Chapter 4.1

RF Design for Wireless

Michael Steer

Abstract

Wireless technologies are and will continue to have profound influences on our society. The concept of communications anywhere anytime and the concepts of communication from person-to-person rather than handset-to-handset will revolutionize our personal relationships as well as the obvious benefits to business and to work-related efficiencies. In this chapter we will discuss aspects of RF and microwave circuit design in the context of wireless communication. The main emphasis will be on Cellular Mobile Radio systems.

4.1.1 INTRODUCTION

Wireless technology has been going through rapid technological development at a pace which is unprecedented in the technological age. This is because several electro-technologies have reached sufficient maturity within the last 5 years or so. From a technology perspective the enabling technologies for the wireless revolution are: communications theory and digital signal processing; VLSI (silicon chip) technology including design technologies; and RF technology. While all of these technologies are important, the cost reductions brought about by advances in VLSI and RF circuits, and their companion design technologies, have ignited the revolution. For example the current analog cellular phone system was fully described by AT&T in a submission to the FCC in 1971. However only recently have the VLSI and RF technologies matured sufficiently to make it all possible. Conventional consumer oriented wireless systems have been available for some time. Mobile telephone service has operated at the 40 MHz, 150 MHz and 450 MHz bands but could support only very few conversations at any one time. The most common forms of non-broadcast wireless communications have been and still are paging services operating in the 40 MHz, 150 MHz, 450 MHz and 900 MHz bands, and cordless telephones at the 45 MHz and 900 MHz bands. Over the last decade cellular radio has become increasingly pervasive.

Since World War II the demand for wireless person-to-person communication far exceeded the available capacity. For example, the wireless telephone service in large metropolitan areas could only support 10 or so two-way conversations at a time in each of the available bands. Thus in a region the size of New York city inly 30 conversations could be supported. The demand was much greater than this

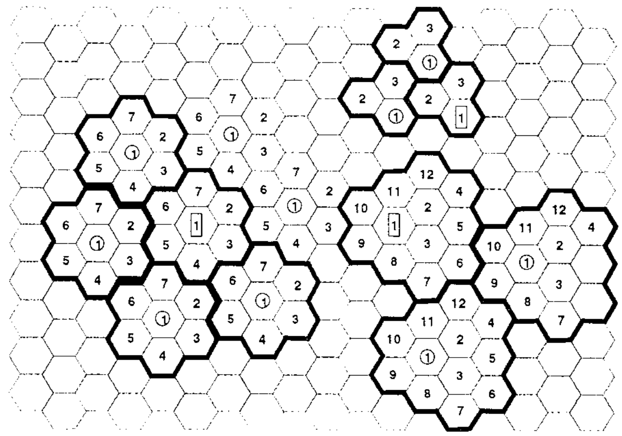

Figure 4.1.1 Cells arranged in clusters.

and was only controlled by the frustration people had in using the system, and the long waits required to be allocated one of the few subscriptions to the service. The demand was believed to be many orders of magnitude greater than could possibly be satisfied using the concepts behind the mobile telephone system. It was clear that only a radically different system could satisfy demand. The solutions arrived at are responsible for the the current "wireless" explosion. The initial cellular radio system, the Advanced Mobile Phone System (AMPS) was described by Bell Laboratories in a submission to the Federal Communications Commission in the U.S.A. In 1979 W.A.R.C. allocated the band 862–960 MHz for mobile radio leading to the FCC releasing, in 1981, 40 MHz in the 800-900 MHz band for cellular land-mobile phone service. The service was defined as

> "A high capacity land mobile system in which assigned spectrum is divided into discrete channels which are assigned in groups to geographic cells covering a cellular geographic area. The discrete channels are capable of being reused within the service area."

The key attributes here are

- High Capacity.
  Prior to the availability of the cellular system other radio systems were used. Users were not always sure of gaining access to the radio network and frequently had to make three or four attempts to gain access.

- The concept of cells.
  The idea is to divide a large geographic area into cells which, in the North American system, are 2–20 kilometers in radius. These cells are arranged in clusters and the total number of channels available are divided among the cells in a cluster and the full set is repeated in each cluster. In Figure 2.1 three, seven and twelve cell clusters are shown. The size of a cell can be reduced, and so increase system capacity, through a process called cell splitting. Here additional basestations are introduced and the basestations and mobile units are operated at lower power levels.

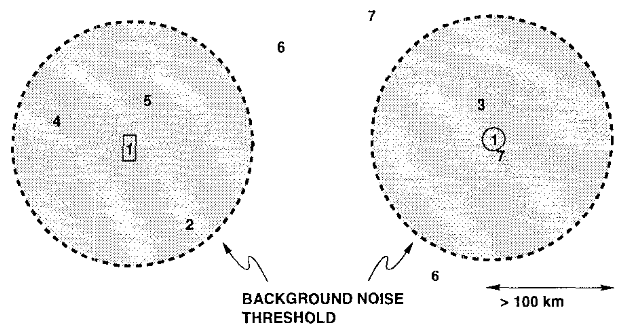

Figure 4.1.2 Interference in a conventional radio system.

