0 \leq t \leq T_s \\ -(1/\pi) \sin(\pi t/T_s) + (t/T_s - 2) \cos(\pi t/T_s), & T_s \leq t \leq 2T_s \end{cases} \quad (2.71)$$

The pulse waveform of SCMSK in the time domain within the twice symbol interval of $2T_s$ is shown in Fig. 2.18, where it is very close to the pulse waveform of SQAM with $A=0.8$, which creates the best PSD and BER performances in both linear and nonlinear channels compared with other A parameters. Therefore, it is expected that their PSDs should be close to each other as well. Figures 2.20 and 2.21 show PSDs in a linear channel and a nonlinear channel, respectively. It can be seen that PSD of the SQAM signal with $A=0.8$ rolls off the fastest, followed by the

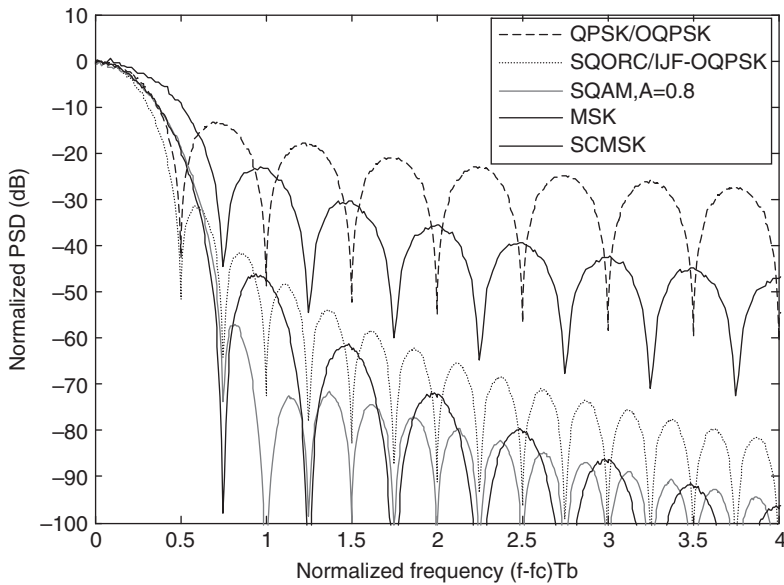


Fig. 2.20 Power spectral densities for QPSK/OQPSK, SQORC/IJF-OQPSK, SQAM, MSK, and SCMSK in a linear channel, where $T_b = T_s/2$ is the bit duration

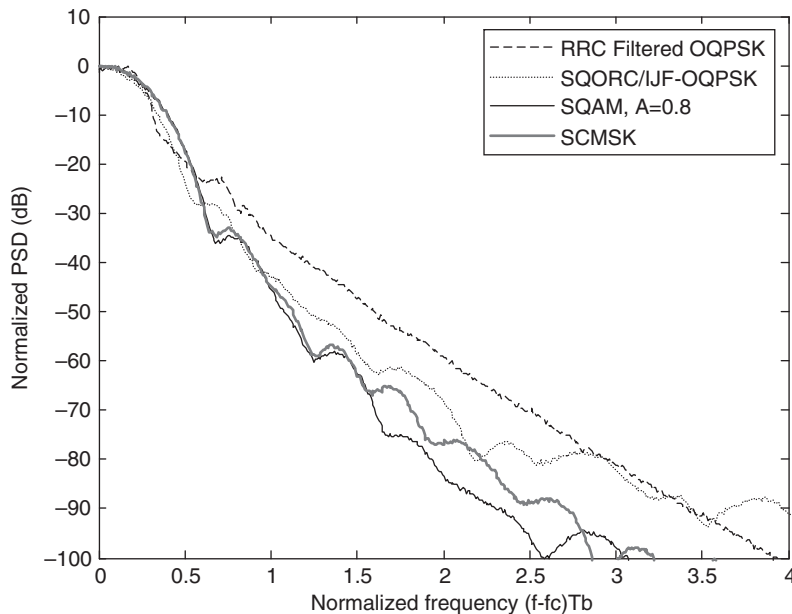


Fig. 2.21 Power spectral densities for OQPSK, SQORC/IJF-OQPSK, SQAM, and SCMSK in a nonlinear channel, where OQPSK is passed through a root-raised cosine filter with $\alpha = 0.35$, where $T_b = T_s/2$ is the bit duration

SCMSK signal in the linear channel. In a nonlinear channel, the SQAM signal with $A = 0.8$ rolls off as fast as the SCMSK signal up to $(f - f_b)T_b = 1.5$. The main lobe of SQAM signal $A = 0.8$ is the same as the SCMSK signal and both of them are equal to the normalized frequency $(f - f_b)T_b = 0.75$, which is wider than the main lobe of QPSK signal, which is equal to $(f - f_b)T_b = 0.5$. Figure 2.22 illustrates E_b/N_o degradation of SQAM signal versus the parameter of A compared with an ideal theoretical QPSK $E_b/N_o = 8.4$ dB at $\text{BER} = 10^{-4}$ in a linear channel. The best performance in linear and nonlinear channels is obtained for $A = 0.8$. With this value, E_b/N_o is degraded by 0.3 dB in a linear channel and 0.5 dB in a nonlinear channel [16]. It has been demonstrated that the SCMSK signal has the same E_b/N_o performance as SQAM signal in both linear and nonlinear channels because of their identical similarities in their pulse waveforms in Fig. 2.18.

It is seen from Fig. 2.18 in the time domain and Fig. 2.20 in the frequency domain that for the cases of SQAM signal with $A = 0.8$ and $A = 1$ (or IJF-OQPSK), the more smoothly the pulse reaches zero in the time domain, the faster the side lobes of the PSD roll off in the frequency domain. On the other hand, the narrower the pulse around the center is in the time domain, the wider the main lobe of the PSD is in the frequency domain. This phenomenon is easily understood from the property of the Fourier transform that the more smoothly the pulse reaches zero, the faster the high-frequency components decay, and the narrower the pulse around the center is, the wider the bandwidth occupied by the low-frequency components is.

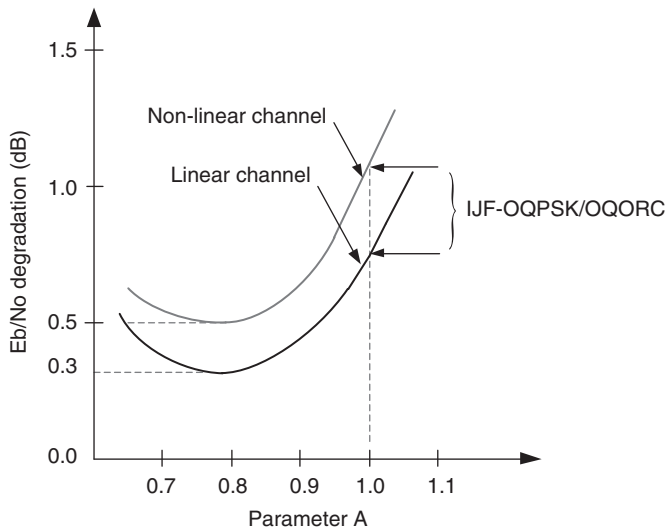


Fig. 2.22 E_b/N_o degradation of SQAM signal from an ideal theoretical QPSK in a linear channel requires $E_b/N_o = 8.4$ dB at $\text{BER} = 10^{-4}$. Redrawn from [16]

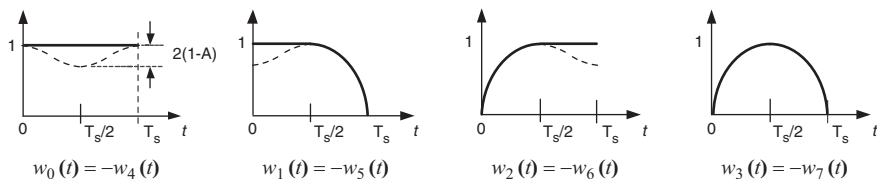


Fig. 2.23 Composite waveforms with one symbol interval T_s , where the solid-curves are for QORC and dashed-curves are for SQAM

2.5.4 Bit Error Rate in Coherent Demodulation

Like the demodulation for QPSK/OQPSK and MSK signal, an optimum cross-correlation-type receiver can be used to demodulate the overlapped pulse-shaping signals of QORC/SQORC and SQAM. The detailed receiver structures are described in [14, 19, 20], where an integrate-and-dump (I&D) filter is used on each correlation branch. Basically, for the cases of QORC/SQORC and SQAM signals there are all eight possible waveforms in any typical one symbol transmission interval $0 \leq t \leq T_s$, as shown in Fig. 2.23 and each of them occurs with equal probability. These eight waveforms are defined as

$$\begin{aligned}
w_0(t) &= 1 - (1 - A) \left(1 - \cos \frac{2\pi t}{T_s} \right), \quad 0 \leq t \leq T_s \\
w_1(t) &= \begin{cases} 1 - (1 - A) \left(1 + \cos \frac{2\pi t}{T_s} \right), & 0 \leq t \leq \frac{T_s}{2} \\ \sin \frac{\pi t}{T_s}, & \frac{T_s}{2} \leq t \leq T_s \end{cases} \\
w_2(t) &= \begin{cases} \sin \frac{\pi t}{T_s}, & 0 \leq t \leq \frac{T_s}{2} \\ 1 - (1 - A) \left(1 + \cos \frac{2\pi t}{T_s} \right), & \frac{T_s}{2} \leq t \leq T_s \end{cases} \\
w_3(t) &= \sin \frac{\pi t}{T_s}, \quad 0 \leq t \leq T_s
\end{aligned} \tag{2.72}$$

where $A = 1$ corresponds to QORC/SQORC or IJF-OQPSK signals and values in the range $0 < A \leq 1.0$ are used for a SQAM signal. Receivers for QORC/SQORC and SQAM signals traditionally perform symbol-by-symbol detection that employs simple integrate-and-dump (I&D) filters as detectors. This type of the detection is referred to as suboptimum detection in [14, 20]. In such a symbol-by-symbol detection, only four positive waveforms of the total eight waveforms are used to perform correlation detection at the suboptimum receiver, where each waveform individually multiplies the received signal on its branch before an integrate-and-dump filter as shown in Fig. 13 in [14]. Note that these four waveforms with the symbol period T_s have different energy within the symbol interval; and thus, the outputs of the I&D filters must be biased by the related waveform segment energy divided by 2 before the detector [20].

In general, the optimum correlation receiver maximizes the output signal energy during the symbol period T_s in a linear channel corrupted by the additive Gaussian noise by minimizing the equivalent noise bandwidth. This is equivalent to maximizing the signal-to-noise ratio (SNR). The objective of the correlation device is not to identify any one of the eight possible waveform segments. Instead it is to decide whether $+1$ or -1 was sent from the transmitter [14]. The optimum correlation receiver will be discussed in Chap. 3 in more detail.

It is shown in [14] that the simplest demodulator with the third-order Butterworth lowpass filter for QORC/SQORC yields a good BER performance in a linear channel corrupted with Gaussian noise, approximately 0.7 dB E_b/N_o degradation from an ideal system performance at $\text{BER} = 10^{-5}$ when compared with the I&D detection based demodulator. In fact, the Butterworth-type filter is used in many practical applications due to its simple implementation and slight BER degradation. Figure 2.24 illustrates the block diagram of a coherent demodulator with Butterworth filters for QPSK/OQPSK type of signals, including the overlapped pulse waveforms with pulse-shaping of QORC/SQORC and SQAM.

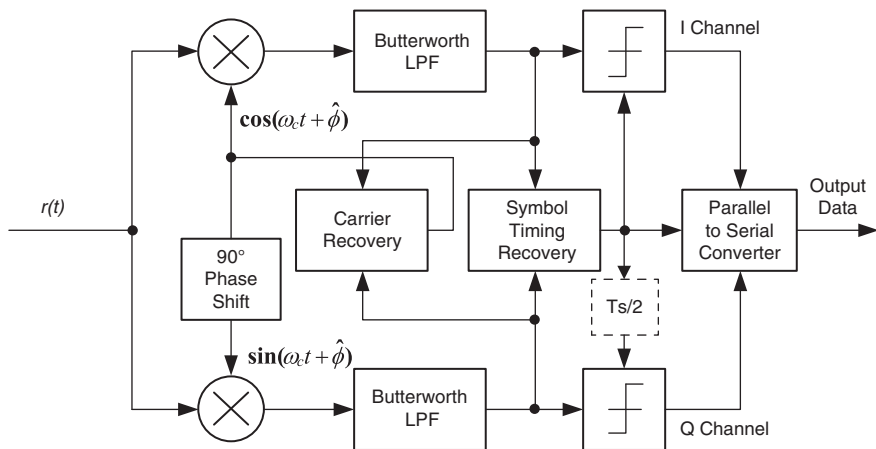


Fig. 2.24 Block diagram of a coherent demodulator with Butterworth filters for QPSK/OQPSK type of signals, where the delay $T_s/2$ is used for OQPSK signal only

The carrier phase estimate $\hat{\phi}$ for the received carrier phase ϕ is used in generating the local reference carrier signal for the coherent demodulator. If the phase error $\phi - \hat{\phi}$ is very small, the transmitted baseband signals on the in-phase and quadrature branches at the outputs of the lowpass filters are recovered. Then the symbol timing recovery circuit synchronizes the local symbol clock with the recovered baseband signals, and then samples the recovered baseband signals on the I-Q branches at the maximum eye-opening points to make decisions through the decision blocks. The sampling clock on the Q branch needs a half-symbol delay relative to the I branch if a type of offset QPSK modulation is used. Finally, the recovered symbol sequences on the I-Q branches are combined into serial output data through the parallel to serial converter.

