2 Linearity Performance of Outphasing Power Amplifier Systems

2.1 Introduction

As discussed in Chapter 1, in an outphasing power amplifier system, the AM of a bandpass waveform is transformed into PM of two signals prior to power amplification, and finally recovered by recombining the two amplified signals. As shown in Figure 1.16, with this approach, two highly efficient nonlinear power amplifiers can be used without distorting the original signal. This is the main advantage of the outphasing approach to linear power amplification. However, the bandlimited feature is no longer preserved—a wideband signal is added to and subtracted from the original source input to construct the two outphased component signals. Since the spectrum of the two additional signal components extends far into the adjacent channels, the system linearity performance relies on the precise cancellation of the wideband signals during power combining, and the tolerance on the matching between the two power amplifier branches becomes critical—a slight mismatch between the two paths can result in significant distortion in the adjacent and alternate frequency bands.

The gain and phase characteristics of the two amplifier branches, which includes the digital baseband, analog, and RF portions, must be well-matched for the acceptably small out-of-band rejection. A failure to achieve sufficient matching generally creates both out-of-band interference and inband distortion. The out-of-band interference is characterized by the ACPR, while the inband distortion may be described in terms of EVM. Various factors can contribute to path mismatch and the consequent linearity
degradation. These include environmental variations such as thermal drift, component aging, and channel transition. A number of coupling effects in integrated circuits—substrate, capacitive, bond wire, and package—as well as process variation during fabrication could also contribute to a significant amount of mismatch. The main contributors to the mismatch are the RF gain and phase errors, the quadrature modulator errors, and the imbalance caused by the SCS. If the SCS is implemented with DSP, the effects of the quantization noise, DSP sampling rate, and reconstruction filter also need to be investigated.

This chapter begins with a brief discussion of the relevant modulation schemes and baseband filtering typically used in mobile communications applications. Next, various techniques for the implementation of the signal component separation will be described. The major sources of imbalance, including the path imbalance, the quadrature modulator error, the SCS quantization, the DSP sampling rate, and reconstruction filtering, are successively investigated. Finally, we summarize some of the practical implementation issues for outphasing power amplifiers.

2.2 Digital Modulation Techniques

Modulation is a process that encodes source information onto a carrier signal for optimized transmission. During the modulation process, the analog or digital source signal is mapped onto the amplitude or phase (or both) of the carrier signal. Digital modulation offers a number of advantages over its analog counterpart, such as increased capacity, tolerance to channel impairments and noise, and accommodation to various digital signal conditioning and processing techniques [1–4]. Various modulation schemes and their variants have been developed. Their names are usually derived from the mapping operation, such as PSK, FSK, and quadrature amplitude modulation (QAM). This section briefly introduces several digital modulation techniques that are used in modern digital communications, including QPSK, OQPSK, $\pi/4$-DQPSK, and QAM. GMSK will be discussed in Section 2.3, together with the Gaussian filter.

2.2.1 QPSK and Its Variations

QPSK belongs to the family of M-ary phase shift keying modulation, in which the digital data is mapped onto M-number of discrete phase states. As shown in Figure 2.1(a), four phase states of $\pm 45^\circ$ and $\pm 135^\circ$ are used in QPSK, and
each phase state represents a single symbol (or two bits) of information. QPSK modulation can be realized by a quadrature modulator with the baseband I/Q data streams as the driving signals. Figure 2.1(a) also shows that the possible
phase transitions are $-90^\circ$, $+90^\circ$, or $180^\circ$. During a $180^\circ$ transition, the signal trajectory passes through the origin, which is also known as a “zero-crossing.” The zero-crossing leads to the well-known problem of spectral regrowth, when the QPSK waveform passes through the baseband filter and is processed by a saturated power amplifier. Baseband filtering, which is discussed in Section 2.3, is used to suppress the sidelobe spectrum, improving the spectral efficiency and minimizing the interference spreading to adjacent channels. However, it introduces a variation on the envelope of the filtered QPSK waveform. Specifically, every time a zero-crossing is encountered, the envelope of the filtered signal drops to zero. This situation is illustrated in Figure 2.1(b). The baseband filter used is a square-root raised cosine filter with roll-off 0.35, described in Section 2.3. A saturated power amplifier possesses high power efficiency but usually exhibits significant nonlinearity when confronted with a high degree of variation in the envelope of the input signal. Power amplifier nonlinearity results in a spectral leakage to the neighboring channels, excited by the signal envelope fluctuation. Hence, QPSK-modulated waveforms generally require a highly linear power amplifier to accommodate the substantial envelope variation.

The zero-crossing may be eliminated if an appropriate time offset is introduced between the I/Q data streams. In OQPSK, the Q-channel data stream is delayed by half the symbol period and the phase transitions are therefore restricted to $\pm 90^\circ$. Figure 2.2(a) illustrates the signal constellation and the phase transitions in OQPSK. As shown in Figure 2.2(b), the filtered OQPSK waveform exhibits much less envelope fluctuation and is therefore much more attractive for linear power amplifier implementation. Due to the similarity, OQPSK has the same BER and power spectral density (PSD) as QPSK. OQPSK is used in the CDMA IS-95 reverse link.

The $\pi/4$-DQPSK modulation technique is another variation of QPSK that differs from the latter in two aspects. First, the data stream is differentially encoded such that the information bits are embedded in the phase change between the successive symbols rather than the absolute phase. As an example, in differential binary PSK (DBPSK) modulation, a digital “1” indicates a $180^\circ$ change on the absolute phase of the carrier, while a “0” results in no phase change, or vice versa. One advantage with $\pi/4$-DQPSK is that the differential encoding enables “noncoherent” detection, which simplifies the receiver implementation and is especially desirable for mobile communications. Second, an additional $45^\circ$ phase shift is inserted in each symbol such that there exist eight possible phase states while the phase transition is restricted to $\pm 45^\circ$ and $\pm 135^\circ$. The constellation and phase transition of $\pi/4$-DQPSK is illustrated in Figure 2.3(a). This scheme is
therefore a compromise between QPSK and OQPSK, and the maximum phase transition of 135° leads to less envelope variation compared to QPSK, as shown in Figure 2.3(b). The 45° phase shift also ensures the continuous

Figure 2.2 (a) OQPSK signal constellation and phase transitions, and (b) a baseband-filtered time domain waveform indicating phase transition and envelope variation.
phase transition for each symbol that allows the proper operation of the timing recovery and synchronization circuits in the receiver. One disadvantage of $\pi\!/4$-DQPSK is the higher BER compared to QPSK.

Figure 2.3 (a) $\pi\!/4$-DQPSK signal constellation and phase transitions, and (b) a baseband-filtered time domain waveform indicating phase transition and envelope variation. I and Q data are encoded.

phase transition for each symbol that allows the proper operation of the timing recovery and synchronization circuits in the receiver. One disadvantage of $\pi\!/4$-DQPSK is the higher BER compared to QPSK.
It should be pointed out that a higher envelope variation of the bandlimited signal, or more precisely, a higher PAP, does not necessarily lead to greater spectral regrowth or intermodulation [5]—for example, when the signal is fed to a power amplifier with modest nonlinearities, like a Class AB or B design. Conventionally, the power amplifier is backed off from its 1-dB compression point approximately by an amount of PAP to accommodate the signal envelope variation. The intermodulation is generated during power amplification when the high input level drives the amplifier near saturation. Therefore, the average amount of intermodulation generated by the power amplifier is dependent on the statistical behavior of the instantaneous signal power above the average point. This time-domain distribution is characterized by the CCDF or envelope distribution function. From [5], although the QPSK-modulated signal possesses a higher value of PAP, it actually exhibits less spectral regrowth than OQPSK- and π/4-DQPSK-modulated signals on the same average power basis. The power amplifier used was a Class AB, and the filters were CDMA IS-95 or IS-54 baseband filters. However, a higher PAP does result in a lower power efficiency in many cases.

2.2.2 QAM

QAM is a modulation technique in which the symbols have both amplitude and phase variations, as compared to PSK modulation techniques where only the phase of the carrier is varied. A M-ary QAM modulation combines every log_2 M bit as an individual symbol, and the symbol rate is hence log_2 M times less than the bit data rate. A typical value for M is in the range of 16–1,024. QAM is a spectrally efficient modulation scheme, and, it is used by computer modems and wireless LAN systems. Figure 2.4 illustrates the constellation of 16-QAM, with each symbol representing four bits. In some applications, the I/Q data streams are offset by half the symbol period, just as with OQPSK.