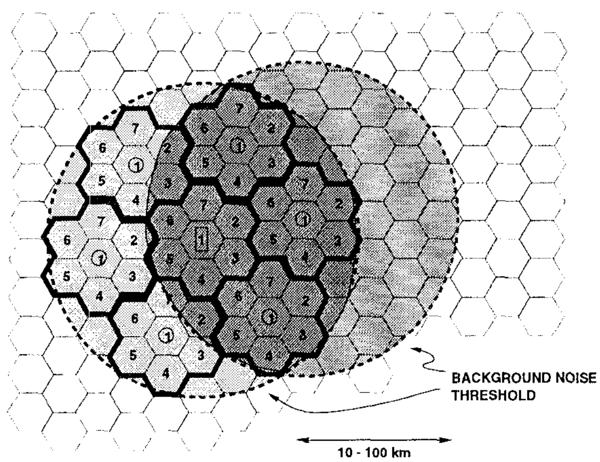

Figure 4.1.3 Interference in a cellular radio system.

- Frequency Reuse

Achieving maximum frequency reuse is essential in achieving high capacity. In a conventional wireless system, be it broadcast or the mobile telephone service, base-stations are separated by sufficient distance that the signal levels fall below a background noise threshold before the same frequencies are reused as shown in Figure 2.2. In this figure several different frequency sets are indicted by number. The signals in Frequency set #1 are separated by a sufficient distance that no signals in the left hand geographic area interfere with signals in the right geographic area. There is clearly wasted spectrum here. The geographic areas could be pushed closer to each other. Interference in the conventional systems is limited by background noise in stark contrast to a cellular system. Consider the interference in a cellular system as shown in Figure 2.3. Now corresponding cells in different clusters do interfere with each other and generally the interference is much larger than that of the background noise. Interference can also be controlled by dynamically adjusting base-station and mobile transmit power to the minimum acceptable if so required. Living with interference from corresponding cells operating at the same frequency

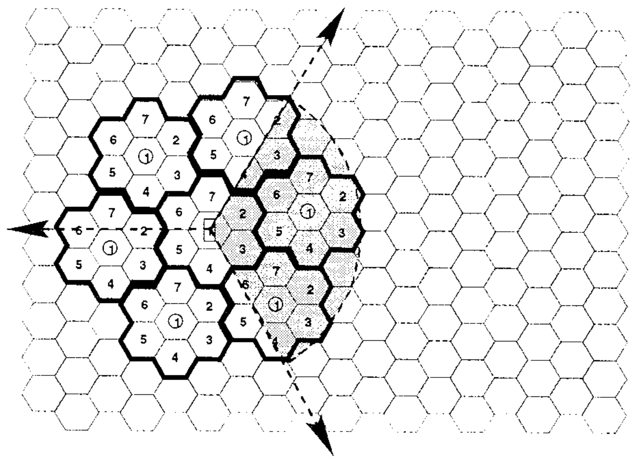

Figure 4.1.4 Interference in a cellular radio system.

in neighboring clusters (so called co-channel interference) is the the wining concept in cellular radio. The interference levels that can be tolerated vary among the various cellular systems. The AMPS system has a qualitative minimum signal-to-interference, S/I, ratio of 17 dB (about a factor of 50) and was determined via subjective tests. The seven cell clustering shown in Figure 2.3 does not achieve the the S/I required in the AMPS system. Using directional antennas at the base-station however brings the S/I up to acceptable levels. The interference pattern using a tri-sectional antenna is shown in Figure 2.4. Traffic engineering tells us that using clusters with a smaller number of cells results in greater spectrum efficiency. Thus systems that can tolerate high levels of interference, such as Qualcomm's CDMA system and and to a lesser extent other digital systems, can reuse frequency channels more efficiently. Indeed in Qualcomm's system the concept of clustering is not required.

From an RF point of view the major distinction between the various phone systems is whether they use analog or digital modulation and how the channels are partitioned. Systems using analog modulation typically use frequency modulation FM (The North American AMPS system) but phase modulation is also commonly used. The main attributes of these analog modulation schemes is that the signal amplitude is constant — only the frequency or phase varies. This means that RF transmitters can use transistor amplifiers operating in high efficiency saturating modes such as class AB, B, C and higher. (Class AB amplifiers are the most common type currently used in cellular radio handsets.) The newer and proposed systems using digital modulation use a variety of modulation schemes which combine frequency, phase and amplitude variations. When information is carried in amplitude variations so that amplitude nonlinearities can significantly affect signal integrity. This presents enormous challenges to the design of efficient transistor amplifiers.