A Butterworth lowpass filter is used here due to its mild properties of amplitude attenuation and group delay variation. A 3 dB bandwidth of the Butterworth lowpass affects the BER performance of the system, and therefore should be optimized. If the bandwidth is too wide, the SNR is reduced due to more additional Gaussian noise passing through the filter. On the other hand, if it is too narrow, ISI is produced due to severe group delay variation within the bandwidth. It has been found in [16] that the optimum 3-dB bandwidth is about $0.55f_s$ or $B_{3\text{dB}}T_s = 0.55$, where $T_s = 1/f_s$ is the symbol duration and $B_{3\text{dB}}$ is a 3 dB bandwidth.

One effective way to reduce the ISI in a narrow-bandwidth case is to add a group delay equalizer after the receive lowpass filter. Thus, the narrow receive filter cascaded with the group delay equalizer can greatly attenuate the noise and meanwhile significantly reduce ISI. Figure 2.25 shows the recovered eye diagrams of SQAM signal with $A = 0.8$ at the output of the receive lowpass filter through a nonlinear channel. It is noted that ISI is reduced after the group delay equalizer and eye diagrams open widely, where the second-order group delay equalizer is used.

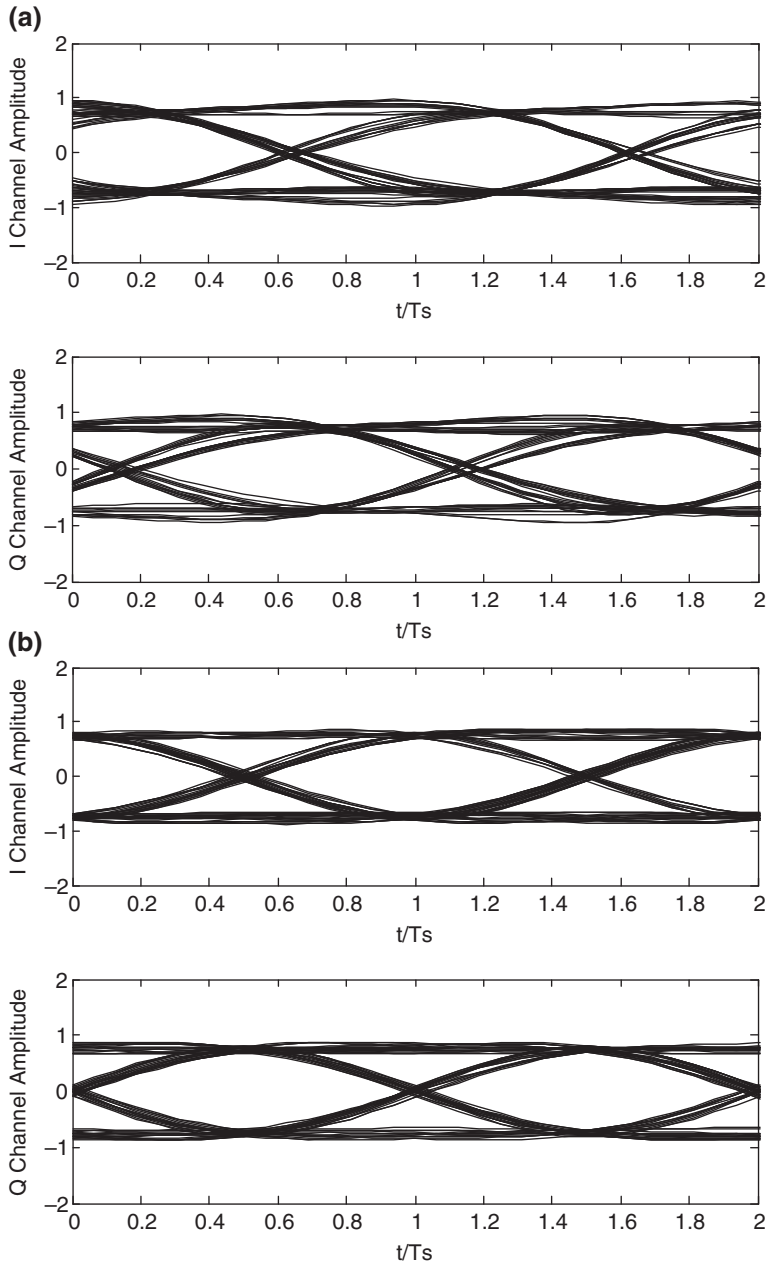


Fig. 2.25 Simulated eye diagrams at the output of the fourth-order Butterworth filter with $BT_s=0.55$ through a nonlinear (hard-limited) channel. (a) SQAM with $A=0.8$ and without group delay equalizer, (b) SQAM with $A=0.8$ and with group delay equalizer

To simulate the BER performance of the SQAM system in a nonlinear channel, we use an ideal hard-limiter to simulate the effect of the nonlinear amplification on the SQAM signal. The ideal hard-limiter represents a good first-order approximation of the saturated high-power amplifier [15]. Figure 2.26 shows the simulated BER performance results of the SQAM signal through a nonlinear channel, where the parameter $A = 0.8$ is used for the SQAM signal due to its best BER performance, as shown in Fig. 2.22. It is also shown in Fig. 2.26 that both SQAM and SCMSK signals have the same BER performance through a nonlinear channel, where the BER performance is degraded by 0.5 dB in the case where no group delay equalizer is cascaded with the Butterworth lowpass filter, and by 0.3 dB in the case where the second-order group delay equalizer is cascaded with the Butterworth lowpass filter. A 0.2 dB improvement is achieved with the group delay equalizer.

2.6 Minimum Bandwidth and ISI-free Nyquist Pulse Shaping

In practical applications, most communication channels are band-limited to some specified bandwidth $2B_w$ Hz for each user in order to efficiently utilize the total available channel bandwidth W . For this case, the channel bandwidth can be equivalent to its baseband bandwidth B_w , or its equivalent lowpass frequency response is non-zero for $|f| \leq B_w$ and zero for $|f| > B_w$. When such a channel is ideal (its amplitude and group delay responses are both constant) for $|f| \leq B_w$, what is the maximum data rate that can be transmitted through the channel without causing intersymbol interference (ISI) in the desired channel and adjacent channel interference (ACI) in the adjacent channels? If such a channel is not ideal for $|f| \leq B_w$, how do we design a compensator or equalizer to compensate for the channel such that the channel impairments can be minimized at the output of the compensator?

The answer to the first question is to design a spectrum shaping pulse that satisfies the *Nyquist criterion* or zero ISI. The solution for the second problem is to design an equalizer to minimize ISI, including the channel impairments caused by both amplitude and phase distortions.

2.6.1 Nyquist Minimum Transmission Bandwidth with ISI-Free

The higher the transmission data rate is, the wider channel bandwidth the transmission system occupies. Nyquist [21] investigated the relationship between the theoretical minimum transmission bandwidth and ISI. He demonstrated that if synchronous *impulse streams* at a transmission rate of f_s symbols per second are

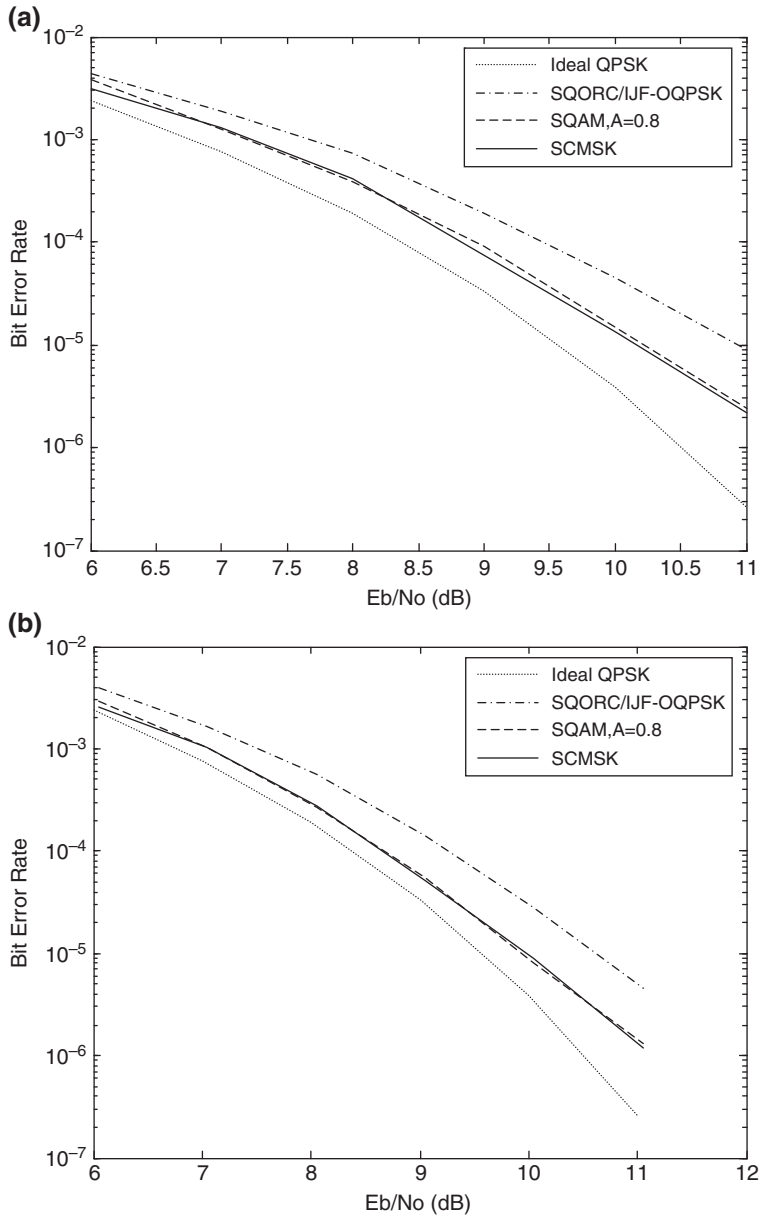


Fig. 2.26 Bit error rate performance of SQORC/IJF-OQPSK, SQAM, and SCMSK signals in a nonlinear channel: (a) without group delay equalizer, and (b) with group delay equalizer. (Note: 0.3 dB degradation for SQAM/SCMSK and 0.9 dB for IJF-OQPSK at 10^{-4})

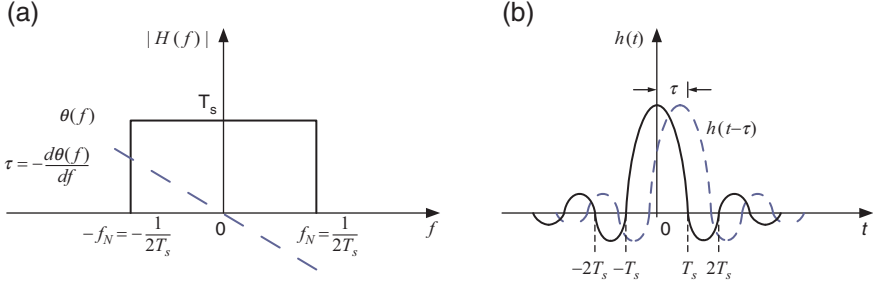


Fig. 2.27 Nyquist channel and its impulse response: (a) ideal brick-wall (Nyquist channel) transfer function, and (b) impulse response

transmitted through a lowpass channel filter with the minimum bandwidth $f_N = f_s/2$ Hz, whose transfer function is characterized by an ideal brick-wall amplitude response and linear phase response, then the output signals of such a channel filter can be detected independently without ISI at the detection instants. The minimum bandwidth $f_N = f_s/2$ is called *Nyquist minimum transmission bandwidth* or *Nyquist frequency*.

Such an ideal lowpass channel transfer function $H(f)$ is in the shape of a brick-wall, as shown in Fig. 2.27a. For baseband transmission systems when the amplitude and phase responses of the overall transfer function $H(f)$ are the same as the one shown in Fig. 2.27a, the transfer function $H(f)$ and its impulse response $h(t)$ are derived as follows:

$$H(f) = \begin{cases} T_s, & |f| \leq f_N = \frac{1}{2T_s} \\ 0, & |f| > f_N = \frac{1}{2T_s} \end{cases} \quad (2.73)$$

$$\begin{aligned} h(t) &= F^{-1}[H(f)] = \int_{-T_s/2}^{T_s/2} T_s e^{j2\pi ft} dt \\ &= \frac{\sin\left(\frac{\pi t}{T_s}\right)}{\frac{\pi t}{T_s}} = \text{sinc}\left(\frac{\pi t}{T_s}\right) \end{aligned} \quad (2.74)$$

where $f_N = 1/(2T_s) = f_s/2$ is known as *Nyquist frequency* and is also equal to the cut-off frequency of the channel filter, where T_s is the symbol duration and f_s is the symbol rate. In binary transmission systems, the symbol duration T_s is equal to the bit duration T_b or $T_s = T_b$. In multilevel M or multi-state M transmission systems, the symbol duration is calculated by $T_s = T_b \log_2 M$. For example, in the QPSK or 4-QAM modulation, we have $M = 4$, and $T_s = 2T_b$.