2.3 Baseband Filtering of Digital Data

The transmission of purely digital data across the wireless communications channel represents an interesting problem for those concerned with the spectral efficiency of the communications system. In this case, the spectral efficiency of a digital signal is defined as the number of bits per
second of data that can be supported by each hertz of channel bandwidth, that is,

\[ h = \frac{R}{B} \]  

where \( R \) is the data rate, and \( B \) is the channel bandwidth. Clearly, a higher value of \( h \) is more desirable than a lower value. Intuitively, as the noise introduced by the channel grows, the maximum value of \( h \) will decline. The maximum spectral efficiency is given by Shannon’s channel capacity formula [6]

\[ h_{\text{max}} = \ln \left( 1 + \frac{S}{N} \right) \]  

where \( S/N \) is the signal-to-noise ratio (SNR) of the received signal. A delightful “proof” of this theorem can be found in [7]. Unfortunately, the total bandwidth of a square-wave digital signal is infinite, containing spectral energy at all of the harmonics of the fundamental frequency. A polar signal (a digital sequence consisting of +1 or -1) consisting of a sequence of random binary data has a spectral response of

\[ P_{\text{polar}}(f) = A^2 T_b \frac{\sin^2(\pi f T_b)}{(\pi f T_b)^2} \]  

Figure 2.4 16-QAM constellation.
and a unipolar signal (a digital sequence consisting of +1 or 0) consisting of a sequence of random binary data has a spectral response of

\[
P_{\text{unipolar}}(f) = \frac{A^2 T_b}{4} \frac{\sin^2(\pi f T_b)}{(\pi f T_b)^2} \left[ 1 + \frac{1}{T_b} \delta(f) \right]
\]

(2.4)

where \(A\) is the amplitude of each bit, \(T_b\) is the time period associated with each bit, and \(\delta(f)\) is the Dirac function. Clearly, this function has significant energy to a frequency much higher than \(1/T_b\). The one-sided spectral densities for these two cases are shown in Figure 2.5. Therefore, a crucial distinction between analog and digital transmission is the requirement for filtering the data in the case of digital transmission to minimize the required bandwidth.

As a result, the digital data is typically filtered prior to modulation and upconversion. This filtering improves the spectral efficiency dramatically, but at the potential cost of another phenomena known as intersymbol interference (ISI). The process of lowpass filtering the digital data narrows its response in the frequency domain, but because of the inverse relation

![Figure 2.5](image-url)

**Figure 2.5** Spectral density of unipolar NRZ data and polar NRZ data.
between the time and frequency domain characteristics of any signal (a narrow signal in the time domain will create a broad signal in the frequency domain, and vice versa), the time response of each pulse is widened. This is often referred to as the uncertainty principle in signal processing—analogous in some ways to Heisenberg’s uncertainty principle in physics. For example, the RC lowpass filter of Figure 2.6 illustrates a random stream of “1’s” and “0’s” passing through the filter. Note that each bit is potentially corrupted by the long “tails” of the responses of the previous bits, and ISI results. If the data was perfectly periodic, then the system would settle down into a steady state response and there would be no ISI, but of course there would be no information transmitted in such a case.

There are several possible solutions to this problem. The most commonly used is known as Nyquist’s first method, where the impulse

![Figure 2.6 ISI due to lowpass filtering: (a) lowpass RC filter, and (b) digital data passed through lowpass filter exhibits ISI.](image-url)
response of the filter used to improve the bandwidth efficiency of the system has the property that

\[ h(kT_b + \tau) = \begin{cases} 
C, & k = 0 \\
0, & \text{otherwise}
\end{cases} \quad (2.5) \]

This feature of the filter eliminates ISI exclusively at times that are integral multiples of the bit period. It is of course impossible to eliminate ISI at all times, but if it can be eliminated at the instant of sampling, then the problem can be dramatically reduced. The class of filters that satisfy this criteria are known as Nyquist filters. One example of an “ideal” Nyquist filter exhibits a \text{sinc} \ ((\sin x)/x) behavior for its impulse response, such that

\[ h(t) = \frac{\sin\left(\frac{\pi t}{T_b}\right)}{\pi t/T_b} \quad (2.6) \]

Note that the impulse response of this filter is zero for all values of \( t = nT_b \), satisfying Nyquist’s criterion. The filter also has the desirable property that it behaves like a “brick-wall” filter in the frequency domain. However, this ideal filter is physically unrealizable, since its impulse response is noncausal. In addition, its peak amplitude decays only as \((1/t)\), and so any mistiming in the clock recovery circuit can generate significant ISI. One of the filters that satisfy Nyquist’s criterion is the \textit{raised cosine} filter family. The raised cosine filter has an impulse response that drops off much faster in time than the \text{sinc} response, at the expense of a small increasing bandwidth, and hence finds its applications in wireless communications.

### 2.3.1 Raised Cosine Filter

The raised cosine filter derives its name from its shape in the frequency domain. The frequency response of a raised cosine filter can be described by

\[
H(f) = \begin{cases} 
T_s, & 0 \leq |f| \leq \frac{1 - \alpha}{2T_s} \\
\frac{T_s}{2} \left\{ 1 + \cos \left( \frac{\pi T_s}{\alpha} \left( |f| - \frac{1 - \alpha}{2T_s} \right) \right) \right\}, & \frac{1 - \alpha}{2T_s} \leq |f| \leq \frac{1 + \alpha}{2T_s} \\
0, & |f| > \frac{1 + \alpha}{2T_s}
\end{cases} \quad (2.7)
\]
where $\alpha$ is the roll-off factor that determines the excess bandwidth of the filter frequency response and in the range of $[0,1]$, and $T_s$ is the sampling or the symbol period. A zero roll-off is essentially a sinc response. As shown in Figure 2.7(a), the filter response has a cosine roll-off shape between $(1 - \alpha)/2 T_s$ and $(1 + \alpha)/2 T_s$ and becomes strictly zero beyond $(1 + \alpha)/2 T_s$. This frequency is thus the bandwidth of the raised cosine response

$$\text{BW} = \frac{1 + \alpha}{2 f_s}$$

(2.8)

where $f_s = 1/T_s$ is the sampling rate or the symbol rate. The corresponding filter impulse response is described by

Figure 2.7 (a) Raised cosine filter frequency response, and (b) impulse response for various roll-off factors.
It is clear that at the sampling instant \( t = kT_s \), the impulse response is essentially zero except at \( k = 0 \). Therefore no ISI will be created. The peak amplitude of the raised cosine impulse response, with a modest value of \( \alpha \), drops off much faster than the \( \text{sinc} \) response. This is due to a more gradual roll-off of cosine shaping at the filter cutoff frequency. Note that the smaller the roll-off factor, the less the absolute bandwidth but the higher the impulse response ripple, as shown in Figure 2.7(b). The raised cosine filtering is widely used in several communication standards, and a value of \( \alpha \) of 0.35 is used in IS-54, 0.5 in PDC, and 0.22 in WCDMA.

In most communications applications, the raised cosine filter is divided into two sections, one placed in the transmitter and the other in the receiver—each one is the so-called square-root raised cosine filter. With this arrangement, the two sections together form the Nyquist filter, while the filter in the receiver side also serves as a matched filter. The matched filter has a response that passes the desired signal, while it rejects the noise and interference and thus maximizes the SNR. In the presence of white noise, the matched filter has an impulse response that is simply the time-reversed transmitted signal pulse shape. The frequency response of a square-root raised cosine filter is the square root of (2.7) and the impulse response is given by

\[
\begin{align*}
\text{for } & \cos \left( \pi \alpha \frac{t}{T_s} \right) \quad \text{and} \quad \sin \left( \pi \frac{t}{T_s} \right) \\
\text{and } & \frac{1}{1 - \left( \frac{2\alpha t}{T_s} \right)^2} \left( -\frac{\pi}{T_s} \right) \\
\end{align*}
\]

(2.9)

The frequency and impulse responses are shown in Figure 2.8. The noise bandwidth of the square-root raised cosine filter is readily calculated with (2.7)

\[
B = \frac{1}{|H_i(0)|^2} \int_0^\infty |H_i(f)|^2 df = \frac{1}{2} f_s
\]

(2.11)

which is simply half the symbol rate.
The impulse response of the raised cosine filter lasts indefinitely and needs to be truncated for use in practical applications. The truncation results in a set of nonzero side lobes in the frequency domain. However, since the pulse response decay is proportional to \(1/t^3\), a proper truncation results in little deviation from the theoretical performance. A smaller roll-off factor \(\alpha\) results in a higher ripple in the impulse response and thus requires a longer pulse shape to achieve the same stopband attenuation. For a stopband attenuation of approximately 40 dB, a rough estimate on the required truncation length for \(0.2 \leq \alpha \leq 0.75\) is given by

\[
L_{\text{min}} = -44\alpha + 33
\]  

(2.12)

where \(L_{\text{min}}\) is the truncated filter impulse response length.

The impulse response of the raised cosine filter lasts indefinitely and needs to be truncated for use in practical applications. The truncation results in a set of nonzero side lobes in the frequency domain. However, since the pulse response decay is proportional to \(1/t^3\), a proper truncation results in little deviation from the theoretical performance. A smaller roll-off factor \(\alpha\) results in a higher ripple in the impulse response and thus requires a longer pulse shape to achieve the same stopband attenuation. For a stopband attenuation of approximately 40 dB, a rough estimate on the required truncation length for \(0.2 \leq \alpha \leq 0.75\) is given by

\[
L_{\text{min}} = -44\alpha + 33
\]  

(2.12)

where \(L_{\text{min}}\) is the truncated filter impulse response length.