Personal Communication Services (PCS) are the current next big challenge in wireless communications. Most activity in PCS is around 2 GHz using licensed bands. All the proposed systems are digital and the systems are very similar to cellular radio. They are really an evolution of digital cellular radio with the key

RF Designer	IC Designer
Maxwell's Equation†	Ohm's Law
AC	DC
Power	Volts
$c=1/\sqrt{\mu_0\epsilon_0}$	$V_T = kT/q$
Smith Chart, Harmonic balance	SPICE
Scattering Parameters	Y Parameters
Noise Figure in dB	Noise in $\mathrm{nV}/\sqrt{Hz}$
GaAs	Si
20-transistor ICs	1+ million transistor ICs
Spectrum Analyzer	Oscilloscope
Distributed Elements	Delay
Parasitics are Dominant	Parasitics are a Nuisance

Figure 4.1.5 RF Designers vs. IC Designers. † RF designers claim that Ohm's law is a special case of Maxwell's equations. Adapted from [7]

conceptual difference being that communication is from person-to-person whereas in cellular radio communication is from terminal-to-terminal. In some systems this is achieve by using a credit-card sized smart card to identify the user. The particular hardware or handset a user is using does not matter. So in PCS each person has a unique phone number. Some of the main features of PCS is that it will be lower cost than cellular radio both to provide service and to design hardware. Also cells are smaller and are called microcells and picocells. The PCS market will be much more competitive. Other important developing wireless applications are wireless local loop, cellular radio service, advanced cordless phone, and enhanced paging. Wireless local area networks will operate in unlicensed and licensed bands. Most activity at 2.4 GHz and 5.7 GHz and also below 1 GHz but systems at 60 GHz are being developed to support multimedia as then upwards of 150 MHz bandwidth is available.

4.1.2 RF Design

There is a big difference in the design of ASIC and RF circuits. A tongue in cheek comparison is given in Figure 2.5. There are several basic concepts that must be learnt before RF and Microwave circuit design can be appreciated or applied:

- Electromagnetic Theory
- Signal Propagation as Waves
- Power Transmission as Waves
- Electric-Magnetic field vs. Voltage-Current views of the world.
- Scattering Parameters and Smith Charts

- RF Design Methodology
- Nonlinear Phenomenon
- RF CAE tools

The importance of RF and microwave computer-aided engineering (CAE) tools cannot be underestimated. We can not build high efficiency, high yield RF circuits without simulation and automated design optimization. Simulation tools were used in RF and microwave circuit design to a greater extent earlier than they were used by in other electrical engineering disciplines. It is absolutely essential to design circuits with precise frequency characteristics and to reduce costs through optimization of designs for maximum yield. The actual performance of active circuits can vary substantially as the active devices themselves vary significantly. Because of the high frequencies and limited performance of active devices feed-back principles cannot be used to stabilize circuit performance. Many RF circuits are "open-loop" so that alternative techniques must be utilized to obtaining functioning circuits given varying processing conditions.

The use of microwave CAE tools profoundly influences design — it has been estimated that design productivity is as follows:

Design Process	Designs per Year
Cut and Try	2 – 4
CAE, beginning user	4 – 6
CAE, advanced user	15 – 20

RF and microwave design methodology can be thought of as a seven step process:

1. Problem Identification
2. Specification Generation
3. Concept Generation
4. Analysis
5. Evaluation
6. Initial Design
7. Final Design

Current RF CAE tools help with the last two aspects — the initial design and the final design. RF CAE tools are having a major impact but back of the envelope calculations are important. RF CAD tools can be used to learn the fundamentals of design. We will concentrate on the fundamentals as these are things that do not change with time. Simulation is an emerging educational technology and we will make some use of it. The primary CAE tools that are utilized are

- Circuit Simulation
  - Linear circuit simulation.
  - Nonlinear circuit simulation: Harmonic Balance, And Spice

- Yield Optimization
- Electromagnetic Field Simulators

Typical problems faced in RF design for wireless are

- Closely spaced frequency bands.
- Nonlinear circuit design.
- Fundamental knowledge of digital coding, modulation and access techniques.
- Digital sampling techniques, combined analog/RF circuit design.
- Wave propagation properties and models, especially for satellite transmission.
- Analytical and numerical field theory
- Antenna analysis and design techniques (e.g. wire antennas vs. planar antennas. Diversity techniques and adaptive systems.

The microwave design environment has changed significantly in recent years from being predominantly military to being predominantly commercial with fundamental change in character:

CHARACTERISTIC	DEFENSE	COMMERCIAL
Character	High Cost, Low Volume	Low Cost, High Volume
Market	Domestic	Global
Performance	Overkill	Just Enough
Application	General Purpose	Application Specific

The result is that business and design pressures have become those typical of the commercial electronics sector: narrow market windows; shorter product cycles; trends towards partnerships; demand for high quality product; and demand for cost reduction. RF and microwave houses have responded with the following changes: shorter product design cycles; concurrent engineering; mixed signal/technology design; and highly skilled design team. A single chip solution minimizes size and cost at the expense of performance however it has signal isolation problems. Multiple chip solution permits optimization of characteristics:

- High efficiency power amplifiers; low noise receive amplifiers; and low loss switching with power FETs
- Low current drain receive functions using E/D or other low power process
- Better control of undesired feed-through around filters.