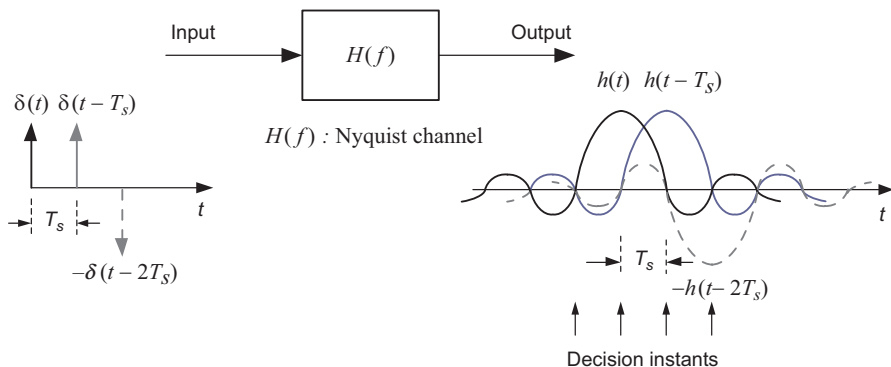


Fig. 2.28 Concept of Nyquist channel achieving zero ISI transmission for input impulse streams

In the derivation above, a zero linear phase for the filter $H(f)$ is assumed. For each input impulse, the impulse response $h(t)$ can be calculated from (2.74) and has its values at the integer multiples of the symbol interval of T_s below

$$h(nT_s) = \begin{cases} 1, & \text{for } n = 0 \\ 0, & \text{for } n = \pm 1, \pm 2, \pm 3, \dots \end{cases} \quad (2.75)$$

Thus, if each impulse response at the output of the received channel filter is of the form as given in (2.75), the received sequences can be detected without ISI. If the ideal Nyquist filter has non-zero phase but linear phase (*dashed line* in Fig. 2.27), the impulse response is shifted by an amount of the filter delay, which is equal to $\tau = -d\phi(f)/df$ and is constant over all frequencies. In this case, the output impulse responses still meet the Nyquist criterion because all impulse streams are delayed by the same delay τ .

When a transmission channel meets the criterion of a Nyquist channel, or the transmission symbol rate of f_s is less than and equal to twice the Nyquist frequency of f_N ($f_s \leq 2f_N$), the outputs of the transmission channel are ISI-free at all decision-making instants for the inputs of impulse streams with the rate of $f_s = 1/T_s$, as shown in Fig. 2.28. On the other hand, if the transmission symbol rate is greater than twice the Nyquist frequency, or $f_s > 2f_N$, the outputs of the transmission channel have ISI at these decision instants.

The names “Nyquist filter” and “brick-wall filter” are often used as alternatives to describe the Nyquist channel for satisfying ISI-free transmission at the decision instants. The impulse response of the Nyquist filter is also called “Nyquist pulse”.

It should be noted that the inputs to the Nyquist filter are impulse streams in order to satisfy ISI-free transmission. However, in most communication systems the inputs are rectangular (or NRZ) pulses instead of impulse streams. For these rectangular-pulse inputs to satisfy ISI-free transmission, a $x/\sin(x)$ -shaped amplitude compensator has to be cascaded with the Nyquist filter. The frequency response of the cascaded amplitude compensation with the Nyquist filter is shown in Fig. 2.29. The reason for that is explained as follows:

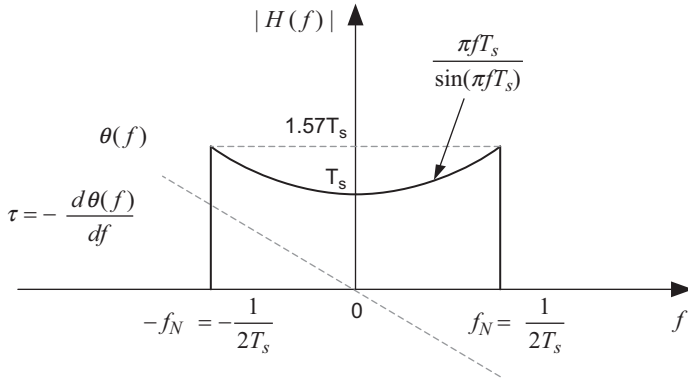


Fig. 2.29 Frequency response of Nyquist channel for rectangular pulse transmission with amplitude compensation

For an impulse input, its Fourier transform is constant over all frequencies. Thus, the output of the Nyquist filter is simply the inverse Fourier transform of the Nyquist filter. For a rectangular-pulse input, its Fourier transform is the form of $\sin(x)/x$. To satisfy the Nyquist ISI-free transmission, the rectangular pulse should be transferred into the impulse before entering the Nyquist filter. If such a rectangular pulse is passed with an amplitude compensator with the transfer function $x/\sin(x)$, then the output of the amplitude compensator becomes an impulse and now satisfies ISI-free transmission.

A practical approach to the amplitude compensation is to use a second-order lowpass with the damping factor less than 0.707 so that its amplitude response increases with the frequency and approximates the shape of $x/\sin(x)$ within the Nyquist bandwidth f_N . This lowpass filter can be added either before or after the Nyquist filter.

Unfortunately, the ideal Nyquist channel filter with the minimum bandwidth is not realizable, and therefore it cannot be implemented with any hardware circuits or components. In the case of approximating it, the approximated Nyquist channel filter may need a larger-order number of filter. In addition to approximating the amplitude response, it is very difficult to achieve a linear phase filter as requested for the Nyquist channel filter.

Fortunately, a practically and widely used function that satisfies the free-ISI is the raised-cosine filter, whose amplitude response is in the form of the raised-cosine shape. The amplitude response of the raised-cosine (RC) filter is expressed as

$$|H_{rc}(f)| = \begin{cases} T_s & 0 \leq |f| \leq \frac{1-\alpha}{2T_s} \\ T_s \cos^2 \left[\frac{\pi T_s}{2\alpha} \left(|f| - \frac{1-\alpha}{2T_s} \right) \right] & \frac{1-\alpha}{2T_s} \leq |f| \leq \frac{1+\alpha}{2T_s} \\ 0 & |f| > \frac{1+\alpha}{2T_s} \end{cases} \quad (2.76)$$

where α is the roll-off factor and determines the raised-cosine filter shape and bandwidth. For $\alpha = 0$, an ideal brick-wall filter having a minimum bandwidth equal to the Nyquist frequency of $f_N = 1/2T_s$ is achieved. However, this brick-wall filter cannot be realized in practice. In practical applications, values between $0.2 \leq \alpha \leq 1.0$ are usually chosen. The Nyquist frequency f_N is also called the *excess bandwidth* and is expressed as a percentage of f_N . For example, $\alpha = 0.35$ corresponds to excess bandwidth of 35%, while $\alpha = 1$ corresponds to excess bandwidth of 100%.

The impulse response $h_{rc}(t)$ or the inverse Fourier transform (2.76) is

$$h_{rc}(t) = \frac{\sin(\pi t/T_s)}{\pi t/T_s} \frac{\cos(\pi \alpha t/T_s)}{1 - 4\alpha^2 t^2/T_s^2} \quad (2.77)$$

Figure 2.30 illustrates the amplitude response of the raised-cosine filter and the corresponding impulses for $\alpha = 0, 0.3$, and 1. The single-side bandwidth is equal to $f = (1 + \alpha)f_N$. It is clearly seen that the tails of $h_{rc}(t)$ decay faster with the increase of α value. The lower the tail of $h_{rc}(t)$ decays, the bigger the ISI is at the decision instants with severe phase jitter.

The expression of the raised-cosine filter (2.76) can also be rewritten

$$|H_{rc}(f)| = \begin{cases} T_s & 0 \leq |f| \leq \frac{1-\alpha}{2T_s} \\ \frac{T_s}{2} \left\{ 1 + \cos \left[\frac{\pi T_s}{\alpha} \left(|f| - \frac{1-\alpha}{2T_s} \right) \right] \right\} & \frac{1-\alpha}{2T_s} \leq |f| \leq \frac{1+\alpha}{2T_s} \\ 0 & |f| > \frac{1+\alpha}{2T_s} \end{cases} \quad (2.78)$$

If the input to the raised-cosine filter is the rectangular (NRZ) pulse streams, a $x/\sin(x)$ -shaped amplitude compensator must be cascaded with the raised-cosine filter. For a rectangular pulse $g(t)$ with the interval T_s and amplitude $1/T_s$, its Fourier transform is

$$\begin{aligned} G(f) &= \int_{-T_s/2}^{T_s/2} g(t) e^{-j2\pi f t} dt = \frac{1}{T_s} \int_{-T_s/2}^{T_s/2} e^{-j2\pi f t} dt \\ &= \frac{\sin(\pi f T_s)}{\pi f T_s} \end{aligned} \quad (2.79)$$

Thus, the amplitude response of the amplitude compensator expressed as the inverse of $G(f)$ in (2.79) is added to the raised-cosine filter (2.78) to form the cascaded amplitude response for ISI-free pulse transmission:

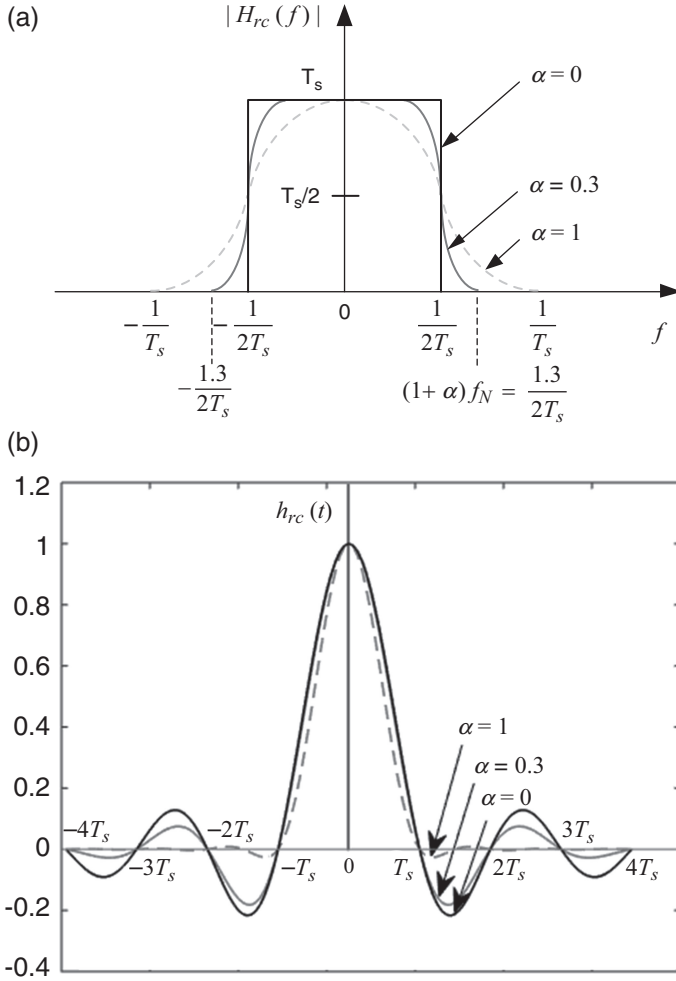


Fig. 2.30 Raised-cosine filter characteristics: (a) amplitude response, and (b) impulse response

$$|H(f)| = \begin{cases} \frac{\pi f T_s}{\sin(\pi f T_s)} T_s & 0 \leq |f| \leq \frac{1 - \alpha}{2T_s} \\ \frac{\pi f T_s}{\sin(\pi f T_s)} \frac{T_s}{2} \left\{ 1 + \cos \left[\frac{\pi T_s}{\alpha} \left(|f| - \frac{1 - \alpha}{2T_s} \right) \right] \right\} & \frac{1 - \alpha}{2T_s} \leq |f| \leq \frac{1 + \alpha}{2T_s} \\ 0 & |f| > \frac{1 + \alpha}{2T_s} \end{cases} \quad (2.80)$$

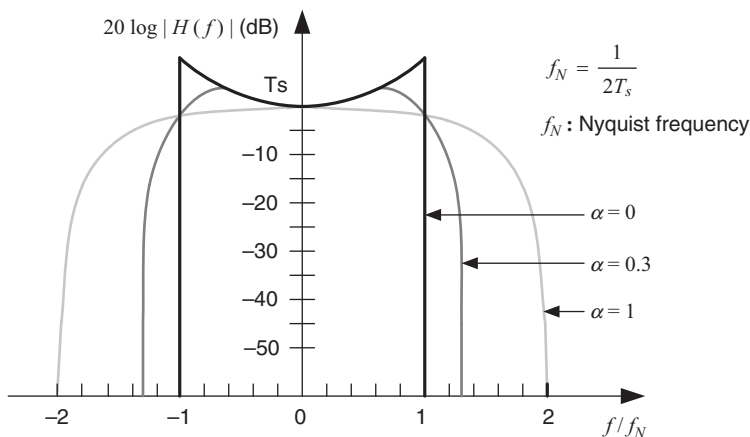


Fig. 2.31 Amplitude response of raised-cosine filter cascaded with $x/\sin(x)$ shaped amplitude compensator for rectangular pulse ISI-free transmission. (Note: Attenuation is $4 - 6 = -2$ dB at $f/f_N = 1$ for $\alpha \neq 0$)

Figure 2.31 illustrates the amplitude response of the cascaded filter with different α values. For $\alpha = 0$, an unrealizable minimum-bandwidth filter or Nyquist filter is obtained, as shown by the dark black line. Theoretically, the attenuation has an infinite value beyond the bandwidth. Practically, attenuation might be in the range from 20 to 50 dB, depending on implementation approach. In an analog filter design approximation, the attenuation level outside the channel should not be designed too large due to severe group delay variation of the analog filter with high order. In a digital FIR filter design approximation, however, the attenuation level can be large by increasing the number of FIR filter coefficients because of its linear property of the group delay.