The raised cosine filter may be implemented within the baseband DSP or with an analog filter. With the baseband DSP approach, the sampled
values of the filter shape are stored in a lookup table (LUT) and read out in sequence to form the desired symbols. The advantage of the analog approach to the realization of the filter is the reduced hardware complexity, since the raised cosine filter also serves as a reconstruction filter following the D/A converter.

### 2.3.2 Gaussian Filter

A Gaussian filter is used in GMSK, which is the modulation approach adopted for the GSM standard [8]. GMSK modulation is minimum phase shift keying (MSK) with Gaussian filtering and can be implemented by passing the nonreturn-to-zero (NRZ) rectangular waveform (which has values of ±1) through a Gaussian filter followed by a frequency modulator with a modulation index of 0.5. GMSK is intrinsically a constant envelope modulation and thus can be applied directly to the nonlinear power amplifier without distortion. The frequency response of a Gaussian filter is described by

\[
H(f) = e^{-\left(\frac{f}{a}\right)^2}
\]

where the parameter \( a \) is a constant that determines the filter 3-dB bandwidth \( B \)

\[
B = \sqrt{\ln 2} \cdot a = 0.5887a
\]

The impulse response of the Gaussian filter is

\[
h(t) = \sqrt{\pi}ae^{-\left(\pi at\right)^2}
\]

Note that both the frequency response and impulse response have the same Gaussian shape. Gaussian filtering is usually specified by the relative bandwidth in terms of the bit rate of the data stream, for example,

\[
BT = \frac{B}{f_b}
\]

where \( f_b \) is the bit data rate. The fact that the Gaussian impulse response exhibits a smooth roll-off with no overshoot or ringing leads to little ISI. Gaussian filtering also suppresses the high-frequency components and
provides a narrower bandwidth compared to MSK modulation without filtering. As $BT$ decreases, however, the filtered pulse tends to span over several symbol periods and more ISI is generated. GSM uses a relative bandwidth of 0.3 as a compromise between the transmission bandwidth and ISI.

### 2.3.3 IS-95 Baseband Filter

The CDMA IS-95 system uses a pulse-shaping filter that does not satisfy Nyquist’s criterion and is included here for completeness. The IS-95 baseband filter uses a Chebyshev roll-off instead of cosine roll-off. Figure 2.9 shows the limit of the normalized frequency response of the baseband filter [9]. The passband edge frequency $f_p$ is 590 kHz, and the stopband edge frequency $f_s$ is 740 kHz. The passband ripples are bounded within $\pm 1.5$ dB, and the stopband attenuation is equal to or greater than 40 dB. The impulse response of the baseband filter $h_c(t)$ satisfies the following equation

$$
\sum_{k=0}^{\infty} [\alpha h_c(kT_s - \tau) - h(k)]^2 \leq 0.03
$$

(2.17)

where the constants $\alpha$ and $\tau$ are used to minimize the mean square error. The constant $T_s$ is one-quarter of a PN chip. The values of the coefficients $h(k)$ are listed in [9] for $k<48$. These values can be directly used as a 48-tap baseband FIR filter with 4× oversampling rate.

![Figure 2.9](image.png)

**Figure 2.9** CDMA IS-95 baseband filter frequency response limits [9].
2.4 Signal Component Separation for Outphasing Amplifiers

Despite the neatness and elegance of the outphasing concept, the practical complexity of realizing the signal component separation has limited its wide acceptance. The difficulty stems from the fact that the generation of the two component signals involves memoryless nonlinear signal processing that requires a high degree of accuracy. During the past decades, various approaches have been proposed to accomplish this function, but difficulties remain in terms of such factors as implementation complexity, bandwidth, and power consumption, due to the stringent distortion and noise requirements.

There are two canonical forms for the signal component separation process—the phase modulation method and the in-phase/quadrature method [10]. In the phase modulation method, the phase modulation $\theta(t)$ and amplitude modulation $r(t)$ of the input are separated first, and two nonlinear phase modulators with inverse sine characteristics are used to transform the input envelope variation into the phase of the desired output signals, as shown in Figure 2.10. The two component signals can be described by [11–13]

\[
S_1(t) = A_c \sin[\omega_c t + \theta(t) + \phi(t)]
\]

\[
S_2(t) = A_c \sin[\omega_c t + \theta(t) - \phi(t)]
\]

where $A_c$ is the amplitude of the carrier signal, and $0 \leq A_c \leq r(t)$. $\phi(t)$ is given by

\[\int r(t) \sin[\omega_c t + \theta(t)] dt\]

\[\approx A_c \sin[\omega_c t + \theta(t)]\]

\[\text{Arcsine phase modulator}\]

\[\text{Envelop detector}\]

\[\text{Arcsine phase modulator}\]

\[S_1(t)\]

\[S_2(t)\]

Figure 2.10 Signal component separation with nonlinear phase modulators [10].
\[
\sin \phi(t) = \frac{r(t)}{A_c} \tag{2.20}
\]

The desired output signal is obtained by subtracting \( S_2(t) \) from \( S_1(t) \) following the power amplification,

\[
S(t) = G S_1(t) - G S_2(t) = 2 G r(t) \cos[\omega_c t + \theta(t)] \tag{2.21}
\]

The SCS in this scheme is typically realized as an analog circuit and operates at some intermediate frequency or directly at the carrier frequency. Figure 2.11 shows an implementation of the inverse sine phase modulator, which is constructed by a linear phase modulator with the necessary feedback [11]. The operation of the circuit can be understood by making the following idealization—the input impedance of the amplifier \( G \) is much higher than \( R_f \) or \( R_i \), so that the current injected into the input of the amplifier is negligible, and the loop gain is sufficiently large so that the amplifier input node \( V_i \) is close to a virtual ground. These assumptions give rise to the following equation

\[
\frac{r(t)}{R_i} + \frac{V_f(t)}{R_f} = 0 \tag{2.22}
\]

The phase modulator modulates the phase of the carrier signal by an amount linearly proportional to the amplifier output \( V_o \).

Figure 2.11  Nonlinear inverse sine phase modulator [11].
\[ S_1(t) = A_c \sin[\omega_c t + \theta(t) + mV_o(t)] \quad (2.23) \]

where \( m \) is the characteristic of the phase modulator and \( A_c \) is the amplitude of the carrier signal. The mixer and lowpass filter act as a phase detector that downconverts the modulator output and removes the original phase modulation term \( \theta(t) \), then

\[ V_f(t) = \frac{1}{2} A_c^2 \sin[mV_o(t)] \quad (2.24) \]

Substituting (2.22) into (2.24), we have

\[ mV_o(t) = -\arcsin \left[ \frac{2R_f}{2R_c} r(t) \right] \quad (2.25) \]

The amplifier output \( V_o \) can be made proportional to the inverse sine of the input envelope by adjusting the values of \( A_c \), \( R_f \), and \( R_l \). As a bonus, the phase modulator output \( S_1(t) \) is one of the constant-envelope vector components with the desired phase modulation. A similar circuit can generate the other component signal \( S_2(t) \) by inverting the sign of the amplifier output \( V_o(t) \). The nonideality of the limiter and envelope detector, the finite loop gain, and the loop noise all contribute to the performance degradation. Another limitation is that the feedback loop delay limits the bandwidth of the modulator.

As described in Chapter 1, the in-phase/quadrature method first computes the I/Q components of the quadrature signal \( e(t) \),

\[ e(t) = js(t) \sqrt{\frac{r_{\text{max}}^2}{r^2(t)}} - 1 \quad (2.26) \]

then the two component signals are obtained by summing and subtracting the quadrature signal from the source input.

\[ S_1(t) = s(t) - e(t) \quad (2.27) \]
\[ S_2(t) = s(t) + e(t) \quad (2.28) \]

A graphic illustration of the I/Q component separation is given in Figure 1.17.

The calculation of the quadrature signal involves complicated functions, such as multiplication, division, and square root, which are generally
difficult to be accurately and cost-effectively implemented. However, the in-phase/quadrature SCS can take the advantage of baseband signal processing, and two quadrature modulators are used to translate the component signals to the desired carrier frequency. Figure 2.12 illustrates an early version of an analog signal component separator that may operate at intermediate frequency (IF) or RF [10].

Another means to realize linear amplification with nonlinear power amplifiers with the outphasing approach is based on a feedback tracking loop to lock the power amplifier output to the input waveforms [14–17]. These approaches can be regarded as the generalization of the feedback power amplifier applied to the outphasing system. Two representative methods that fall into this category are CALLUM [14] and the vector-locked loop (VLL) [15], in which the amplifier output is compared to the source input and the error signal is used to control the relative phases of two VCOs. The former technique (CALLUM) is realized in Cartesian form while the latter (VLL) is realized in polar form. Since no explicit SCS is involved, and the feedback loop automatically corrects the path imbalance and improves the system linearity, these two structures will be discussed in Chapter 3.