4.1.3 Call Processing

Call processing is an important aspect of wireless communications. The guiding concept behind call processing is to provide substantial capacity through efficient frequency allocation and support maximum frequency reuse while maintaining acceptable interference levels. It is important to provide "transparent" call origination and call set-up. The idea is that a service similar to that available to plain old telephone service (POTS) customers should be provided. This requires that a

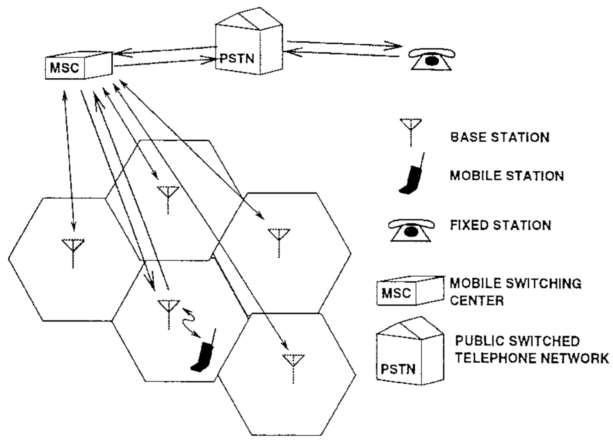

Figure 4.1.6 Depiction of call processing for mobile-to-fixed call setup.

user dials his or her own calls and receives calls without operator intervention. A user has a directory number just like a POTS subscriber. The cellular radio system must also provide full transparent mobility by providing seamless, automatic and transparent hand-off from one cell to another. More than this, call processing must provide seamless intersystem roaming. The user also expects some of the more advanced POTS services.

A call is originated from a mobile when a user enters a number and presses the send button, see Figure 2.6. If it has not already done so, the mobile scans all of the control channels and selects the strongest forward control channel as this is mostly probably associated with the basestation of the cell the user is located in. This is the process in the North American analog and digital systems. In PCS using the GSM protocol, the mobile selects the strongest channel and locks onto the control time slot. Access is requested by the mobile and, in the AMPS and NADC systems, the mobile waits for the idle bit in the forward control channel and, after a random wait, sends a request for access on the reverse control channel. The system then acknowledges the mobile by sending it a channel assignment. The mobile acknowledges, switches to the voice channel, and conversation begins.

A similar process occurs when the call originates from the public switched telephone network, see Figure 2.7. The mobile scans channels and selects the strongest forward channel and monitors the channel continuously for it to be paged. When it recognizes that it is being paged, the mobile requests access. As before the mobile waits for the forward control channel to be idle and, when this is detected, sends an access request on the reverse control channel notifying the system of its presence. Using the forward control channel, the system sends the voice channel assignment to the mobile, the mobile acknowledges and switches to the assigned voice channel. Conversation then begins.

A combination of the above procedures is used in mobile-to-mobile communication, Figure 2.8.

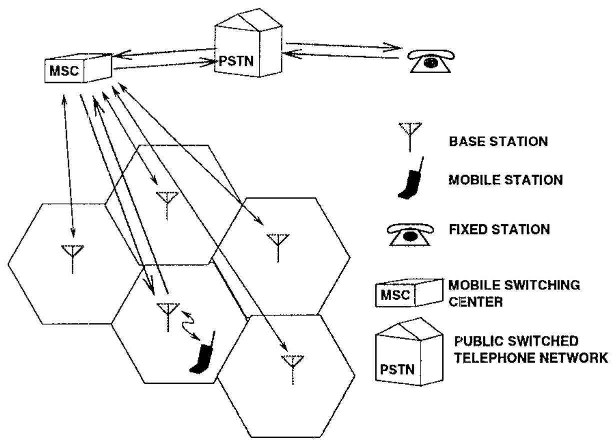

Figure 4.1.7 Depiction of call processing for fixed-to-mobile call setup.

4.1.4 Wireless Systems

Modulation schemes in wireless can be categorized as being either analog modulation or digital modulation. In analog modulation the RF signal can have a continuous range of values but in digital modulation the output has a number of prescribed discrete states. Let's consider some systems in detail. The main difference between the major wireless radio systems from an RF design point of view is whether they use analog or digital modulation and how the channels are accessed. A summary of the these attributes for the various systems is given in Table 2.1. Below is a key to the major modulation schemes.

FM	Frequency modulation
PM	Phase modulation
MSK	Minimum shift keying
GMSK	Minimum shift keying using Gaussian filtered data. Constant RF envelope (can use saturating mode amplifiers).
pi/4 DQPSK	*pi* on 4 Differential Encoded Quadrature Phase Shift Keying. Not constant RF envelope.