The simulated eye diagrams for NRZ data filtered by a raised-cosine filter either with $x/\sin(x)$ -shaped amplitude aperture compensation or without it are illustrated in Fig. 2.32. In the case with $x/\sin(x)$ compensation (Fig. 2.32a,b) eye diagrams are ISI-free at the maximum opening instants. In the case without $x/\sin(x)$ compensation (Fig. 2.32c), the eye diagram has ISI at the maximum opening instants. Therefore, the amplitude aperture compensator is necessary for rectangular (NRZ) pulse streams to achieve ISI-free transmission. It is noted that for $\alpha = 1$ in Fig. 2.32a, the eye diagram has zero data polarity transmission jitter at $t/T_s = 0.5$ or 1.5 . With $\alpha = 0.3$ there is a significant data polarity transmission jitter, which results in severe timing jitter in the recovered timing-clock signal. Furthermore, large tails corresponding to small α values exhibit more sensitivity to timing errors and thus result in severe BER degradation due to ISI. On the other hand, a large α value will result in a large excess bandwidth.

Considering that the overall channel filter consists of a transmitter filter and a receiver filter, the raised-cosine filter in (2.76) can be split into two parts:

$$H_{rc}(f) = H_t(f)H_r(f) \quad (2.81)$$

If the receiver filter is matched to the transmitter filter or $H_r(f) = H_t(f)$, the desired magnitude response of the filter, called the *square root of raised-cosine (SRRC)* filter, is obtained:

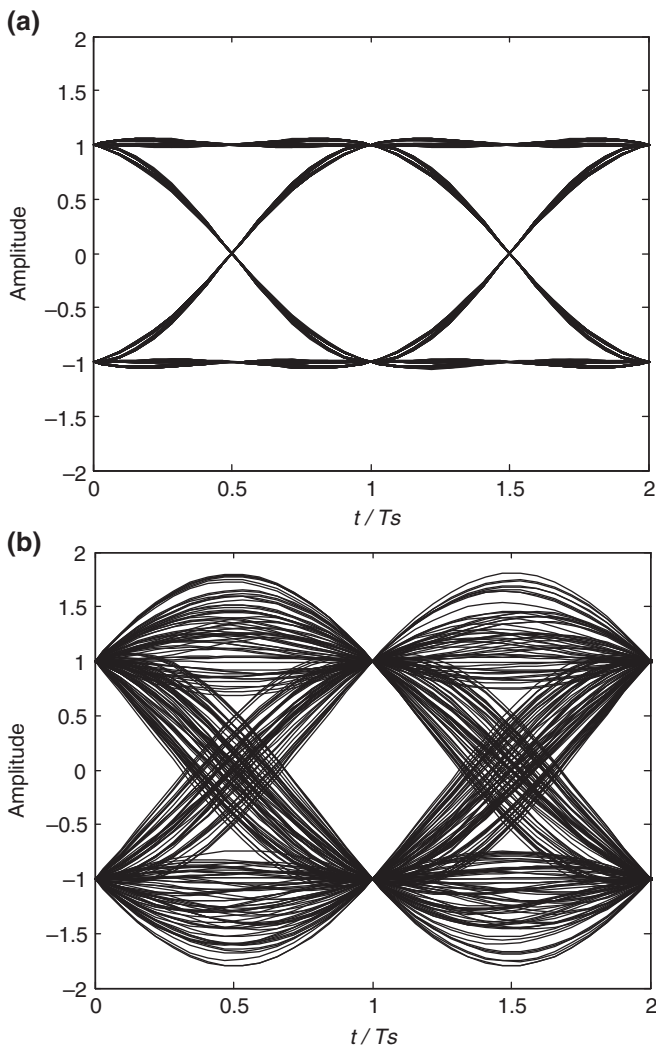


Fig. 2.32 Eye diagrams of NRZ data filtered by raised-cosine filter: (a) roll-off factor $\alpha = 1$ with $x/\sin(x)$, (b) roll-off factor $\alpha = 0.3$ with $x/\sin(x)$ and (c) roll-off factor $\alpha = 1$ without $x/\sin(x)$

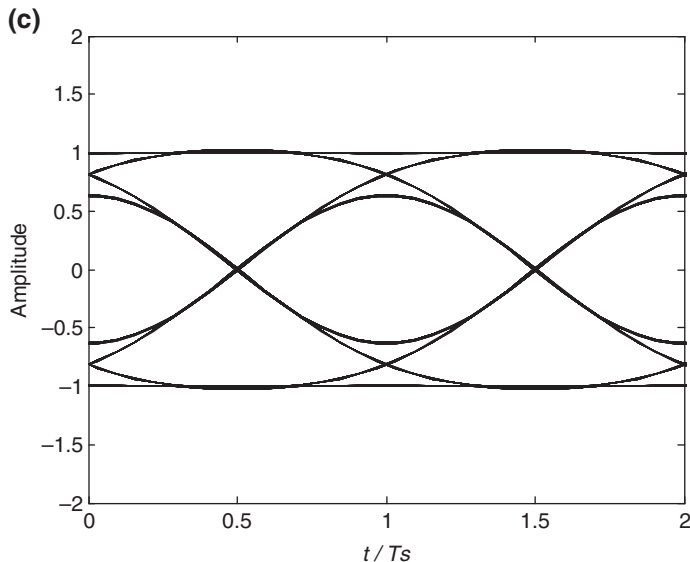


Fig. 2.32 (continued)

$$\begin{aligned}
 |H_t(f)| &= |H_r(f)| = H_{\text{srrc}}(f) = \sqrt{|H_{\text{rc}}(f)|} \\
 &= \begin{cases} \sqrt{T_s} & 0 \leq |f| \leq \frac{1-\alpha}{2T_s} \\ \sqrt{T_s} \cos \left[\frac{\pi T_s}{2\alpha} \left(|f| - \frac{1-\alpha}{2T_s} \right) \right] & \frac{1-\alpha}{2T_s} \leq |f| \leq \frac{1+\alpha}{2T_s} \\ 0 & |f| > \frac{1+\alpha}{2T_s} \end{cases} \quad (2.82)
 \end{aligned}$$

The abbreviation $\sqrt{\alpha}$ is used to stand for SRRC in some literature. In this case, the receiver filter $H_r(f)$ is called the *match filter* to the transmitter filter $H_t(f)$. If the input data to the transmitter filter $H_t(f)$ is NRZ data, the amplitude aperture compensator with the form $x/\sin(x)$ should be cascaded with $H_t(f)$ together for ISI-free transmission. It should be noted that the amplitude aperture compensator $x/\sin(x)$ is unnecessarily needed in the digital FIR filter designs for either a RC filter or a SRRC filter due to the input of the approximate impulse streams after performing up-sampling rate of N by inserting $N - 1$ zero between two adjacent data sequences.

In practical designs, considering the physical realization of the filter, (2.82) is usually written as

$$H_t(f) = \sqrt{|H_{\text{rc}}(f)|} e^{-j2\pi f \tau} \quad (2.83)$$

where τ is some certain constant delay that ensures the peaks of the impulse response $h_t(t) = F^{-1}[H_t(f)]$ occurring after $t \geq 0$.

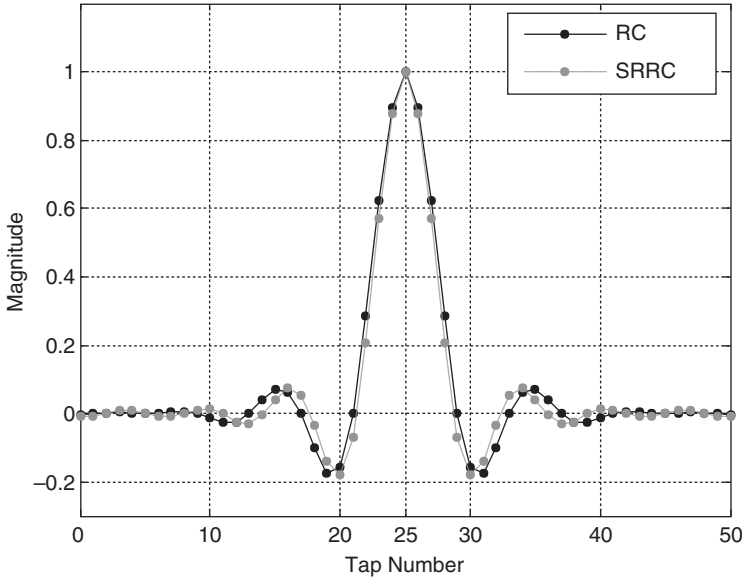


Fig. 2.33 Normalized impulse responses of raised-cosine filter and square root raised-cosine filter with $\alpha = 0.3$

The impulse response of the SRRC filter is obtained by taking the inverse Fourier transform of (2.82), which is expressed as

$$h_{\text{srcc}}(t) = \frac{1}{\sqrt{T_s}} \frac{1}{1 - (4\alpha t/T_s)^2} \left\{ \frac{\sin[(1 - \alpha)\pi t/T_s]}{\pi t/T_s} + \frac{4\alpha \cos[(1 + \alpha)\pi t/T_s]}{\pi} \right\} \quad (2.84)$$

The normalized impulse responses of the RC and SRRC filters are depicted in Fig. 2.33, where four samples per symbol duration T_s are marked by dots. It is clear that the impulse response of the SRRC filter has non-zero crossing at every four sample instants starting from the middle peak instant, while that of the RC filter has zero crossing at these instants. At these sampling instants, which are also called decision-making instants, other adjacent signals reach their peak values. Thus, the SRRC-filtered signal at the transmitter has ISI, while the signal at the output of the matched filter or another SRRC filter at the receiver is ISI-free due to the composite full response of the overall system.

2.6.2 Analog Filter Approximation to SRRC Filter

Analog filters play an important role in spectrally efficient transmission systems. To achieve spectrally efficient transmission, some communication systems use the

analog filter to approximate a raised-cosine filter as the pulse-shaping filter at the transmitter or the matched filter at the receiver due to its simple structure and low cost. This was especially true in the before 1990, when digital signal-processing chips were not affordable, available, or practical due to their high cost and speed limitations. For example, in the late 1980s most digital microwave transmission systems and digital satellite/earth station transmission systems were equipped with analog raised-cosine filters to achieve spectrally efficient transmission. With the development of digital signal-processing technologies over the past two decades, the raised-cosine filter in modern digital communications can be easily realized by a digital FIR filter with linear phase or constant group delay within a 3-dB corner frequency. However, analog filters are still needed to filter out out-band noise, interference signals, and image components, such as a reconstruction filter after a digital-to-analog converter at the transmitter, and a channel selection filter or an anti-aliasing filter before the analog-to-digital converter at the receiver. In today's wireless communication systems, such as 2G GSM and 3G wideband code-division multiple access (WCDMA) systems, the analog filter still has an advantage over the digital filter in wireless handset devices due to its low cost and low current consumption. For example, in WCDMA systems, a fifth-order analog filter is used at the receiver to achieve functionality in both channel selection filtering and approximately matched filtering [22].

In this section, we introduce one example for practical analog filter design approximation to the amplitude response of SRRC filter. Due to the non-constant group delay characteristics of the analog filter, an allpass filter as a group delay equalizer needs to be cascaded with the analog filter to reduce group delay variation within the required bandwidth. This analog approach can meet the need for low-cost, low-power consumption, and low-form-factor user equipment (UE). For detailed design procedures of group delay equalization, the interested reader can refer to Appendix D.

2.6.2.1 Amplitude Approximation

In a white Gaussian noise (WGN) channel, either a RC or a SRRC filter is used to achieve spectrum shaping and ISI-free transmission in band-limited channels. Previously (from the 1990s to the 2000s), the RC and SRRC filters were mostly approximated by using Butterworth and Chebyshev analog filters, or other types of analog filters whose parameters are optimized by computer to minimize the amplitude error between the amplitude response of the analog filter and that of the target RC or SRRC filter. Even today, in the 3GPP WCDMA system, the analog filter is still used to approximate the SRRC filter as part of the channel selection filtering at the receiver, matched to the transmitter-side SRRC filter to achieve the minimal ISI and the maximum adjacent channel interference (ACI) rejection [22–24].

When the analog filter is used to approximate the SRRC filter, the attenuation requirement of the analog filter at the transmitter is determined by the adjacent channel power ratio (ACPR), which is the ratio of the average power in the main

channel and adjacent channels, while the attenuation requirements of the analog filter at the receiver are decided by both in-channel SNR and adjacent channel interference (ACI). The attenuation of the analog filter, on the other hand, is primarily determined by filter prototype, 3-dB corner frequency, and filter order. Four common filter prototypes are Butterworth, Chebyshev, Inverse Chebyshev, and elliptic. In terms of a given order, the larger the attenuation of the filter, the worse the group delay variation of the filter. The elliptic filter provides the largest attenuation and a sharper transition region among the four types of filters but the worst group delay variation within a 3-dB corner frequency. The Butterworth filter shows the smallest group delay variation, but the smallest attenuation as well.

The 3-dB corner frequency of the analog filter is generally determined by the Nyquist frequency f_N or the half-symbol rate $f_s/2$ as described in Sect. 2.6.1. An actual 3-dB corner frequency can be obtained through an approximation to a SRRC filter with a specified roll-off factor, corresponding to a certain symbol rate. A design example at the transmitter will be introduced in the following section while the other one at the receiver will be illustrated in Sect. 7.5.2.