The in-phase/quadrature method is most common and practical means for SCS, whether implemented in analog or digital circuitry. The rest of this section reviews the different ways to generate the quadrature signal, or in a sense, realize the inverse sine or cosine function. Figure 2.13 shows an inverse sine phase modulator, which consists of a balanced modulator, a summer,
and an envelope-limiting amplifier [18]. The circuit takes a baseband envelope as the input and modulates the phase of an IF carrier approximately in accordance with the \( \arcsin \) characteristic. The modulator output and its conjugate can then be used along with the original PM of the input signal to provide the two constant envelope signals. The analysis [18] shows if the loop delay \( \tau \) is small and satisfies

\[
\omega_c \tau = n\pi + \frac{\pi}{2},
\]

where \( \omega_c \) is the carrier radian frequency and \( n \) is an arbitrary integer; then the phase of the modulator output consists of a linear term and a distortion term proportional to \( \tau^2 \) such that

\[
\phi(t) \approx (-1)^n \arcsin \left\{ r(t) + \frac{r(t) r'(t) \tau^2}{2[1 - r^2(t)]^2} \right\}
\]

where \( r'(t) \) is the derivative of the signal envelope. The second-order effect of the loop delay in the above expression could result in a considerable improvement on distortion and bandwidth. Note that no filter is used in this modulator.

Another analog technique for generation of the quadrature signal is illustrated in Figure 2.14 [19]. The AM and PM of the input are separated and summed together to pass through a comparator or a hard limiter, whose output is either +1 or −1—the polarity of the driving signal. The in-phase component of the quadrature signal is formed after bandpass filtering the comparator output; that is,

\[
S_1(t) = \frac{4}{\pi} \cos[\omega t + \theta(t)] \int_0^\infty \frac{\cos[r(t)x]}{x} J_1(x) dx
\]

\[
= \frac{4}{\pi} \cos \{\arcsin[r(t)]\} \cos[\omega t + \theta(t)]
\]

\[
(2.31)
\]
where $J_1$ is the first-order Bessel function of the first kind. The envelope $r(t)$ is assumed to be normalized to $r_{\text{max}}$. As a by-product, the inverse sine is obtained by lowpass filtering the comparator output—that is,

$$
S_0(t) = \frac{2}{\pi} \int_0^\infty \frac{\sin[r(t)x]}{x} J_0(x) dx = \arcsin[r(t)]
$$

(2.32)

The quadrature signal, as its name implies, always has a 90° phase difference with respect to the input. This fact suggests that the quadrature signal can be generated by rotating the input signal by 90° and with a proper scaling. According to (2.26), the scaling can be realized by a variable gain amplifier (VGA) with a gain factor of

$$
\text{gain} = \sqrt{\frac{r_{\text{max}}^2}{r^2(t)} - 1}
$$

(2.33)

It is readily seen that this scheme demands that the VGA gain approaches infinity as the input envelope approaches to zero. This is of course impossible; hence it is important to ensure that there is no zero-crossing in the modulated signal. Fortunately, this requirement is exactly what is required for some linear modulations, including $\pi/4$-DQPSK and OQPSK. Figure 2.15 shows an implementation of this VGA-based scheme with power feedback [20]. The power of the two component signals are added and then subtracted from a reference signal. The error signal is fed to a lowpass filter and an amplifier to drive the VGA. In other words, the feedback loop ensures the constant envelope of the two output signals. The SCS chip was
implemented in a 0.8-\(\mu\)m BiCMOS technology and operates at 200-MHz IF. The tested bandwidth is 1.8 MHz, due to on-chip implementation of the SCS.

The VGA gain may be directly calculated based on (2.33), and that results in an improved bandwidth over the feedback approach. Figure 2.16 illustrates the block diagram of this scheme [21, 22]. The VGA control signal is obtained by the use of a squaring circuit, a lowpass filter, a reference signal, and a key analog circuit that performs division and “square-root” functions.
The key circuit is implemented based on a translinear core, as shown in Figure 2.17. Six NPN bipolar transistors \( Q_1 - Q_6 \) form a translinear loop. The translinear principle states that in a translinear loop the product of the collector current density in the clockwise direction is equal to the product of the collector current density in the counterclockwise direction [23], which can be easily proved by adding the base-emitter voltages of the transistors in each direction. The diode-connected transistors like \( Q_1 \) and \( Q_6 \) are used to take the square of the current \( I_c \) and \( I_{out} \), which transforms the control current to a voltage and takes the square root, respectively. Thus

\[
I_{out} = \sqrt{\frac{I_c^2(I_{ref} - I_{sq})}{I_{sq}}} \quad (2.34)
\]

A VGA control voltage proportional to the output current is then generated. The major error source of the above expression is the finite current gain of the transistors, particularly as \( I_{sq} \)—proportional to the instantaneous input level—approaches a small value. Some compensation circuits can be used to improve the accuracy of the calculation [21]. The SCS chip set of [21] was

![Figure 2.17 A bipolar translinear core [21].](image-url)
fabricated in a 0.8-μm BiCMOS process and operated at a 200-MHz IF. A CMOS SCS chip set of the similar scheme is given by [22], in which the key circuit is based on the voltage translinear circuit principle to perform the square root and division [24].

So far, the discussion of the SCS is restricted to that of an analog circuit implementation. An apparent disadvantage of the analog implementation is the difficulty to obtain highly accurate modulated component signals to meet the stringent linearity requirements imposed by the wireless standards. The advent of modern DSP and the application-specific DSP (ASDSP) technology makes a digital implementation of the SCS feasible [25]. With this method, all the signal processing is implemented at the baseband, and two quadrature modulators are used to translate the two component signals to the desired carrier frequency, as shown in Figure 2.18. However, the computational load required by the DSP imposes a practical constraint on the performance in terms of bandwidth and power consumption. There are also other constraints and trade-offs associated with the DSP/ASDSP approach, such as the word length of the DSP, the sampling rate, and the reconstruction filter selection. These factors will be discussed in Chapter 3. One disadvantage of this approach is that the reconstruction filter cannot be the pulse-shaping filter due to the nonlinear transformation in the SCS.

A digital SCS method based on a look-up-table (LUT) was suggested in [26], in which a two-dimensional LUT stores the complex quadrature signal, and each LUT entry is addressed by the in-phase and quadrature components of the baseband digital input. The only instructions required are a table lookup and two complex additions for every processed sample. A practical

![Figure 2.18 Digital signal component separator with frequency translation.](image-url)
constraint of this method is the extremely large memory requirement on the LUT. As an example, if the complex input and quadrature signals are quantized to nine bits, the LUT would have at least $2^9 \times 2^9 \times 2 \times 9 \approx 4.7$ Mbits memory to store all the possible values of the quadrature signal. However, by utilizing the symmetry along the real and imaginary axes, the memory requirement would be reduced by a factor of four. Note that a further advantage of symmetry can be taken, as shown in Figure 2.19. The vector $s'$ and $s$ are symmetric with respect to the $45^\circ$ axis, and their quadrature vector $e'$ and $e$ are symmetric with respect to the $-45^\circ$ axis, according to the following expressions,

$$\theta + \theta' = 90^\circ$$

(2.35)

$$\alpha + \alpha' = \theta + \theta' + 270^\circ$$

(2.36)

Hence, the following relations can be readily derived

$$s'_I = s_Q$$

(2.37)

$$s'_Q = s_I$$

(2.38)

$$e'_I = -e_Q$$

(2.39)

$$e'_Q = -e_I$$

(2.40)

Figure 2.19  LUT symmetry with respect to the $45^\circ$ axis.
which further reduces the total memory requirements. The total memory reduction of a factor of eight would result in approximately 590 Kbits—still a very large number, but not prohibitive for many applications.

An improved approach, using a one-dimensional LUT, was proposed in [27]. The LUT stores the amplitude-scaling factor $\sqrt{\frac{r_{\text{max}}^2}{r^2(t)}} - 1$ — the equivalent VGA gain in (2.33) and the table entry is addressed by the input power $r^2(t)$. The real and imaginary parts of the quadrature signal $e(t)$ are obtained by rotating the input vector $s(t)$ by $90^\circ$ and multiplying it by the addressed scaling factor. Simulation can be used to minimize the table resolution (the number of table entries) and precision (bit length of the table entry) for the required linearity. As an example, less than 14 Kbits of memory was required to achieve $-70$-dBc ACI with two Class F power amplifiers. The huge reduction in the required memory is achieved at the expense of an increased algorithmic overhead, for now several multiplication instructions are required for each processed sample.