The FM, GMSK and PM modulation techniques produce constant RF envelopes. There is no information contained in the amplitude of the signal. Therefore errors introduced into the amplitude of the system are of no significance and so efficient saturating amplifiers such as class C can be used. In contrast the MSK and $\pi/4$ DQPSK modulation techniques do not result in constant RF envelopes and so information is contained in the amplitude of the RF signal. For these modulation techniques reasonably linear amplifiers are required. Class A amplifiers have the highest linearity but are not efficient. As a compromise class AB amplifiers are used. The various amplifier classes differ according to where they are biased. The output characteristics of a transistor together with a load line are shown in Figure 2.9. The choice of operating point is the first and most important, in terms of efficiency and linearity, of the various RF design decisions that are influenced by high level system design, in this case the choice of modulation formats.

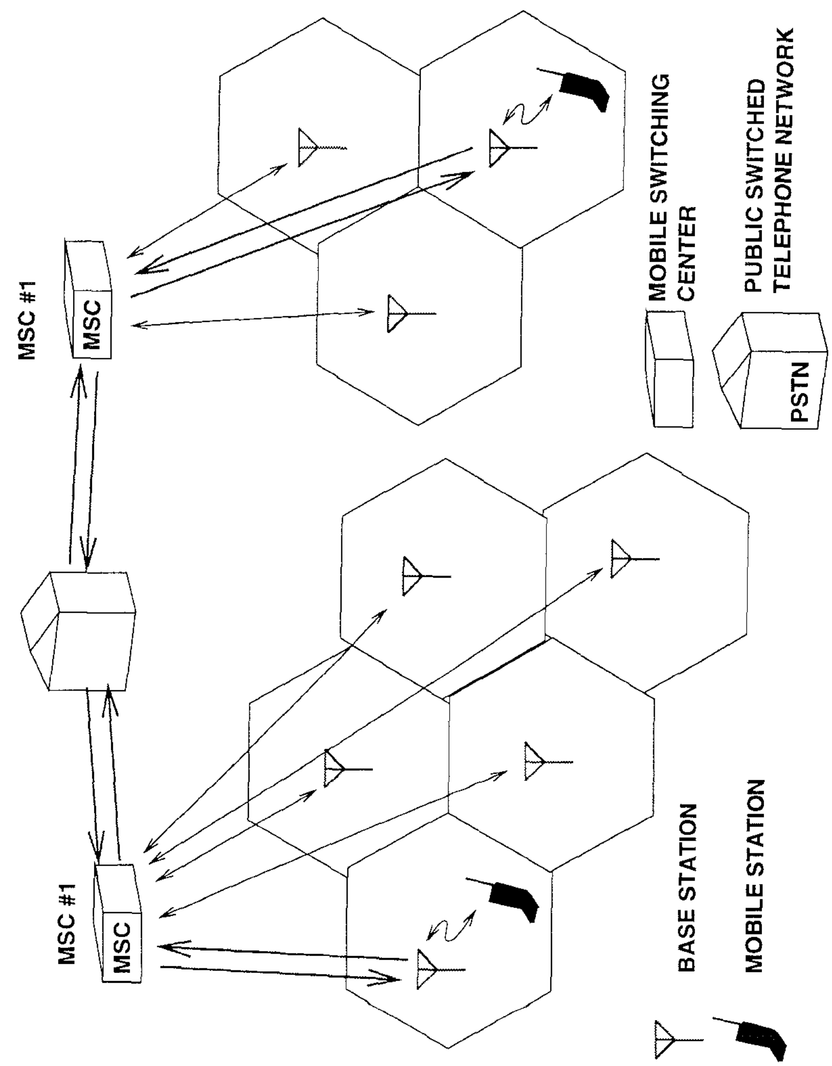

Figure 4.1.8 Depiction of call processing for mobile-to-mobile call setup.