2.6.2.2 Group Delay Compensation

In general, due to natural characteristics of the large group delay fluctuation of the analog filter within the 3-dB corner frequency, an allpass filter needs to be cascaded to reduce group delay fluctuation as much as possible; meanwhile it does not change magnitude response of the analog filter. Because of the spectral-efficiency requirement, most communication channels are usually characterized as band-limited linear filters if all active circuits operate at linear regions. Then, the transfer functions of such channels may be expressed by their frequency response

$$H_c(j\omega) = |H_c(j\omega)|e^{j\theta_c(\omega)} \quad (2.85)$$

where $|H_c(j\omega)|$ is the amplitude response and $\theta_c(\omega)$ is the phase response. Another characteristic, which is more useful to describe the channel's behavior, is the group delay, which is obtained from the negative derivative of the phase, or

$$\text{GD}_c(\omega) = -\frac{d\theta_c(\omega)}{d\omega} \quad (2.86)$$

A channel is said to be ideal if $|H_c(j\omega)|$ and $\text{GD}_c(\omega)$ both are the constant within the transmitted signal bandwidth. Otherwise, the transmitted signal may be distorted through such a channel, depending on how much worse $|H_c(j\omega)|$ and $\text{GD}_c(\omega)$ are. The signal distortion caused by non-ideal $|H_c(j\omega)|$ is called *amplitude distortion*, and that caused by non-ideal $\text{GD}_c(\omega)$ is called *delay distortion*.

The group delay equalizer or compensator can only compensate for the delay distortion by introducing extra delay in such a way that the overall group delay

variation is minimized as much as possible. Suppose the transfer function of the equalizer is expressed by its frequency response

$$H_e(j\omega) = |H_e(j\omega)|e^{j\theta_e(\omega)} \quad (2.87)$$

To compensate for the group delay of $H_c(j\omega)$, the equalizer needs to be cascaded with $H_e(j\omega)$. Thus, the overall transfer function $H(j\omega)$ is

$$\begin{aligned} H(j\omega) &= H_c(j\omega) \times H_e(j\omega) \\ &= |H(j\omega)|e^{j\theta(\omega)} \end{aligned} \quad (2.88)$$

where the amplitude and phase responses are expressed by

$$\begin{aligned} |H(j\omega)| &= |H_c(j\omega)| \times |H_e(j\omega)| \\ &= |H_c(j\omega)| \end{aligned} \quad (2.89)$$

and

$$\theta(\omega) = \theta_c(\omega) + \theta_e(\omega) \quad (2.90)$$

In (2.89), the amplitude response of the group delay equalizer is assumed to be a constant and is normalized to 1. The group delay of $H(j\omega)$ is given by

$$\begin{aligned} \text{GD}(\omega) &= -\frac{d\theta(\omega)}{d\omega} = -\frac{d\theta_c(\omega)}{d\omega} - \frac{d\theta_e(\omega)}{d\omega} \\ &= \text{GD}_c(\omega) + \text{GD}_e(\omega) \end{aligned} \quad (2.91)$$

where $\text{GD}_e(\omega)$ is the group delay of the equalizer. For an ideal case, $\text{GD}(\omega)$ is the constant or nearly constant $\text{GD}(\omega) \approx C$ within the interested frequency band. To achieve such a nearly constant group delay, computer optimization programs can be used to minimize the peak-to-peak variation of the overall group delay within the interested bandwidth B_w or

$$\Delta\text{GD}_{\text{pp}}(\omega) = \text{GD}_{\text{MAX}}(\omega) - \text{GD}_{\text{MIN}}(\omega), \quad \omega \leq B_w \quad (2.92)$$

The question is how small $\Delta\text{GD}_{\text{pp}}(\omega)$ should be. Usually, $\Delta\text{GD}_{\text{pp}}(\omega)$ is determined by the requirement of EVM tolerance at the receiver. In general, a first-order allpass filter or a second-order allpass filter is used as a fundamental unit of the group delay equalizer. A higher-order group delay equalizer can be constructed by cascading such multiple fundamental sections together, in which the first-order section is needed to achieve the odd order of the equalizer.

A characteristic analysis of the first-order and second-order allpass filter sections as a group delay equalizer is introduced in Appendix D.

We start with an analog filter design approximation to a SRRC filter in the transmitter. To achieve ISI-free transmission, the analog filter needs not only to

approximate the amplitude response of the SRRC filter, but also to have small group delay fluctuation. Depending on practical applications, a filter with small group delay fluctuation can be created by cascading a one-stage or multiple-stage group delay equalizer with the approximated SRRC filter. Designing a group delay equalizer may be complicated compared with the design of the filter because several parameters in the equalizer need to be carefully adjusted in order to achieve smaller group delay variation.

Design Example 2.1 In an SCPC satellite earth station system, data are transmitted through a QPSK modulation format at the rate of 64 kbps. A SRRC filter with $\alpha = 0.5$ is used either at the transmitter to perform spectrum-shaping transmission in a limited bandwidth of 45 kHz or at the receiver to match the SRRC filter of the transmitter to attenuate outside channel Gaussian noise and adjacent channel interferers as well as to minimize the in-channel ISI before the decision. Design an analog lowpass at the transmitter to approximate such an ideal SRRC filter without a shape of $x/\sin(x)$ as amplitude compensation. If it is needed, a second-order lowpass filter with a *damping factor* $\zeta < 0.707$ as an amplitude compensation shape of $x/\sin(x)$ can be cascaded with the SRRC filter, which is described in Sect. 2.6.4. A digital SRRC filter is assumed in the receiver to match the SRRC filter of the transmitter.

Solution To achieve such an approximation to a SRRC filter, we choose a fourth-order Butterworth lowpass filter due to its mild group delay characteristic to approximate the amplitude response of the SRRC filter with $\alpha = 0.5$ and a second-order allpass filter as a group delay equalizer to compensate for the group delay fluctuation of the Butterworth lowpass filter in order to achieve small ISI as much as possible.

For the QPSK signal transmission at the bit rate $f_b = 64$ kbps, the Nyquist frequency f_N is equal to 16 kHz due to $f_N = f_s/2 = f_b/4$. After approximation to the amplitude response of the SRRC with $\alpha = 0.5$ and $f_N = 16$ kHz, the fourth-order Butterworth lowpass with the cut-off frequency of 17.2 kHz can achieve the best amplitude approximation, as shown in Fig. 2.34a.

The analog filter closely approximates the ideal SRRC filter down to the 10-dB attenuation point. Beyond -10 dB, distortion caused by approximation errors may be ignored in practice, depending on the requirement of the desired signal leakage to the adjacent channels. To achieve smaller group delay variation, a second-order allpass filter is chosen as the group delay equalizer; its group delay response is shown in Fig. 2.34b. The group delay fluctuation of the fourth-order Butterworth lowpass filter is significantly reduced from 12 μ s to less than 1 μ s within the cutoff frequency of 17.2 kHz after the equalizer. For the detailed design procedure and circuit implementation, the interested reader can again refer to Appendix D.

Figure 2.35 illustrates the simulated eye diagram at the output of the RX digital SRRC filter, which is matched to the TX analog approximation SRRC filter designed above. It can be seen that the received signal has very small ISI at the decision-making instants. Therefore, such an analog approximation to the SRRC filter is satisfied in a low-cost and low-power consumption application.

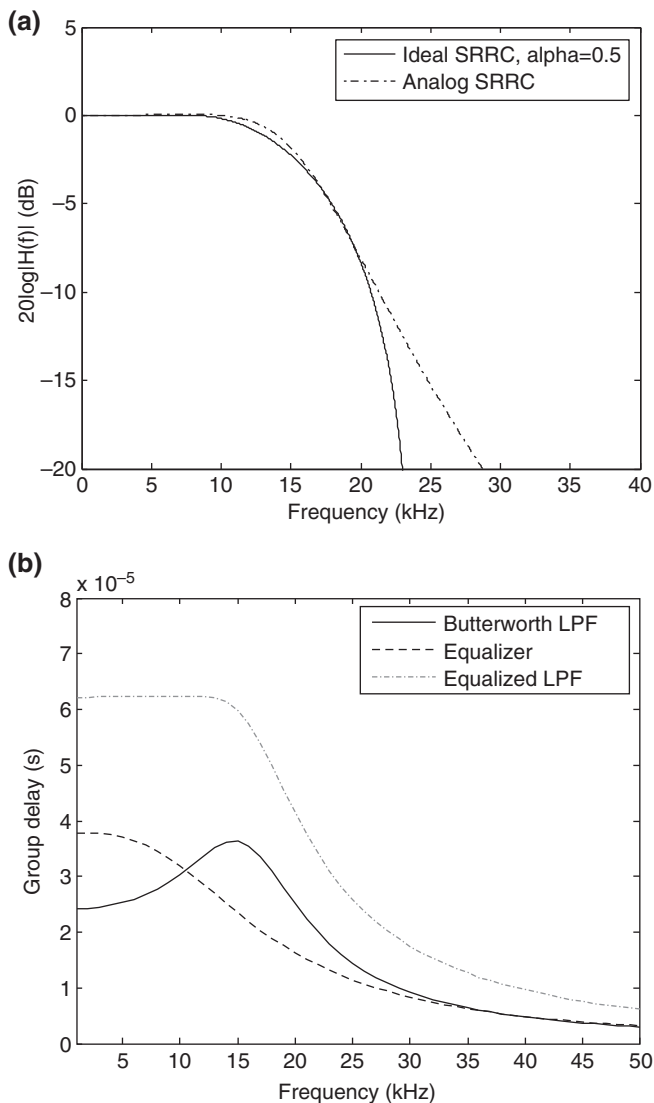


Fig. 2.34 Frequency response of the fourth-order analog filter approximation to a SRRC filter with $\alpha = 0.5$: (a) amplitude response and (b) group delay response

2.6.3 Digital Filter Approximation to Raised-Cosine Filter

In digital communication systems, the strictly band-limited and ISI-free requirements of a wireless communication channel demand the use of a pulse-shaping filter. These requirements are very difficult to meet by using an analog filter approximation to a RC filter plus a group delay equalizer. Therefore, a digital filter

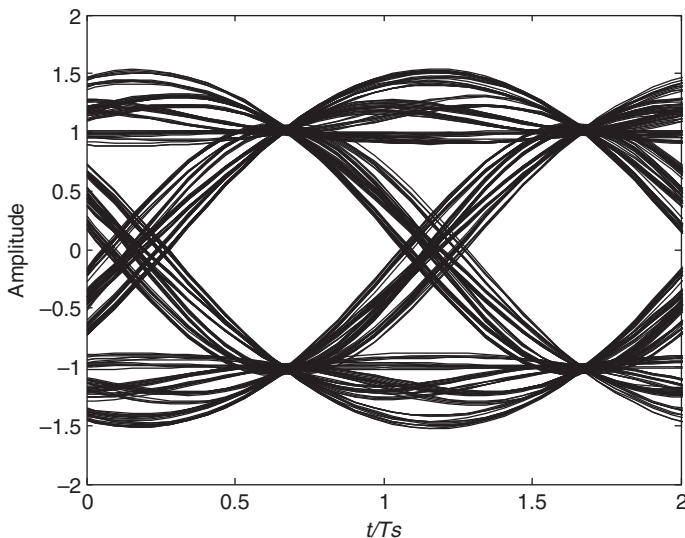


Fig. 2.35 Received eye diagram at the output of the RX digital SRRC filter with $\alpha=0.5$, which is matched to the TX SRRC filter approximated by a fourth-order Butterworth lowpass filter cascaded with a second-order equalizer

approximation to an RC filter is still a dominant design approach due to its accurate approximation to both amplitude and phase characteristics, especially in the transmitter.

From the impulse response of the raised-cosine filter in time domain, it can be seen that the impulse response has an *infinite duration*, even though it decays very little as time increases. Hence, it is impossible to implement either a RC or SRRC filter with an infinite duration of the impulse response. Also it is unnecessary due to the power assumption and cost. The simplest way to approximate an ideal RC or SRRC filter is to design a FIR filter, due to its symmetric impulse responses and linear phase. The coefficients of the FIR filter approximation to a RC or SRRC filter can be obtained from either its impulse response in the time domain or its frequency response in the frequency domain. Some basic design procedures using these methods will be introduced in the following sections.

2.6.3.1 Filter Design by Window

The simplest method of FIR filter design in the time domain is called the *window method*. A basic method to approximate an ideal filter is to truncate the ideal impulse response $h(n)$. After truncation, the designed filter with impulse response $h_d(n)$ is given by

$$h_d(n) = \begin{cases} h(n), & n = 0, \pm 1, \dots, \pm (N-1)/2 \\ 0, & \text{otherwise} \end{cases} \quad (2.93)$$

If a finite duration rectangular window $w(n)$ is used, the impulse response $h_d(n)$ in (2.93) can be expressed as

$$h_d(n) = h(n)w(n) \quad (2.94)$$

where

$$w(n) = \begin{cases} 1, & n = 0, \pm 1, \dots, \pm (N-1)/2 \\ 0, & \text{otherwise} \end{cases} \quad (2.95)$$

Besides the rectangular window, other windows with a less abrupt truncation property may also be used to achieve small side-lobes in the frequency domain. However, this is achieved at the price of a wider main lobe. It is also well known that the *Gibbs phenomenon* can be moderated through the use of such a window with the property of a smooth transition to zero [25]. A useful sinusoid window can be used to truncate the impulse response of the RC or SRRC, and expressed as

$$w(n) = \begin{cases} \sin\left(\frac{\pi[n + (N-1)/2]}{N-1}\right), & n = 0, \pm 1, \dots, \pm (N-1)/2 \\ 0, & \text{otherwise} \end{cases} \quad (2.96)$$

For the rectangular window, we can calculate the impulse response of the digital RC filter from (2.77), or

$$\begin{aligned} h_{rc}(n) &= h_{rc}(t)|_{t=nT_{\text{sam}}} \\ &= \frac{\sin(\pi n T_{\text{sam}}/T_s)}{\pi n T_{\text{sam}}/T_s} \frac{\cos\left(\frac{\pi \alpha n T_{\text{sam}}}{T_s}\right)}{1 - \left(\frac{2\alpha n T_{\text{sam}}}{T_s}\right)^2}, \quad n = 0, \pm 1, \dots, \pm (N-1)/2 \end{aligned} \quad (2.97)$$

where N is the length of the filter. The minimum number of the samples per symbol is 2, corresponding to the sampling duration $T_{\text{sam}} = T_s/2$. Usually four samples per symbol are used in most transmitter filter designs, or $T_{\text{sam}} = T_s/4$.