2.5 Path Imbalance and Its Effects on Linearity

The path imbalance is a manifestation of the gain and phase mismatch between the two parallel power amplifier paths. This imbalance mainly comes from the mismatch between the gain and phase characteristics of the two power amplifiers, possibly arising from the impedance-matching and biasing circuits as well as mismatches in the individual transistors themselves. The nonideality of the power combiner and the misalignment of the metal wires and transmission lines also contributes to the path imbalance. When the matching components are realized on an integrated circuit with lithographic techniques, a variety of effects can cause the effective size of the components to be different from the size of the layout mask. A matching error of 0.1 to 1% can be achieved with careful layout. Matching accuracy becomes worse with discrete components, and at best 3 to 5% matching can be achieved. Due to the constant envelope feature of the power amplifier driving signals, the path imbalance is straightforward to analyze in terms of the amplitude and phase difference between the two amplified component signals. The path imbalance has been investigated in [28, 29] and needs to be corrected for most practical applications, since it contributes directly to ACPR and EVM degradation. This section derives a closed-form expression for the out-of-band spectral regrowth for a two-tone test. A simple means to estimate the ACI for a linearly modulated signal is also given.
2.5.1 Two-Tone Linearity Analysis of an Outphased Amplifier with Path Mismatch Effects

The classic two-tone test is the simplest and yet most effective method to evaluate the nonlinearity of an amplifier and can illustrate both magnitude and phase distortion. The two-tone test is almost universally accepted, although alternative techniques such as white noise or multicarrier test signals are more suitable for characterizing the nonlinearities in digital modulation transmitters. During the two-tone test, two equal-amplitude sinusoidal signals are applied to the amplifier, and the signal envelope experiences complete variation between zero and the maximum value, by means of which the amplifier characteristic is examined. Because of the zero-crossing envelope, the ACI performance excited by two-tones is generally worse than the ACI generated by some linearly modulated signal with modest envelope variation. If the two signal frequencies are assumed to be at frequencies $\omega_0 \pm \Delta \omega$, we have the complex input signal

$$s(t) = Ae^{j(\omega_0 - \Delta \omega)t} + Ae^{j(\omega_0 + \Delta \omega)t}$$  \hspace{1cm} (2.41)

where $A$ is the amplitude of each tone. The instantaneous envelope of the input two tone is given by

$$r(t) = \sqrt{s(t)s^*(t)}$$

$$= 2A|\cos \Delta \omega t|$$  \hspace{1cm} (2.42)

The two-tone signal level should be arranged such that the peak envelope power (PEP) is equal to that of the full-power rating of the amplifier, \textit{i.e.}, $r_{\text{max}} = 2A$ in this case. Hence the ideal quadrature signal $e(t)$ is calculated by

$$e(t) = js(t)\sqrt{r_{\text{max}}^2/r^2(t)} - 1$$

$$= js(t)|\tan \Delta \omega t|$$  \hspace{1cm} (2.43)

Now, we introduce the effects of path mismatch on the amplifier performance. We assume that the gain and phase delay of the power amplifier branches are $G_0$ and $\phi_0$ and that the gain and phase imbalance of the bottom amplifier branch with respect to the upper one are $\Delta G/G_0 << 1$ and $\Delta \phi << 1$, respectively. The final output signal would be the desired signal plus an interference term
\[ S(t) = G_0 e^{i\phi_0} [s(t) - e(t)] + G_0 \left( 1 + \frac{\Delta G}{G_0} \right) e^{i(\phi_0 + \Delta \phi)} [s(t) + e(t)] \]
\[ \approx 2G_0 e^{i\phi_0} s(t) + G_0 e^{i\phi_0} \left( \frac{\Delta G}{G_0} + j\Delta \phi \right) e(t) \]  
(2.44)

The quadrature signal \( e(t) \) is a wideband signal, and its spectrum in the two-tone case can be easily calculated. Rearranging \( e(t) \), we have

\[ e(t) = 2A \cos \Delta \omega t \left| \tan \Delta \omega t \right| e^{j\omega_0 t} \]
\[ = r_{\text{max}} e_s(t) e^{j\omega_0 t} \]  
(2.45)

where

\[ e_s(t) = \cos \Delta \omega t \left| \tan \Delta \omega t \right| \]  
(2.46)

Hence \( e(t) \) can be regarded as a sinusoidal signal modulated by a slow variation term \( e_s(t) \). Taking the Fourier series expansion of \( e_s(t) \), we have

\[ e_s(t) = \sum_{n=-\infty}^{\infty} a_n e^{jn\Delta \omega t} \]  
(2.47)

where the Fourier series coefficient is given by

\[ a_n = \frac{1}{T} \int_{-T}^{T} e_s(t) e^{-jn\Delta \omega t} dt \]
\[ = \frac{2 \sin(n\pi/2) [n - \sin(n\pi/2)]}{\pi} \frac{1}{n^2 - 1} \]  
(2.48)

Obviously all even-order terms remain zero, and only the odd-order terms are left. Both out-of-band spurs and inband distortion are generated—the \( n = \pm 1 \)-order terms are inband distortion, and the rest of the \( n = \text{odd} \)-order terms are out-of-band interference. A close examination reveals that the periodic 180° phase discontinuity of the two-tone signals as the envelope reaches zero is responsible for the out-of-band spurs. Clearly, the most significant spurs occur at \( \omega_0 \pm 3\Delta \omega \), in analogy to intermodulation distortion in a weakly nonlinear amplifier.

In a two-tone test, the ACI, or out-of-band rejection, is the ratio of the magnitude of the most significant spurs to the magnitude of the desired signals. In this case, ACI is given by
\[
ACI = 20 \log_{10} \left[ \frac{1}{\pi} \sqrt{\left( \frac{\Delta G}{G_0} \right)^2 + \Delta \phi^2} \right]
\]
\[
= 10 \log_{10} \left[ \left( \frac{\Delta G}{G_0} \right)^2 + \Delta \phi^2 \right] - 9.9 \text{ (dB)}
\] (2.49)

Figure 2.20 displays the spectrum of the outphasing system output, in the presence of 2° phase imbalance and 0.3-dB gain imbalance. It is interesting to note that the adjacent two spurs have the same magnitude and opposite sign.

### 2.5.2 ACI Estimation with Gain and Phase Mismatch

In an outphasing system, the band-limited input is decomposed into two component signals, then amplified and recombined. In the presence of gain and phase imbalance, the combined output would be
\[ S(t) = S_1(t) + S_2(t) \]

\[ = [s(t) - e(t)] + \left( 1 + \frac{\Delta G}{G_0} \right) e^{j\Delta \phi} [s(t) + e(t)] \]

\[
\begin{align*}
\text{desired term} & = \left[ 1 + \left( 1 + \frac{\Delta G}{G_0} \right) e^{j\Delta \phi} \right] s(t) + \left\{ \left( 1 + \frac{\Delta G}{G_0} \right) e^{j\Delta \phi} - 1 \right\} e(t) \\
\text{error term} & = \left\{ \left( 1 + \frac{\Delta G}{G_0} \right) e^{j\Delta \phi} - 1 \right\} e(t)
\end{align*}
\]

(2.50)

The signal has been normalized to \( G_0 e^{j\phi_0} \)—the gain and phase characteristics of the upper amplifier branch. From (2.50), it is clear that the gain and phase imbalance modify the magnitude and phase of the output signal. The first term in (2.50) constitutes the desired output signal with slightly modified amplitude and phase, while the second term creates both the out-of-band interference and inband distortion. Equation (2.50) can be simplified and rearranged in the case of a small amount of the gain and phase imbalance.

\[ S(t) \approx 2s(t) + \left( \frac{\Delta G}{G_0} + j\Delta \phi \right) e(t) \]  

(2.51)

The power spectrum is thus the weighted sum of the desired spectrum of \( s(t) \) and the spectrum of \( e(t) \). The spectrum of the quadrature signal \( e(t) \) is difficult to calculate and strongly dependent on the modulation—knowing the bandwidth of the source signal is not adequate to predict the bandwidth of the quadrature signal. The quadrature signal is generally wideband, and its spectrum extends far into adjacent and alternate channels [30]. The second term in (2.51) is evidence of the incomplete cancellation of the quadrature signal during power combining, and it is responsible for the out-of-band spectrum regrowth. Note that, in the case of small amount of gain and phase imbalance, the inband distortion generated by this term hardly affects the spectrum characteristics of the desired signal. Thus, a simple way to estimate the out-of-band rejection is to do a simulation of the quadrature signal and compute its spectrum first, then compare it to the spectrum of the source signal and obtain the ratio of the peak spectral density outside the channel of the quadrature signal spectrum to the inband spectral density of the source signal—denoted as \( P_e \). Finally, the ACI of the output signal can be approximated by
ACI = 20 \log_{10} \left[ \frac{1}{2} \sqrt{\left( \frac{\Delta G}{G_0} \right)^2 + \Delta \phi^2 \times P_e} \right] \\
= 10 \log_{10} \left[ \left( \frac{\Delta G}{G_0} \right)^2 + \Delta \phi^2 \right] + P_e - 6.0 \quad (\text{dB}) \quad (2.52)

Figure 2.21 displays the spectrum of the desired signal as well as the quadrature signal for various modulations, filtered with the square-root raised cosine with a roll-off factor of 0.35. The sampling rate was 16× higher than the symbol rate. The inband power spectrum density has been normalized to 0 dB, and the spectra of the quadrature signals are offset accordingly. It can be seen that for the quadrature signal, a significant portion of the power extends into adjacent and alternate channels. One aspect of this is that an input signal with a wider range of envelope variations—such as 16-QAM—results in a lower PAP ratio on the quadrature signal. The result of this is a higher inband distortion power for the quadrature signal. As we can see, $P_e$ is roughly $-9$ dB for QPSK, OQPSK and $\pi/4$-DQPSK, and $-6$ dB for 16-QAM. With this value in mind, it is easy to estimate the ACI of

![Figure 2.21 Simulated spectrum for $s(t)$ and $e(t)$ for various modulations.](image-url)
the outphasing system with the knowledge of the path imbalance. Compared to the two-tone case, we see that generally the ACI for bandlimited signals is 2–5 dB better than that for the two-tone case, depending on the modulation.