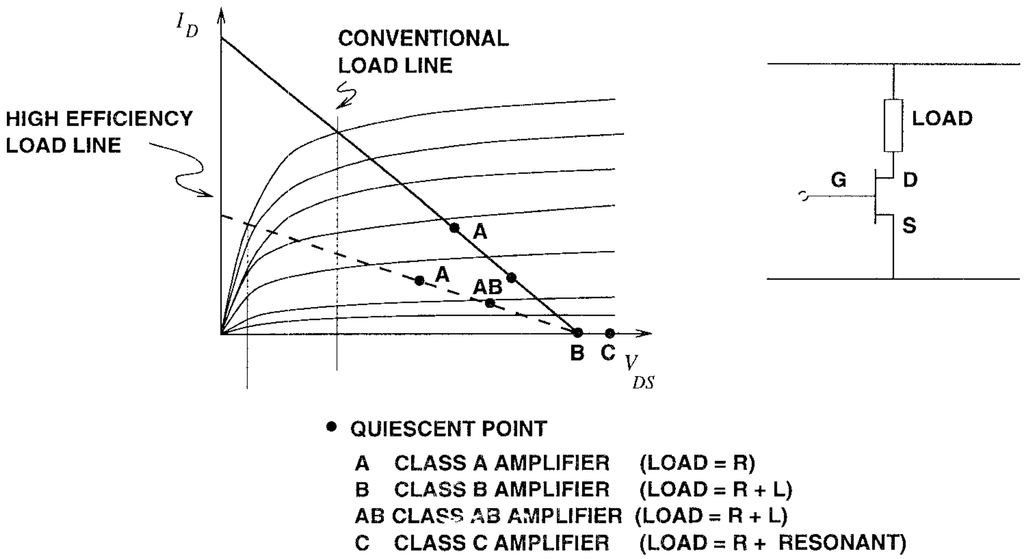

Figure 4.1.9 I-V of Power amplifier transistor showing quiescent point of various amplifier classes.

Table 1.1 Modulation schemes and channel partition approaches of the major cellular wireless systems.

SYSTEM	MODULATION SCHEME	CHANNEL ACCESS	FREQUENCY TO/FROM MOBILE	CHANNEL SPACING
AMPS	FM	FDMA	869–894/824–849	30 kHz
TACS	FM	FDMA	935–960/890–915	25 kHz
MATS-E	FM	FDMA	935–960/890–915	25 kHz
C-450	PM	FDMA	935–960/890–915	25 kHz
NMTS	PM	FDMA	463–467.5/453–457.5 935–960/890–915	25 kHz 25 kHz
NTT	FM	FDMA	870–885/925–940 860–875/915–930	25 kHz 25 kHz
GSM	GMSK	TDMA /FDMA		200 kHz
	13 kbps speech sent in 270.8 kbps bursts			
NADC IS-54	$\pi/4$ DQPSK	TDMA /FDMA	869–894/824–849	30 kHz 8 kbps speech
NDC	$\pi/4$ DQPSK	TDMA /FDMA	810–826/940–956 1477–1489/1429–1441 1501–1513/1453–1465	25 kHz
IS-94	QPSK	CDMA	869–894/824–849	1.25 MHz
	Direct Sequence Spread Spectrum — Qualcomm			

Table 1.2 Attributes of AMPS System. Adapted from [8].

PROPERTY	ATTRIBUTE
Number of Channels	832, 2 groups of 416 channels each group includes 21 signaling channels
Cell Radius	2-20 km
Base-to-Mobile Frequency	869-894 MHz
Mobile-to-Base Frequency	824-849 MHz 45 MHz between transmit and receive channels.
Channel Spacing	30 kHz
Modulation	FM with peak frequency deviation of ±12 kHz Can send data at 10 kbps
Base Station ERP	100 W per channel (maximum)
Coding	
Base-to-Mobile	40:28 BCH (i.e. 28 bits of information and 12 bits of parity, error checking)
Mobile-to-Base	48:36 BCH (i.e. 36 bits of information and 12 bits of parity, error checking) (Difference is because there is more processing power in the base.)
RF Specifications of Mobile Unit	
Transmit RF power	3 W maximum (33 dBm)
Transmit power control	10 steps of 4-dB attenuation each. Minimum power is -4 dBm
Receive Sensitivity	-116 dBm from a 50 W source applied at antenna terminals
Receive Noise Figure	6 dB measured at antenna terminals
Receive Spurious Response	-60 dB from center of the passband
Number of synthesizer channels	832

The AMPS (Advanced Mobile Phone Service) System is the analog system in North America. The attributes of this system are given in Table 2.2. This system is gradually being replaced by the North American Digital Cellular (NADC) system. The NADC system was designed to provide a transition from the current analog system to a fully digital system. It is expected that this will be achieved over a 10 year period and it will be done reusing existing spectrum. The idea is that system providers can allocate a few of their channels for digital radio out of the total available. As analog radio is phased out more of the channels can be committed to digital radio. The main motivation for this is that the NADC system has provides 3-5 times the capacity of the analog system. This contrasts with the main motivation of the European GSM system is to provide a common pan-European standard so that a user can roam throughout Europe. The GSM system provides eight times the capacity of current analog systems and has compatibility with ISDN. ISDN is something that NADC does not support. The Japanese standard, NDC, is similar to NADC but with the addition of a new band in the 1.4 to 1.5 GHz range. It is designed to be compatible with the NTT system which as a 25 kHz channel spacing as oppose to the American system which has a 30 kHz spacing.

The unique feature of $\pi/4$ DQPSK modulation is that the quadrature I and Q channels are coded independently. This can be explained by considering Fig. 2.10.