Similar to the design of $h_{rc}(n)$, the calculation of the impulse response of the digital SRRC filter can be obtained from (2.84):

$$\begin{aligned}
h_{\text{srcc}}(n) &= h_{\text{srcc}}(t)|_{t=nT_{\text{sam}}} \\
&= \frac{1}{\sqrt{T_s}} \frac{1}{1 - (4\alpha nT_{\text{sam}}/T_s)^2} \left\{ \frac{\sin[(1 - \alpha)\pi nT_{\text{sam}}/T_s]}{\pi nT_{\text{sam}}/T_s} \right. \\
&\quad \left. + \frac{4\alpha \cos[(1 + \alpha)\pi nT_{\text{sam}}/T_s]}{\pi} \right\}, \quad n = 0, \pm 1, \dots, \pm (N - 1)/2
\end{aligned} \tag{2.98}$$

For practical implementations, the impulse response of the RC filter or SRRC filter should be delayed by $(N - 1)/2$. Thus, we need to transfer (2.97) and (2.98) to $h_{\text{rc}}(n) = h_{\text{rc}}[n - (N - 1)/2]$ and $h_{\text{srcc}}(n) = h_{\text{srcc}}[n - (N - 1)/2]$, $n = 0, 1, \dots, N - 1$, respectively, after obtaining them.

2.6.3.2 Filter Design by Impulse Invariance

In the filter design based on the concept of *impulse invariance* [25], we know that a discrete-time system can be defined by sampling the impulse response of its corresponding continuous-time system. In the impulse invariance design procedure for a bandlimited system, the impulse response of the discrete-time filter is chosen to be proportional to equally spaced samples of the impulse response of the corresponding continuous-time filter: i.e.,

$$h(n) = T_{\text{sam}} h_c(nT_{\text{sam}}) \tag{2.99}$$

where T_{sam} is a sampling interval. Note that we use T_{sam} as the scaling factor to multiply $h_c(nT_{\text{sam}})$ in order to normalize its frequency response.

In this design we are interested in the relationship between the frequency responses of the discrete-time and continuous-time filters. The frequency response of the discrete-time filter is related to that of its continuous-time filter by

$$H(e^{j\omega}) = \sum_{k=-\infty}^{\infty} H_c[j(\omega - k\omega_{\text{sam}})] \tag{2.100}$$

where $\omega_{\text{sam}} = 2\pi f_{\text{sam}} = 2\pi/T_{\text{sam}}$ is the sampling frequency in radians/s. If the continuous-time filter is bandlimited to ω_B , or

$$H_c(j\omega) = 0, \quad \omega_B \leq |\omega| \leq \pi \tag{2.101}$$

Then, (2.100) becomes

$$H(e^{j\omega}) = H_c(j\omega), \quad |\omega| \leq \omega_B \tag{2.102}$$

or

$$H(e^{j\omega}) = H_c(f) \quad |f| \leq f_B \quad (2.103)$$

If $H_c(f)$ is sampled at equally spaced points in the frequency domain with a frequency step $\Delta f = f_{\text{sam}}/N$, where N is odd number, then (2.100) can be rewritten as

$$H(e^{j\omega}) = H_c(m\Delta f) = H_c(mf_{\text{sam}}/N) \quad (2.104)$$

The relationship between the frequency response and impulse response of the digital raised-cosine (RC) filter with the length of N is given by

$$H_{\text{rc}}(e^{j\omega}) = \sum_{n=-(N-1)/2}^{(N-1)/2} h_{\text{rc}}(n) e^{-j\omega n T_{\text{sam}}} \quad (2.105)$$

Substituting (2.104) into (2.105), we can express (2.105) as

$$H_{\text{rc}}(mf_{\text{sam}}/N) = \sum_{n=-(N-1)/2}^{(N-1)/2} h_{\text{rc}}(n) e^{-j2\pi mn/N} \quad (2.106)$$

Then, the inverse transform of (2.106) is

$$h_{\text{rc}}(n) = \sum_{m=-(N-1)/2}^{(N-1)/2} H_{\text{rc}}\left(\frac{mf_{\text{sam}}}{N}\right) e^{j2\pi mn/N}, \quad n = 0, \pm 1, \dots, \pm (N-1)/2 \quad (2.107)$$

If the RC filter is realized by cascading the transmitter filter with the receiver filter, which is matched to the transmitter filter, each of them is expressed as

$$H_t(f) = H_r(f) = \sqrt{H_{\text{rc}}(f)} \quad (2.108)$$

Similar to the derivation of (2.107), the impulse response of the SRRC filter is obtained by solving

$$\begin{aligned} h_t(n) &= h_r(n) \\ &= \sum_{m=-(N-1)/2}^{(N-1)/2} \sqrt{H_{\text{rc}}\left(\frac{mf_{\text{sam}}}{N}\right)} e^{j2\pi mn/N}, \quad n = 0, \pm 1, \dots, \pm (N-1)/2 \end{aligned} \quad (2.109)$$

2.6.3.3 Digital Design Implementation

Now we give an example by using these two different design methods to design a SRRC filter at the transmitter to achieve spectrally efficient transmission through a linear channel.

Design Example 2.2 Create a digital FIR implementation of the transmitter SRRC filter with $\alpha = 0.3$ and $N = 63$, and implement it in a *Xilinx* chip with an 11-bit fixed point.

Solution In order to simplify hardware design, we choose the sampling rate $f_{\text{sam}} = 4/T_s$, or four samples per symbol interval, making the total length of the FIR filter with $N = 63$ spans $(63 + 1)/4 = 16$ symbols. We both window and impulse invariance methods to design the FIR filter; their impulse responses and frequency responses are illustrated in Fig. 2.36. It can be seen that the impulse responses obtained by using two different design methods are identical, while their frequency responses are very close to each other up to the normalized frequency $f/f_N = (1 + \alpha) = 1.3$. To estimate the design accuracy relative to an ideal filter, we also plot the frequency response of the SRRC FIR filter, with $N = 263$ in Fig. 2.36b, which is used to closely approach such an ideal SRRC filter. Frequency responses of the FIR filters with both window and impulse invariance methods are very close to that of the SRRC FIR filter, with $N = 263$ until -30 dB attenuation. Therefore, using either of these two methods to design the digital SRRC filter can achieve a satisfactory approximation to an ideal SRRC filter.

For a comparison between different window truncation functions, Fig. 2.37 illustrates the frequency responses of the designed SRRC filter with different windows. It is clear that significant small side-lobes are obtained at the price of a slightly wider main lobe.

The coefficients of the impulse response of the SRRC filter calculated from (2.98) and (2.109), respectively, are listed in Table 2.4: each coefficient has two different values; where the first one is obtained by using the window method and the second one is obtained by using the impulse invariance method. Because the impulse response is symmetric, the impulse response values of the transmitter $h_t(n)$ at the positive index are the same as those at the negative index. Therefore, they are not listed except for the coefficients of h_1 and h_{31} .

Assume that 11-bit fixed point, denoted by $C_{10}C_9 \dots C_0$, is used to represent the coefficients in the Xilinx implementation, where C_{10} is a sign bit. After the center coefficient h_0 is normalized to 1023 in decimal, the rest values are listed in Table 2.5. Xilinx Virtex2 Chip has a Coregen function of the FIR filter. By using these coefficients, the FIR Coregen can realize such a SRRC filter.

2.6.4 Amplitude Compensation for a SINC Function

As we discussed in earlier, the $x/\sin(x)$ amplitude compensation is required to cascade with the raised-cosine filter at the transmitter as expressed in (2.80) if the

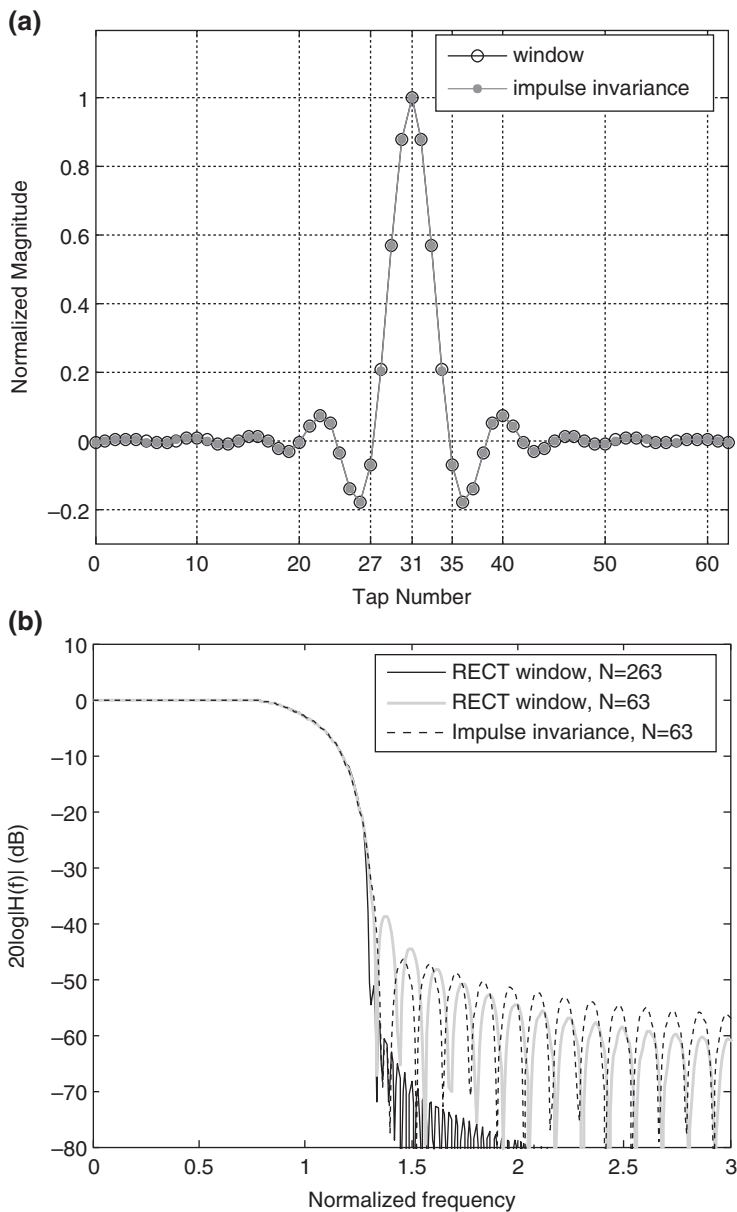


Fig. 2.36 Impulse response and frequency response of FIR filter with a tap length of $N=63$ to approximate to an idea SRRC filter with $\alpha=0.3$: (a) impulse response, and (b) frequency response

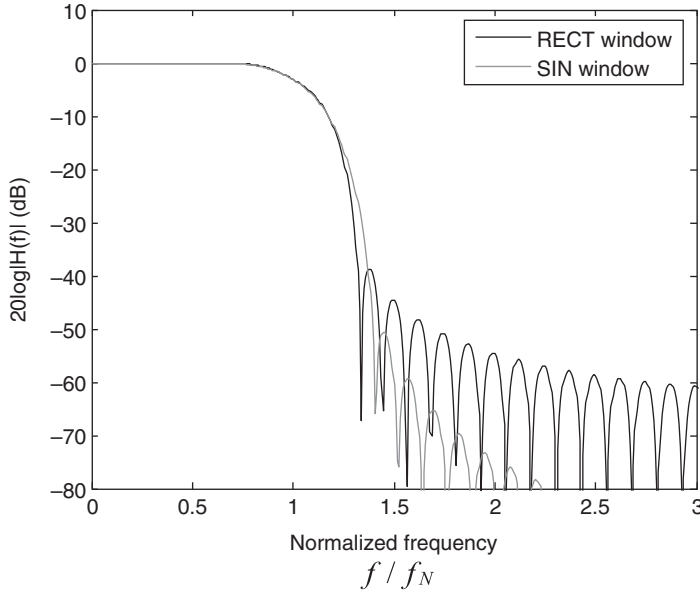


Fig. 2.37 Frequency responses of FIR SRRC filter with $\alpha = 0.3$ and tap length $N = 63$ truncated by rectangular and sinusoid windows