Figure 2.22 shows similar spectra for the CDMA IS-95 uplink. The modulation here is OQPSK. Two spectra with different matching conditions are also illustrated for comparison. One spectrum is obtained under a 10% gain imbalance and a 10° phase imbalance, and the other is obtained under a 3.5% gain imbalance and a 2° phase imbalance. Their spectra are normalized to have the same peak inband density. According to (2.52) the ACI with mismatching is \(-25\) dBc and \(-38\) dBc, respectively, which is in good agreement with the simulation results.

2.6 Effect of Quadrature Modulator Errors on Linearity

Modern mobile systems frequently use a quadrature modulator to generate complex vector modulation. To translate the complex baseband waveform to
the desired carrier frequency band, two general topologies are used in a transmitter—indirect and direct upconversion. An indirect upconversion system implements the modulation at an IF and then translates that result to a higher frequency at the RF, requiring a local oscillator (LO) and IF filtering. A direct upconversion system performs the data modulation and frequency translation at the same time, hence eliminating the IF block. The resultant more compact structure and reduced power consumption is suitable for monolithic integration, but care must be taken to keep the resulting spurious signals outside the band of interest.

The block diagram of a typical quadrature modulator is shown in Figure 2.23, in which the balanced baseband I/Q waveforms modulate two carriers driven in quadrature phase, and the resultant signals are summed. Various factors and implementation imperfections contribute to the degradation of the resulting modulation accuracy, and the effects can be collectively represented by the gain error, phase error, and LO leakage [31]. The quadrature modulator gain error describes the gain mismatch between the I and Q channels. The phase error mainly comes from the imperfection of the 90° phase shifter. The LO-RF feedthrough causes LO leakage. Besides, the DC offsets in the I/Q channels give rise to a similar effect to LO leakage, and hence are ascribed to that term. One way of visualizing the impairment of quadrature error is to consider a baseband sinusoidal single-tone input. The modulator output trajectory in the complex I/Q plane becomes an ellipse—rather than an ideal circle—in the case of gain and phase error. The phase error further rotates the principal axes of the ellipse by 45°. The LO leakage term simply shifts the whole trajectory by an equal amount. By use of the complex envelope notation in [31], the complex envelope of the modulated RF signal can be derived as follows.
\[
R(t) = \frac{1}{2}(1 + C)A(t) + \frac{1}{2}(1 - C)A^*(t) + p
\]  
(2.53)

where \(A(t)\) is the complex baseband waveform and \(A^*(t)\) is the complex conjugate. The constant \(p\) is a complex number to represent the LO leakage term. The constant \(C\) characterizes the gain and phase mismatch of the quadrature modulator in the Q channel with respect to the I channel; that is,

\[
C = (1 + g)e^{j\delta}
\]  
(2.54)

where \(g\) is the gain error and \(\delta\) is the phase error. In the case of the small quadrature error, (2.53) approximates to

\[
R(t) = \left[ 1 + \frac{1}{2}(g + j\delta) \right] A(t) - \frac{1}{2}(g + j\delta)A^*(t) + p
\]  
(2.55)

The first term is the desired signal with a small amplitude and phase modification. The second term is the spurious/image signal, which represents the counterclockwise rotating version of the input. The image rejection ratio is thus given by [31]

\[
IR \approx 10\log_{10}(g^2 + \delta^2) - 6.0 \quad (dB)
\]  
(2.56)

### 2.6.1 Quadrature Modulator Error Minimization

The quadrature modulator error generates a small variation in the amplitude and phase of the modulated signal. When this signal passes through a nonlinear power amplifier, intermodulation terms are generated, and their spectra expand into adjacent channels, causing interference to other users [32]. This effect is readily seen by taking the cube in (2.53) and collecting the \(A^3\) term and \((A^2A^*)\) term. The leakage power is proportional to the square of the quadrature errors. The improvement of quadrature error is reduced when a digital predistorter is included preceding the quadrature modulator. Even a relatively small amount of quadrature error compromises the improvement gained through predistortion [32, 33]. Quadrature modulator errors have a similar detrimental effect to outphasing power amplifiers; the details will be covered in Section 2.6.2.

Great efforts have been made to reduce the quadrature modulator errors in different aspects. One important research activity applies to
the generation of quadrature LO signals. An RC-CR network is the simplest quadrature signal generator, as shown in Figure 2.24. It is easily proven that the two output signals have 90° phase difference at all frequencies but have equal amplitude only at the pole frequency $\omega = 1/RC$. A polyphase filter, shown in Figure 2.25, is often used as a broadband 90° phase shifter [34]. The number of filter stages is determined by the required bandwidth. This structure is less sensitive to the variations of the component values compared to the simple RC-CR network. One disadvantage is the increased thermal noise. An alternative approach is RC-CR pole frequency tuning, in which varactor diode capacitors or transistor junction capacitors are changed by an appropriate control voltage. Havens’ quadrature circuit exploits the fact that the two diagonals of a rhombus are perpendicular to each other [35]. In other words, the summation and difference of two equal-amplitude vectors always have a 90° phase difference, as illustrated in Figure 2.26. The limiters ensure that the two input and resulting signals have equal amplitude; otherwise a phase error will result. It is easy to show that the amplitude mismatch in percentage generates approximately the same amount of phase error in radians. For example, every 0.1-dB amplitude error contributes to a 0.7° phase error. The quadrature oscillator is another way to generate quadrature LO signals with equal amplitude [36].

Despite those efforts, however, the quadrature modulator error is expected to change with environmental variation, channel frequency, temperature, component aging, supply voltage, and biasing. Hence, some sort of continuous tracking and compensation may be necessary to achieve exceptional linearity performance. Several schemes have been proposed to systemically track and correct the quadrature modulator error, including the gain and phase mismatch and carrier leakage [37–40]. These methods use an envelope detector, which checks the modulation output, as a feedback means

Figure 2.24 Quadrature signal generation with RC-CR network.
to guide the adaption of the digital correction circuit. With the knowledge of quadrature error, the compensation is accomplished by simply adjusting the amplitude, phase, and dc offset of the baseband I and Q signals. The estimation of the quadrature error is based either on the predefined test vectors [37, 39, 40] or on the comparison of the modulation data and the envelope feedback [38, 40], as shown in Figure 2.27(a, b), respectively. The requirement of training in the former methods restrains their application, although they provide relatively simple adaption. The latter methods enable background operation, by use of Newton-Raphson algorithm or least squares algorithm. In addition to their computational load, the algorithm complexity grows greater when quantization error and loop delay have to be taken into account.

Figure 2.25 A two-stage polyphase filter [34].

Figure 2.26 Havens’ quadrature signal generator [35].
The effect of quadrature modulator error on outphasing power amplifiers is intrinsically different from that of the same error on digital predistorters. In the case of a digital predistorter, the quadrature modulator generates a slightly scaled version of the desired signal and a small amount of counter-rotating signal. Before the modulated signal passes through the nonlinear amplifier, the counter-rotating signal has no effect on the spectral leakage to adjacent channels, and the spectrum is not expanded although the EVM is increased. The excess amount of spectral regrowth results from the quadrature modulator error, in conjunction with the amplifier nonlinearity. By contrast, in outphasing power amplifiers, the band-limited waveform is spectrally expanded prior to the quadrature modulator, and the out-of-band rejection relies on the precise cancellation of two equal errors. Consider a simple outphasing system with the power amplifier removed. By taking

![Diagram](image-url)

Figure 2.27 (a) Quadrature modulator error correction by test vectors, and (b) with background adaption.