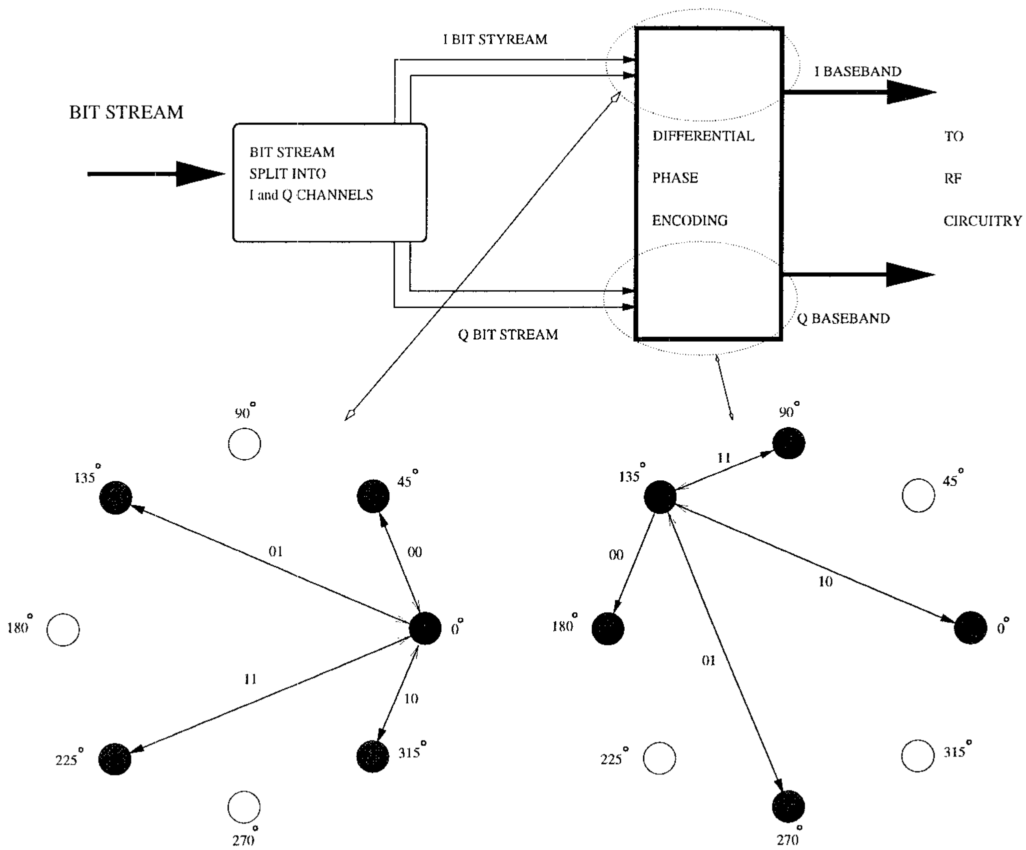

Figure 4.1.10 Differential coding of $\pi/4$ DQPSK.

A four bit stream is divided into two quadrature nibbles of two bits each. These nibbles independently control the I and Q encoding respectively. Hence the term dual quadrature modulation. The generalized quadrature modulation equation

$$x(t) = i(t)\cos[\omega_c t + \phi_i(t)]q(t)\sin[\omega_c t + \phi_q(t)] \tag{1}$$

is used, where $i(t)$ and $q(t)$ embody the particular modulation rule for amplitude, $\phi_i(t)$ and $\phi_q(t)$ embody the particular modulation rule for phase, and ω_c is the carrier radian frequency. In simpler modulation schemes the I and Q signals are 90^o out of phase and the result is that the carrier is suppressed. Only information bearing signals are transmitted. This is true even if $(\phi_q(t) - \phi_i(t))$ is 90^o only on average.

4.1.5 Access Protocols

In analog radio the separation of channels is by frequency separation or frequency division of the available spectrum. The frequency division multiple access protocol is illustrated in Figure 2.11. Here a voice channel is assigned a narrow slice of the spectrum or channel and is the sole occupant of that channel for the duration of the call. There are really only three ways to increase the spectrum efficiency of this system. One is to make the channels closer in frequency but this is