Table 2.4 Coefficients of impulse response of a FIR filter with 63-tap

h_{-31}	h_{-30}	h_{-29}	h_{-28}	h_{-27}	h_{-26}	h_{-25}
0	$6e-4$	$6e-4$	$-7e-4$	$-2.1e-3$	$-1.7e-3$	$1.5e-3$
$-3.6e-3$	$-2.8e-3$	$1.2e-3$	$4.6e-3$	$3.6e-3$	$-1.7e-3$	$-6.7e-3$
h_{-24}	h_{-23}	h_{-22}	h_{-21}	h_{-20}	h_{-19}	h_{-18}
$5e-3$	$5e-3$	$-9e-4$	$-9.4e-3$	$-1.3e-2$	$-5.6e-3$	$1.1e-2$
$-6.2e-3$	$5e-4$	$8.0e-3$	$9.0e-3$	$1.7e-3$	$-7.9e-3$	$-1.01e-2$
h_{-17}	h_{-16}	h_{-15}	h_{-14}	h_{-13}	h_{-12}	h_{-11}
$2.54e-2$	$2.34e-2$	$-1e-4$	$-3.36e-2$	$-5.23e-2$	$-3.47e-2$	$1.85e-2$
$-1.3e-3$	$1.1e-2$	$1.32e-2$	$-1.7e-3$	$-2.37e-2$	$-3.06e-2$	$-5.8e-3$
h_{-10}	h_{-9}	h_{-8}	h_{-7}	h_{-6}	h_{-5}	h_{-4}
$7.74e-2$	$9.64e-2$	$4.46e-2$	$-6.57e-2$	$-1.724e-1$	$-1.889e-1$	$-5.15e-2$
$4.16e-2$	$7.48e-2$	$5.23e-2$	$-3.52e-2$	$-1.42e-1$	$-1.801e-1$	$-6.93e-2$
h_{-3}	h_{-2}	h_{-1}	h_0	h_1	...	h_{31}
$2.364e-1$	$5.924e-1$	$8.863e-1$	1.0	$8.863e-1$	...	0.0
$2.058e-1$	$5.687e-1$	$8.784e-1$	1.0	$8.784e-1$	...	$-3.6e-3$

input to the raised-cosine filter is a NRZ signal. In the digital FIR implementation, however, the amplitude aperture compensator $x/\sin(x)$ may be unnecessarily needed because the input impulse streams are performed with an up-sampling rate of N by inserting $N - 1$ zero between two adjacent data sequences before passing through a digital RC or SRRC filter, especially in the case of $N > 4$.

Table 2.5 Coefficients represented by 11-bit fixed points

h_{-31}	h_{-30}	h_{-29}	h_{-28}	h_{-27}	h_{-26}	h_{-25}
0	1	1	-1	-2	-2	2
-4	-3	1	5	4	-2	-6
h_{-24}	h_{-23}	h_{-22}	h_{-21}	h_{-20}	h_{-19}	h_{-18}
5	5	-1	-10	-13	-6	11
5	1	8	9	2	-8	-11
h_{-17}	h_{-16}	h_{-15}	h_{-14}	h_{-13}	h_{-12}	h_{-11}
26	24	0	-34	-54	-36	19
1	11	14	-2	-26	-31	5
h_{-10}	h_{-9}	h_{-8}	h_{-7}	h_{-6}	h_{-5}	h_{-4}
79	98	46	-67	-176	-193	-53
43	77	54	-36	-145	-184	-71
h_{-3}	h_{-2}	h_{-1}	h_0	h_1	...	h_{31}
242	606	907	1023	907	...	0
211	582	900	1023	900	...	-4

The discrete signal at the output of the RC or SRRC filter is transferred to the continuous signal through a digital-to-analog converter (DAC). If the output of the RC or SRRC filters or the input of the DAC is a sequence of samples, $y_d(n)$, an impulse train $y_i(t)$ to the input of the zero-order hold block can be formed as [25] follows:

$$y_i(t) = \sum_{n=-\infty}^{\infty} y_d(n) \delta(t - nT_{\text{sam}}) \quad (2.110)$$

where T_{sam} is the sampling interval associated with the sequence $y_d(n)$. Thus, the output signal $y_z(t)$ of the zero-order hold block is the convolution of the input $y_i(t)$ of the zero-order hold block and the impulse response $h_z(t)$ of the zero-order hold in the time domain, or

$$\begin{aligned}
 y_z(t) &= y_i(t) * h_z(t) \\
 &= \left(\sum_{n=-\infty}^{\infty} y_d(n) \delta(t - nT_{\text{sam}}) \right) * h_z(t) \\
 &= \sum_{n=-\infty}^{\infty} y_d(n) h_z(t - nT_{\text{sam}})
 \end{aligned} \quad (2.111)$$

The impulse response $h_z(t)$ of the zero-order hold is expressed as

$$h_z(t) = \begin{cases} 1, & 0 < t < T_{\text{sam}} \\ 0, & \text{otherwise} \end{cases} \quad (2.112)$$

Its Fourier transform is

$$H_z(\omega) = \frac{\sin(\omega T_{\text{sam}}/2)}{\omega/2} e^{-j\omega T_{\text{sam}}/2} \quad (2.113)$$

Thus, the Fourier transform of (2.111) becomes

$$\begin{aligned} Y_z(\omega) &= Y_i(\omega) H_z(\omega) \\ &= Y_i(\omega) \frac{\sin(\omega T_{\text{sam}}/2)}{\omega/2} e^{-j\omega T_{\text{sam}}/2} \end{aligned} \quad (2.114)$$

To recover the input signal $Y_i(\omega)$, the output signal $Y_z(\omega)$ of the zero-order hold needs to be passed through a reconstruction filter with a transfer function of $H_r(\omega)$. The Fourier transform of the reconstruction filter output is

$$\begin{aligned} Y(\omega) &= Y_z(\omega) H_r(\omega) \\ &= Y_i(\omega) H_z(\omega) H_r(\omega) \end{aligned} \quad (2.115)$$

It can be seen from (2.115) that the production of the last two items should be equal to the ideal lowpass filter $H_l(\omega)$ in order to recover $Y_i(\omega)$, or

$$H_z(\omega) H_r(\omega) = H_l(\omega) \quad (2.116)$$

$$H_l(\omega) = \begin{cases} T_{\text{sam}}, & |\omega| < \omega_c \\ 0, & \text{otherwise} \end{cases} \quad (2.117)$$

Thus, the reconstruction filter is

$$\begin{aligned} H_r(\omega) &= \frac{H_l(\omega)}{H_z(\omega)} \\ &= H_c(\omega) H_l(\omega) \\ &= \begin{cases} \frac{\omega T_{\text{sam}}/2}{\sin(\omega T_{\text{sam}}/2)} e^{j\omega T_{\text{sam}}/2}, & |\omega| < \omega_c \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (2.118)$$

Note from (2.118) that the reconstruction filter $H_r(\omega)$ is obtained by cascading a compensation filter $H_c(\omega)$ and an ideal lowpass filter $H_l(\omega)$. Its advance time shift of $T_{\text{sam}}/2$ seconds can be compensated if the ideal filter $H_l(\omega)$ has a delay time shift of τ , in which τ meets the condition $\tau \geq T_{\text{sam}}/2$. The compensation filter $H_c(\omega)$ is

$$H_c(\omega) = \frac{1}{T_{\text{sam}}} \frac{\omega T_{\text{sam}}/2}{\sin(\omega T_{\text{sam}}/2)} e^{j\omega T_{\text{sam}}/2} \quad (2.119)$$

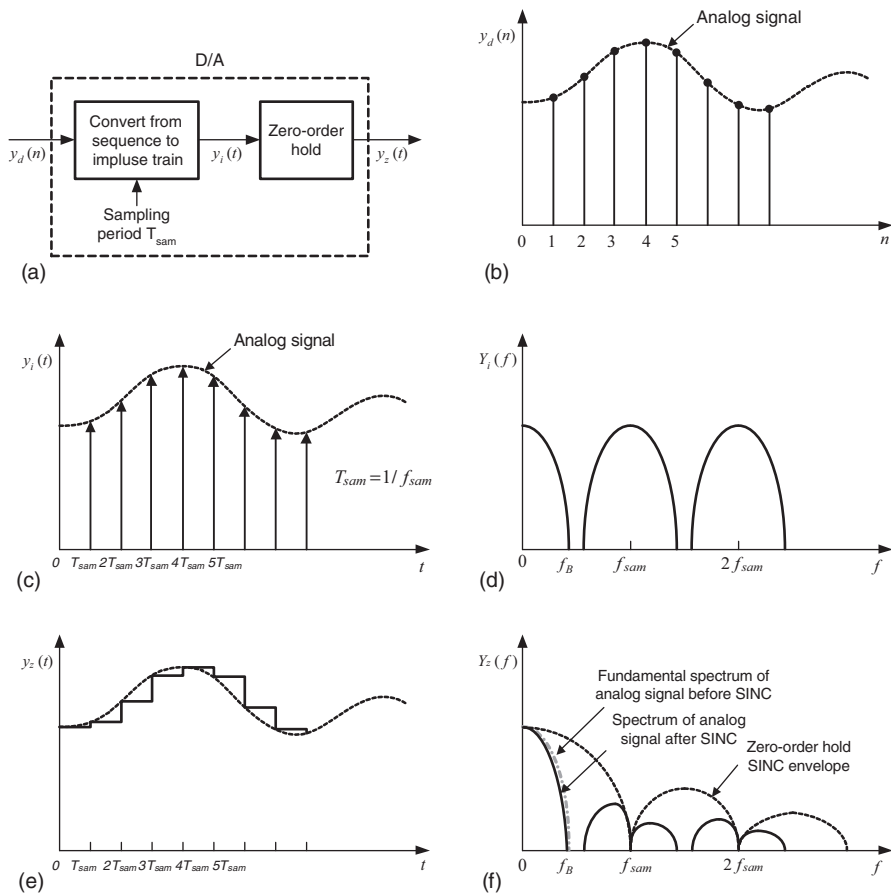


Fig. 2.38 Input and output of DAC in time and frequency domains: (a) block diagram of DAC, (b) discrete-time sequence, (c) impulse train, (d) spectrum of the impulse train, (e) output of the zero-order hold, and (f) Spectrum of the zero-order hold output. f_B is the maximum frequency for a bandlimited analog signal, and f_{sam} is the sampling frequency. Referenced from [25, 26]

The $x/\sin(x)$ shape amplitude compensation given in (2.119) is scaled by $1/T_{sam}$, which is used to cancel a scaling factor of T_{sam} in the ideal filter $H_1(\omega)$ as expressed in (2.117). It is also noted that such an advance time shift of $T_{sam}/2$ in (2.118) can be ignored without any effect on the performance.

At this point, it can be seen from (2.113) that the zero-order hold block introduces $\sin(x)/x$ shape amplitude distortion in the frequency domain. Thus, the amplitude of the signal spectrum at the output of the DAC is multiplied by a function $\sin(x)/x$ or $\text{Sinc}(x)$. As a result, $\text{Sinc}(x)$ function acts as amplitude distortion for the DAC output signal.

Figure 2.38 illustrates the amplitude distortion caused by the zero-order hold characteristic of the DAC, showing the input and the output of the DAC in the time

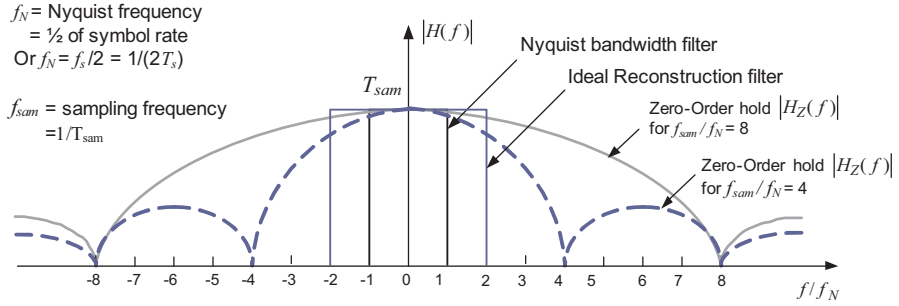


Fig. 2.39 Amplitude response of zero-order hold with sampling frequency $f_{\text{sam}} = 4f_s = 8f_N$. The maximum bandwidth of raised-cosine filter is $2f_N$, or $f/f_N = 2$

domain and their corresponding spectrum patterns in the frequency domain. Figure 2.38f shows that the frequency response of the zero-order hold acts as a lowpass filter that attenuates not only image components but also the fundamental component of the desired signal. Attenuation on the desired signal, however, is frequency dependent, where the attenuation becomes severe when the sampling frequency f_{sam} decreases. For example, the fundamental component before the zero-order hold circuit is shown in the dash-dot line, while after the zero-order hold circuit, the fundamental component is represented by the solid line.

Figure 2.39 shows the magnitude of the frequency response of the zero-order hold circuit when the sampling frequency is either four times or eight times the Nyquist frequency. It is obvious that an amplitude compensator is needed when the sampling frequency is twice as large as the symbol rate or four times as large as the Nyquist frequency, as shown by the dashed line. To compensate for a $\sin(x)/x$ shape distortion, it is natural for the amplitude compensator to have the opposite frequency response of the zero-order hold function, or a $x/\sin(x)$ -shaped frequency response. Thus, the combination of their cascaded frequency response is constant. Usually it is enough for the overall amplitude response to be constant within the bandwidth of $(1 + \alpha)f_N$.