### 2.6.2 Quadrature Modulator Error Effects on Outphasing Systems

The effect of quadrature modulator error on outphasing power amplifiers is intrinsically different from that of the same error on digital predistorters. In the case of a digital predistorter, the quadrature modulator generates a slightly scaled version of the desired signal and a small amount of counter-rotating signal. Before the modulated signal passes through the nonlinear amplifier, the counter-rotating signal has no effect on the spectral leakage to adjacent channels, and the spectrum is not expanded although the EVM is increased. The excess amount of spectral regrowth results from the quadrature modulator error, in conjunction with the amplifier nonlinearity. By contrast, in outphasing power amplifiers, the band-limited waveform is spectrally expanded prior to the quadrature modulator, and the out-of-band rejection relies on the precise cancellation of two equal errors. Consider a simple outphasing system with the power amplifier removed. By taking
the quadrature modulator gain and phase mismatch into account, the output signal can be described by

\[
S_o(t) = \frac{1}{4}(1 + C_1)S_1(t) + \frac{1}{4}(1 - C_1)S_1^*(t) + \frac{1}{4}(1 + C_2)S_2(t) + \frac{1}{4}(1 - C_2)S_2^*(t)
\]

\[
= s(t) + \frac{1}{4}(\alpha_1 + \alpha_2)(s - s^*) - \frac{1}{4}(\alpha_1 - \alpha_2)(e - e^*)
\]  

(2.57)

where

\[
\alpha_{1,2} = C_{1,2} - 1 \\
\approx g_{1,2} + j\delta_{1,2}
\]  

(2.58)

The second term—\((\alpha_1 + \alpha_2)(s - s^*)/4\)—will not contribute to the out-of-band spectrum because it is linearly proportional to \(s\). However, since the quadrature modulator errors generally cannot simply cancel out, the third term in (2.57) is responsible for the spectral regrowth even without the presence of a nonlinear power amplifier. The leakage power is proportional to the square of the quadrature modulator error. This simple analysis shows that quadrature modulator error probably has more detrimental effects on outphasing power amplifier performance than the same error has on other linearization techniques.

A quadrature modulator error degrades the constant envelope characteristics of the two component signals in an outphasing system. While the power amplifier may be driven into saturation for maximum efficiency, the variation of the signal envelope nevertheless creates a devastating effect on the overall system performance. The linearity degradation is dependent upon both quadrature modulator error and amplifier nonlinearity. The envelope variation can be estimated by examining (2.53). It can be shown that in the case of small quadrature gain and phase mismatch, the envelope of the first two terms in (2.53) is bounded by

\[
\left(1 + \frac{1}{2}g - \frac{1}{2}\sqrt{g^2 + \delta^2}\right)|A| \leq |(1 + C)A + (1 - C)A^*| \\
\leq \left(1 + \frac{1}{2}\delta + \frac{1}{2}\sqrt{g^2 + \delta^2}\right)|A|
\]  

(2.59)
Hence, the worst-case envelope variation in the presence of quadrature gain and phase mismatch is given by

$$\frac{\Delta R}{R} \leq \frac{1}{2} \sqrt{\delta^2 + \rho^2} + |\rho|$$

(2.60)

As an example using (2.60), with 0.3-dB gain error, 2° phase error, and −30-dBC LO leakage, we get an envelope variation of roughly ±6% or ±0.5 dB. Such an envelope fluctuation may or may not cause the power amplifier to generate a significant amount of spectral regrowth, depending on the selection of the power amplifier. If the amplifier generates little spectral regrowth due to input envelope variation, then the overall spectral regrowth of the system may be estimated with (2.57).

A semianalytical treatment is given in [41], which elaborates on the coupled effects of the quadrature modulator error, path imbalance, and the power amplifier nonlinearity. This method originates from [32] in the case of constant envelope modulation such as GMSK. The procedure expresses the amplifier nonlinearity into a memoryless Taylor series, substitutes the modulated signal for each amplifier branch, sums the branches together, and finally identifies the dominant distortion terms. The contribution of each term is calculated and summed together. Since the power amplifier operates at rather modest input dynamic range—±0.5 dB in the previous example—it is straightforward and sufficient to represent it with a power series by the measured AM-AM and AM-PM characteristics. A power series expansion to the fifth order is usually sufficient. Most digital modulations are based on some sort of symmetrical quadrature constellation; as a result many cross-correlation products become zero. Nonzero crosscorrelation terms are typically not dominant, unless they are almost the same size and no other terms dominate. By ruling out almost all the cross terms, the spectrum is simply the power addition of each weighted term, and the result is usually close to the real simulation within 1 dB. The paper shows a strong detrimental effect of quadrature modulator error—with a 0.1-dB gain error, 0.3° phase error, and a −40-dBC carrier leakage, which represents the state-of-the-art technology available today; the resultant spectral regrowth with a Class C power amplification is around −50 dBC, still rather high.

Two points should be addressed. First, in contrast to the band-limited baseband inputs applied to a quadrature modulator in the usual case, the quadrature modulator in an outphasing system is producing a constant envelope output signal. This may result in a simpler approach for tracking and correcting quadrature modulator errors than is applicable in the more
general case. Second, since the information is carried on the zero-crossing of two component signals, two limiters may be used to flatten the envelope fluctuation due to quadrature modulator error, before the signals are sent to the power amplifiers. The limiter used must have sufficient bandwidth so as not to introduce extra AM-PM distortion. The limiter approach may result in improved spectral purity with little added expense. The use of switching-mode power amplifiers such as Class D or Class E in replacement of the saturated Class B or Class C amplifier may benefit from a similar rationale.

2.7 SCS Quantization Error Effects on Outphasing Systems

The rapid evolution of modern DSP technology makes it an attractive candidate for the implementation of the SCS in an outphasing power amplifier system. With this approach, the signal separation is done digitally in baseband and converted to analog waveforms prior to upconversion and power amplification. Because of the finite word-length representation of the data stream in the quantizer, quantization noise is inevitably generated. It is important to understand this mechanism and how the quantization noise affects the linearity performance of the outphasing system. By use of application specific DSPs (ASDSPs), the algorithm of the arithmetic blocks and registers can be minimized, and the word length can be optimized to reduce the hardware requirements. This will result in reduced power consumption and cost, and increased bandwidth.

During signal component separation, the quadrature signal is being added to, and subtracted from, the source signal to obtain two constant-envelope component signals. A direct consequence of a DSP implementation of this process is that both the source signal and the quadrature signal are quantized. The source signal is quantized prior to the SCS; thus the amplifier output is a scaled replica of the quantized source signal and the quantization noise sets the final output noise floor. The quantization of the quadrature signal has a different mechanism that gives rise to a random envelope ripple on the power amplifier driving signal. This effect is further enhanced by the nonlinear power amplifier through the AM-AM and AM-PM conversions, which are not canceled. Consequently, the overall SNR is further degraded.

2.7.1 Error Effects of Quantization of the Source Signal

The quantization error of a complicated signal is generally considered to be a stationary random process and is uncorrelated with the original signal [42].
The probability density is uniformly distributed over the range of quantization error. Consequently, the quantizer can be modeled as an additive white noise source. The SNR after a $B_s$-bit complex quantizer is given by [43]

$$\text{SNR}_s = 6.02B_s + 1.76 - \text{PAP} \quad (\text{dB}) \quad (2.61)$$

where the PAP can be calculated from the probability density function,

$$\text{PAP} = -10 \log_{10} \left[ \int_0^1 \rho(r)r^2 dr \right] \quad (\text{dB}) \quad (2.62)$$

where $\rho$ is the probability density function of the modulation and $r$ is the normalized signal amplitude. Note that the normalized average signal power is equal to the inverse of the PAP of the signal. The noise power in (2.61) is doubled in a complex quantizer, due to the fact that the quantization takes place independently on the real and imaginary parts of the source signal.

Considering that the signal power is distributed within the bandwidth of $B$, which is the single-sided equivalent noise bandwidth of the modulation, while the additive white noise power is distributed up to the bandwidth of half sampling rate, the following expression is readily proven [43]

$$\text{ACI}_s = \frac{2B}{f_s \cdot \text{SNR}_s} \quad (2.63)$$

where $f_s$ is the sampling rate. The equivalent noise bandwidth is dependent on the modulation and baseband pulse shaping. As shown previously, the noise bandwidth of a modulated signal with square-root raised cosine filtering is equal to half of the symbol rate. According to (2.63), for every one-bit increase in word length, the SNR and hence ACI will be improved by 6 dB, while doubling the sampling frequency improves ACI by 3 dB.