From Fig. 2.39 we can see that this magnitude compensator may be neglected since the amplitude drops slightly as the sampling frequency increases. For example, the amplitude drops by only $2\sqrt{2}/\pi$ (or -0.91 dB) at the symbol rate or twice the Nyquist frequency ($f/f_N = 2$, which is equivalent to $\omega = \pi/(2T_s)$ in (2.113) when the sampling frequency of f_{sam} is four times the symbol rate of f_s ($f_{\text{sam}} = 4f_s = 8f_N$), as indicated by the light-dark solid line. It drops about $2/\pi$ (or -3.92 dB) when the sampling frequency of f_{sam} is twice the symbol rate f_s ($f_{\text{sam}} = 2f_s = 4f_N$), as indicated by the light-dark dashed line. In the former case, this amplitude compensation may be neglected due to -0.91 dB attenuation. In the latter case, the amplitude compensation is needed because of -3.92 dB attenuation. The spectrum bandwidth of the raised-cosine filtered signal is within the range of $f_N < f \leq 2f_N$, corresponding to the alpha value range $0 < \alpha \leq 1$.

Increasing the sampling clock rate of the DAC not only reduces the attenuation effect of the zero-order hold on the desired signal, but also lowers the quantization noise floor and relaxes attenuation requirements for the reconstruction filter [26]. A DAC with a higher clock rate also increases the design cost and power consumption. In some cases, it is very complicated to design a DAC with a high clock rate due to a wider bandwidth of the transmission data. For example, in the ultra-wideband (UWB) system [27] it costs much more to design a DAC with an over-sampling rate of 1056 MHz because the bandwidth of an OFDM signal is about 264 MHz.

In the case where the sampling rate is very difficult to increase, the amplitude compensation technique is an effective method to deal with the amplitude distortion caused by the SINC-function. Amplitude compensation can be achieved with either the digital or the analog filter. In the former case, the pre-distortion is created before the DAC, while in the latter the post-compensation is generated after the DAC. In both cases, the amplitude response of the compensator is the inverse of the SINC-function or $1/\text{Sinc}(x)$. In practice, the overall amplitude response is required to be flat only within the bandwidth of the desired signal, which is equal to $(1 + \alpha)f_N$.

For detailed design information regarding the pre-distortion-based equalizer or compensator, the interested reader can refer to [26]. Here, we introduce the post-compensation method in more detail. In the post-compensation method, an analog filter whose frequency response is approximately equal to the inverse of the SINC-function is inserted either before or after the reconstruction filter. To reach such an inverse shape of the SINC-function, the analog filter needs to have a peak around the Nyquist frequency. The second-order analog filter can realize such a peak with a proper damping factor. The transfer function of the second-order lowpass filter is

$$H_c(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (2.120)$$

where ω_n is the *natural frequency* of the filter and ζ is the *damping factor*. As we know, the amplitude of the frequency response of the second-order lowpass filter has a peak around the natural frequency when $\zeta < 0.707$ is met. Therefore, we can use this peak property of the frequency response to approximate the inverse of the SINC-function.

Considering that a baseband signal is passed through either a RC filter or SRRC filter approximated by a FIR filter, we only need to compensate the amplitude distortion caused by the SINC-function up to the bandwidth $B_w = (1 + \alpha)f_N$. Hence, the amplitude *gain* of the compensation transfer function relative to the DC amplitude should be approximately equal to the amplitude *attenuation* of the SINC-function up to B_w . Usually the natural frequency $f_n = \omega_n/2\pi$ is set greater than B_w , depending on the ratio of the sampling frequency to the signal bandwidth.

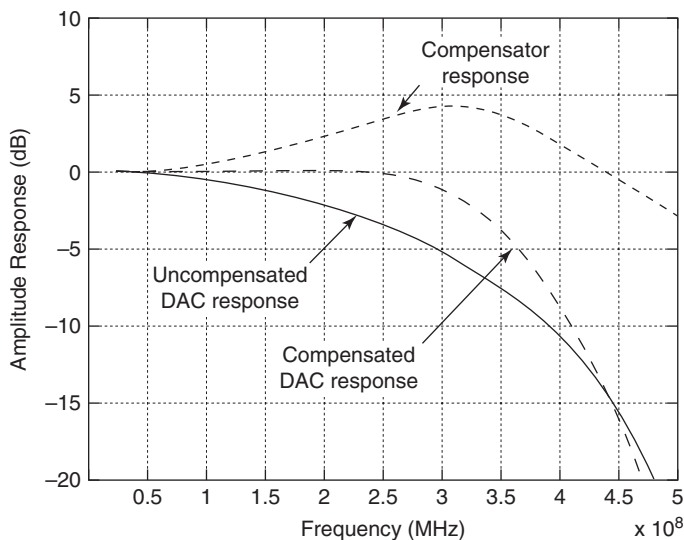


Fig. 2.40 Amplitude response curves of the SINC function and amplitude compensator, where sampling frequency is 528 MHz and Nyquist frequency is 264 MHz

For example, in the UWB OFDM system the single-side bandwidth of the OFDM baseband signal is about 260 MHz [27]. Due to the channel spacing of 528 MHz and the signal bandwidth of 260 MHz, we can have the minimum clock frequency (or sampling frequency) of 528 MHz operate for the DAC. The attenuation of the SINC-function at the half-sampling frequency of $f_{\text{sam}}/2 = 264$ MHz is -3.92 dB from (2.113). Thus, the spectrum attenuation of the OFDM baseband signal approximates to -3.92 dB at the bandwidth edge frequency of 260 MHz. To compensate for amplitude distortion, we set the natural frequency $f_n = 350$ MHz and the damping factor $\zeta = 0.33$ in (2.120) so that the compensator has a gain of about 3.7 dB at the half-sampling frequency of 264 MHz, as shown in Fig. 2.40. After the amplitude compensation, the amplitude response of the compensated DAC at a half the sampling frequency is about $-3.92 + 3.7 = -0.22$ dB.

Figure 2.41 illustrates the PSD of the OFDM signal with a bandwidth of 260 MHz at the output of the reconstruction filter. It can be seen that the PSD of the OFDM signal around the bandwidth edge frequency of 260 MHz in Fig. 2.41b increases after the amplitude compensation so that the overall spectrum becomes flat. Compared with digital-filter-based compensation, analog-filter-based compensation is simple in structure, inexpensive, and has low power consumption.

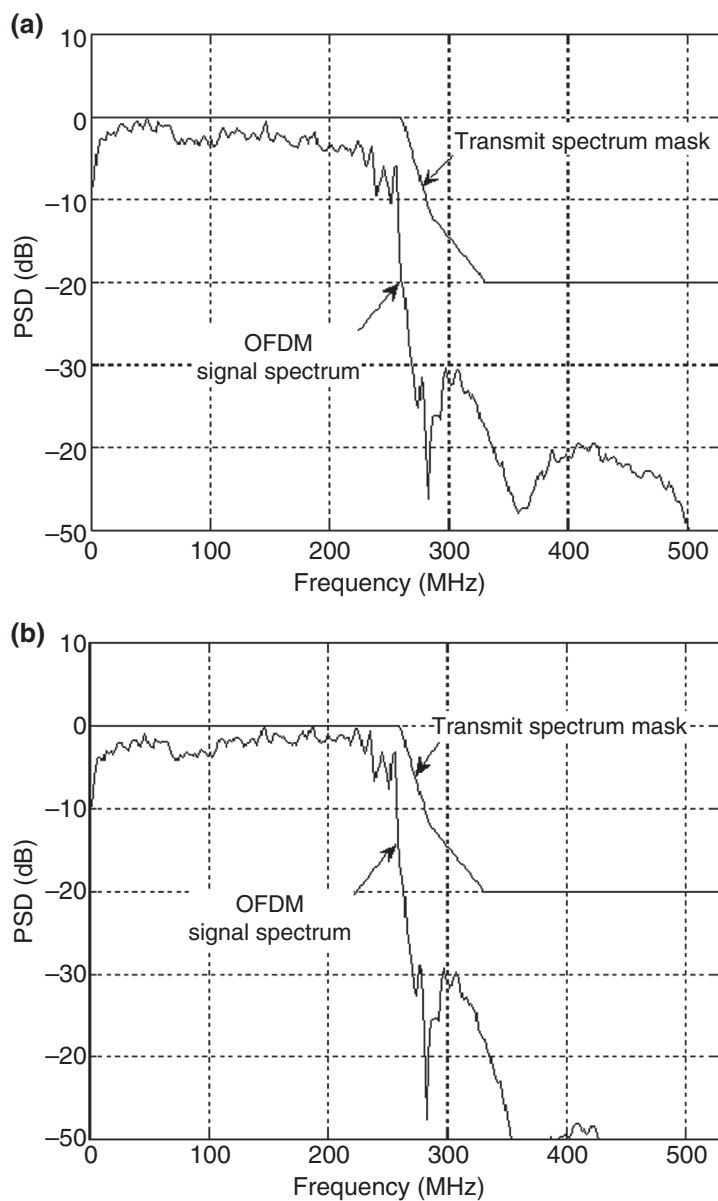


Fig. 2.41 Power spectral density of Multiband OFDM signal at the transmitter: (a) without compensator and (b) with compensator, where sampling frequency is 1056 MHz

References

1. Sklar, B. (2002). *Digital communications: Fundamentals and applications* (p. 168). India: Pearson Education Asia.
2. Wang, A. Y., & Sodini, C. G. (2006). On the energy efficiency of wireless transceivers. In *ICC 2006 Proceedings* (pp. 3783–3788).
3. McCune, E. (2015). A technical foundation for RF CMOS power amplifiers. *IEEE Solid-State Circuits Magazine*, 8(2), 75–82.
4. Joung, J., Ho, C. K., Adachi, K., & Sun, S. (2015). A survey on power amplifier centric techniques for spectrum- and energy-efficient wireless communication. *IEEE Communication Surveys and Tutorials*, 17(1), 315–333.
5. Cripps, S. C. (1999). *RF power amplifiers for wireless communications* (p. 49). Norwood, MA: Artech House.
6. Wimpenny, G. (2012, February 6). Envelope tracking power amplifier characterization (White Paper). Nujira limited. Retrieved from www.nujira.com
7. Application Report, “GC5325 Envelope Tracking,” *Texas Instruments*, SLWA058B, April 2010.
8. Cripps, S. C. (2002). *Advanced techniques in RF power amplifier design* (pp. 4–5). Norwood, MA: Artech House.
9. Zappone, A., & Jorswieck, E. (2015). Energy efficiency in wireless networks via fractional programming theory. *Foundations & Trends in Communications and Information Theory*, 11 (3–4), 185–396.
10. Buzzi, S., Chih-Lin, I., Klein, T. E., Vincent Poor, H., Yang, C., & Zappone, A. (2016). A survey of energy-efficient techniques for 5G networks and challenges ahead. *IEEE Journal on Selected Areas in Communications*, 34(4), 697–709.
11. Ziemer, R. E., & Tranter, W. H. (2002). *Principles of communications: Systems modulation and noise* (5th ed.). New York: Wiley.
12. Feher, K. (1995). *Wireless and digital communications: Modulation & spread spectrum applications*. Upper Saddle River, NJ: Prentice-Hall PTR.
13. ZigBee Alliance. (2006, December). ZigBee Specifications (Version 1.0). Retrieved from <http://www.zigbee.org/>
14. Austin, M. C., & Chang, M. U. (1981). Quadrature overlapped raised-cosine modulation. *IEEE Transactions on Communications*, COM-29(3), 237–249.
15. Le-Ngoc, T., & Feher, K. (1983). Performance of IJF-OQPSK modulation schemes in a complex interference environment. *IEEE Transactions on Communications*, COM-31(1), 137–144.
16. Seo, J. S., & Feker, K. (1985). SQAM: a new superposed QAM modem technique. *IEEE Transactions on Communications*, COM-33(3), 296–300.
17. Gao, W., Ju, D., & Wu, Y. Self-convolving minimum shift keying (SCMSK) modem for satellite system. In *Proceedings of IEEE Singapore ICCS'88* (pp. 341–345).
18. Gao, W., & Feher, K. (1996). All digital reverse modulation architecture based carrier recovery implementation for GMSK and compatible FQPSK. *IEEE Transactions on Broadcasting*, 42(1), 55–62.
19. Proakis, J. G. (1995). *Digital communications* (3rd ed.). New York: McGraw-Hill.
20. Simon, M. K. (2001, June). *Bandwidth-efficient digital modulation with application to deep-space communications*. Deep-space communications and navigation series.
21. Nyquist, H. (1928). Certain topic in telegraph transmission. *Transactions of the AIEE*, 47(2), 617–644.

22. Tenbroek, B., Strange, J., Nalbantis, D., Jones, C., Flowers, P., Brett, S., et al. (2008). Single-chip tri-band WCDMA/HSDPA transceiver without external SAW filters and with integrated TX power control. In *ISSCC 2008* (pp. 202–204).
23. Jussila, J. (2003, June). Analog baseband circuits for WCDMA direct conversion receivers. PhD Dissertation, Helsinki University of Technology, Finland.
24. Li, Z., Li, M., Zhao, D., Ma, D., Ni, W., & Ouyang, Z. (2010). TD-SCDMA/HSDPA transceiver and analog baseband chipset in 0.18- μm CMOS process. *IEEE Transactions on Circuits and Systems-II*, 57(2), 90–94.
25. Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999). *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice Hall.
26. Yang, K. (2006, April 13). Flatten DAC frequency response. *END Magazine*, 65–74.
27. Multiband OFDM physical layer proposal for IEEE 802.15 Task Group 3a (2004, September 14). MBOA-SIG multiband OFDM alliance SIG.

Energy and Bandwidth-Efficient Wireless Transmission

Gao, W.

2017, XXIII, 485 p. 268 illus., 55 illus. in color.,

Hardcover

ISBN: 978-3-319-44220-4