### 2.7.2 Error Effects of Quantization of the Quadrature Signal

When the quadrature signal $e(t)$ is quantized, the quantization error $\Delta e$ seems to cancel out during the signal combining, as illustrated in Figure 2.28. However, since nonlinear power amplifiers are used instead to improve the power efficiency, the quantization error results in a variation of the magnitude of the signal applied to the power amplifiers, and consequently, the power amplifier gain and phase characteristics are changed due to AM-AM
and AM-PM conversion. Assuming the changes of the gain of the two amplifiers are $\Delta G_1$ and $\Delta G_2$, respectively, the amplified output signals are

\[
\hat{S}_1 = (s - e - \Delta e)(G + \Delta G_1) \\
\hat{S}_2 = (s + e + \Delta e)(G + \Delta G_2)
\] (2.64)

where $G$ is the gain of the amplifiers evaluated at the constant envelope $r_{\text{max}}$ without quantization. The gain $G$ is a phasor, including both the AM-AM and AM-PM conversion. The quantization error $\Delta e$ causes both the amplitude deviation and phase deviation of the component signal from the desired value. The effect of the amplitude ripple is enhanced by the nonlinear power amplifiers. However, the amplitude ripple and the phase ripple directly caused by the quantization error $\Delta e$ cancel out in the first order. This effect is evident from the combined output without the explicit term of $\Delta e$,

\[
\hat{S}_{\text{out}} = \hat{S}_1 + \hat{S}_2 \\
\approx 2G \cdot s + (\Delta G_1 + \Delta G_2) \cdot s - (\Delta G_1 - \Delta G_2) \cdot e
\] (2.65)

The error term is evaluated by assuming that the quantization error is uniformly distributed. Then the average noise power is given by [43]

\[
\sigma_e^2 = \frac{1}{3} \cdot \frac{1}{2^{2\beta}} \int_0^1 \rho(r)|G'(r)|^2\{1 - \cos[4 \arccos(r)]\}dr
\] (2.66)
and the signal power is
\[ \sigma_s^2 = \int_0^1 \rho(r)|2rG(r)|^2 dr \] (2.68)

The SNR due to quadrature signal quantization becomes
\[ \text{SNR}_q = \frac{\sigma_s^2}{\sigma_e^2} \] (2.69)

Again assuming that the quantization error appears as white additive noise, the resulting ACI is given by
\[ \text{ACI}_q = \frac{2B}{f_s \cdot \text{SNR}_q} \] (2.70)

In a similar manner to the source signal quantization, for every one-bit increase in the word length representation, the SNR and hence ACI will be improved by 6 dB. Doubling the sampling frequency improves ACI by 3 dB.

The quantization error is generally a small fraction of the signal magnitude. For example, the amplitude variation will be $\pm 0.07$ dB after a 7-bit quantizer. With this range of ripple, the power amplifier AM-AM and AM-PM distortion can be well-approximated by a straight line. Assuming that the amplifier has a complex gain $G_r(r)e^{j\theta(r)}$, we can calculate
\[ |G'(r)|^2 = G_r^2 + G_r^2\theta^2 \] (2.71)

The quantization noise power is then approximated by
\[ \sigma_e^2 = \frac{1}{3 \cdot 2^{2B}} |G'(r)|^2 \int_0^1 \rho(r)(r^2 - r^4) dr \] (2.72)

and the SNR is
\[ \text{SNR}_q = \text{SNR}_s - 10 \log_{10} \left[ \left( \frac{G_r}{G_r} \right)^2 + \theta^2 \right] - 10 \log_{10} \int_0^1 \rho(r)(r^2 - r^4) dr \] (dB) (2.73)

The SNR could be infinity in the case of perfectly linear power amplifiers. This result shows that it is important to design power amplifiers with nearly linear AM-AM and nearly flat AM-PM characteristics near the operating...
point, such that the quadrature signal quantization will not degrade the overall noise performance. Switching-mode power amplifiers may perform better than the classical saturated power amplifiers in this regard, as long as the bandwidth is sufficiently large. It can be shown that the SNR in this case is given by

$$\text{SNR}_q = \text{SNR}_s - \text{PAP} - 10 \log_{10} \int_0^1 \rho(r)(r^2 - r^4)dr \quad \text{(dB)} \quad (2.74)$$

### 2.8 Linearity Effects of Reconstruction Filter and DSP Sampling Rate

As shown in Figure 2.18, in a digital SCS the signal separation is accomplished inside the processor, and the resultant digital streams are fed to the digital-to-analog (D/A) converters and reconstruction filters to be transformed into analog waveforms. The sampling rate of the DSP is a crucial design parameter and should be kept as low as possible to accommodate the high modulation bandwidth and lower the power consumption. However, a low sampling rate increases the complexity of the reconstruction filter since a low filter order generally requires a high sampling rate and vice versa. Thus, these conflicting requirements must be properly traded off.

The effect of the D/A conversion in the absence of the quantization error is equivalent to the simple block diagram shown in Figure 2.29, in which the original analog waveform is multiplied by an impulse train, followed by a zero-order hold circuit. The sampling images are generated during the impulse train modulation. The zero-order hold circuit has a \( (\sin x/x) \) response. The sampling images are the periodically repeated copies of the desired spectrum with frequency shifted by integer multiples of the sampling rate. The sampling images are considered to be interference terms and shall be eliminated by the reconstruction filters. In addition, these images

![Figure 2.29](image-url) Effects of impulse train modulator and zero-order hold.
and interference terms are attenuated by the sinc response of the zero-order hold circuit in the D/A converter. Thus, the spectrum of the two component signals after passing the reconstruction filters can be described by [44]

\[ S_{1,2}(f) = \sum_{k=-\infty}^{\infty} H_z(f) \cdot H_s(f) \cdot S_{1,2}(f + kf_s) \tag{2.75} \]

where \( H_s(f) \) is the frequency response of the reconstruction filters, \( f_s \) is the sampling rate, and \( H_z(f) \) is the “sinc” response of the zero-order hold circuit,

\[ H_z(f) = \frac{1}{f_s} e^{j\pi f / f_s} \cdot \text{sinc}(\pi \frac{f}{f_s}) \tag{2.76} \]

Even though aliasing of the wideband quadrature signals could occur during the D/A conversion, the signal recombining still results in the complete cancellation of the wideband signals owing to the fact that they are ideally 180° out of phase. Now, the final output is

\[ S_{\text{out}}(f) = S_1(f) + S_2(f) \]

\[ = 2 \sum_{k=-\infty}^{\infty} H_z(f) \cdot H_s(f) \cdot s(f + kf_s) \tag{2.77} \]

Note that the sampling images of the source signal are added in-phase, after being attenuated by the reconstruction filters and the “sinc” response. Figure 2.30 shows the spectrum for the upper amplifier branch signal \( S_1(t) \). The reconstruction filter used is a Butterworth filter, and the sampling rate is 4× oversampling. The cutoff frequency of the filter is fixed at 2.5 MHz. Figure 2.31 gives the output spectrum with three different filter orders for a CDMA IS-95 waveform.

In addition to the various filter types, the sampling images may be suppressed by increasing the sampling rate, increasing the filter order, or decreasing the filter cutoff frequency. However, note that the two component signals have a much wider bandwidth than the desired signal. As the filters cut more deeply into the signal at the higher frequencies, the envelope of the two component signals will start to ripple and introduce intermodulation through the saturated power amplifiers by the mechanisms we have already discussed. As usual, this intermodulation term may be mitigated with saturated- or switching-mode power amplifiers. Both the sampling images
and intermodulation contribute to the out-of-band interference and should be kept below the specified ACI. It is generally a cumbersome task to find the optimum combination of sampling rate and filter cutoff frequency. A one-dimensional search method based on simulation was proposed in [44] and proved to be useful. This method assumes that an increase of the cutoff frequency decreases the intermodulation and that the residual spectrum near the sampling frequency is dominated by the sampling image. For a specified filter type and order, the optimal cutoff frequency was found by iterative calculation and simulation to suppress both the sampling image and intermodulation below the required ACI. Butterworth and Bessel filters were investigated; the former performs considerably better due to its sharper cutoff frequency.

2.9 Summary

The major factors contributing to the linearity degradation in outphasing power amplifier systems have been discussed. The path imbalance between
the two power amplifier branches results in incomplete cancellation of the wideband quadrature signal and hence creates both out-of-band interference and inband distortion. The misalignment of the I/Q modulators includes the gain and phase error between the I and Q channels and the carrier leakage. These factors introduce amplitude variation to the two phase-modulated signals and create intermodulation through the AM-AM and AM-PM conversion of the nonlinear power amplifiers. The quadrature modulator errors may have more detrimental effects on the linearity performance of an outphasing system than on other linearization techniques. These error effects are more troublesome, since they cannot be compensated for as easily as the path imbalance effect. However, the quadrature modulator in an outphasing system is ideally producing a constant envelope output signal, and this fact may lead to a simpler and more effective approach to track and correct the quadrature modulator errors. The linearity performance is also degraded by the digital SCS due to quantization effects, sampling, and reconstruction filtering. The quantization noise of the source signal directly adds to the output as a constant background. The quantization of
the quadrature signal causes variation of the gain and phase characteristics of the nonlinear power amplifiers and hence creates interference terms. The effects of the reconstruction filters and sampling rate have also been addressed. The filter type, order, and cutoff frequency generally need to be traded off with the sampling rate to achieve the optimum performance. In the case of the envelope fluctuation induced by the quadrature modulator errors, reconstruction filtering, and DSP sampling, two limiters with sufficient bandwidth or two switching-mode power amplifiers may mitigate the linearity impairment created by the amplifier nonlinearities.

References