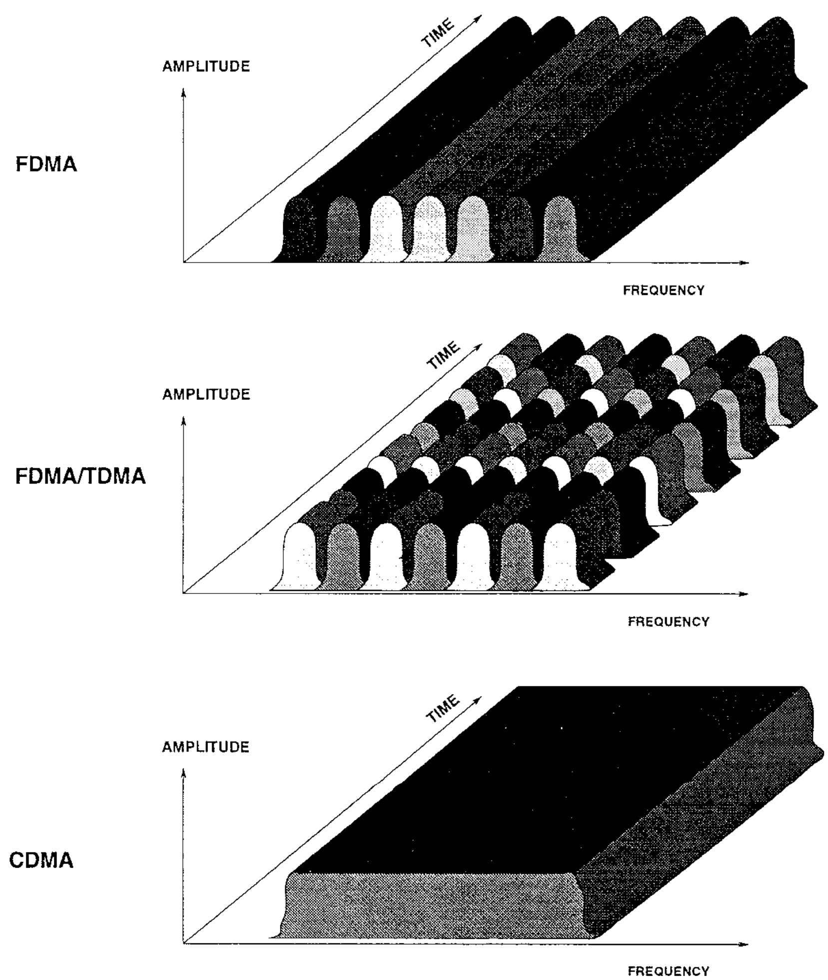

Figure 4.1.11 Comparison of channel assignment in FDMA — frequency division multiple access; and TDMA/FDMA — time division/frequency division multiple access; and CDMA — code division multiple access.

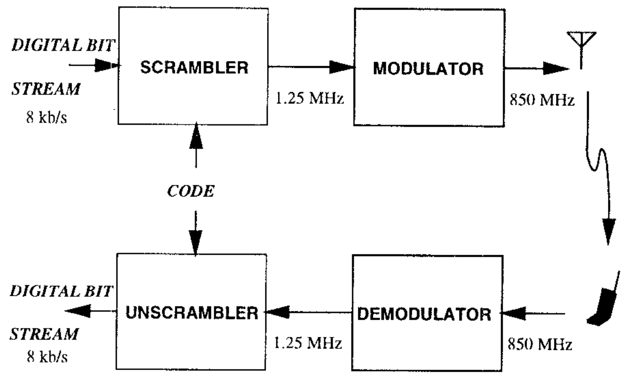

Figure 4.1.12 I-V of Power amplifier transistor showing quiescent point of various amplifier classes.

difficult to do, and another is to use digital techniques and in effect fill one physical channel with a bit stream. A third approach is to use spread spectrum (another digital technique) which will be discussed below. In the first digital technique the bit stream is divided among a few users using the same physical channel. In the current North American Digital Cellular system a physical channel is divided into six time slots and a user is allocated every third time slot so that a total of three users can be supported in each physical channel. Thus the logical channels are divided in both frequency and time however we refer to this access protocol as simply time division multiple access (TDMA). In a TDMA system the base station transmits a continuous stream of data containing frames of time slots for multiple users. The mobile listens to this continuous stream, extracting and processing only the time slots assigned to it. On the reverse transmission, the mobile transmits to the base station in bursts only in its assigned time slots. This is yet another complication to the RF design of TDMA systems, as well as dealing with, in general, a non constant RF envelope and hence the requirement for linear amplifiers, the RF circuitry operates in a burst mode with additional constraints on settling time and frequency spreading. Future extensions will use advanced coding techniques to increase the number of users per physical channel to six with each user using only one of the available time slots.

The second option to increasing bandwidth utilization is to use spread spectrum techniques. In this technique a baseband signal is mixed with a broadband coding signal to produce a broadband signal which is then used to modulate an RF signal. This process is illustrated in Figure 2.12. Only with the unique code used to "scramble" the original baseband signal can the original signal be recovered. The system can support many users with each user assigned a unique code hence the term code division multiple access (CDMA). There are two types of CDMA systems: frequency-hopping and direct sequence.

4.1.6 Power Descriptions

At RF and microwave frequencies the power of a signal is everything. Voltage and current of signals are not that meaningful as not much information is conveyed by these terms. Indeed it is very easy to change voltage and current levels using

Table 1.3 Common reference power designations.

P_{REF}	Bell Units	Decibel Units
1 W	BW	dBW
1 mW	Bm	dBm

transformers. Also there is a philosophical issue as to what voltage really means, especially in free-space. Instead we always want to talk in terms of the power of a signal. Also a noise source such a resistor or galactic radiation generates a finite amount of noise power and the signal-to-noise ratio (which is the square root of the signal power to noise power ratio) is one of our design parameters. While we can talk about power in absolute terms such as watts or milliwatts it much more useful to use a logarithmic scale. To use a logarithmic scale we must take a ratio. The ratio of two power levels in bells is

$$P(B) = \frac{P}{P_{\text{REF}}} \tag{2}$$

where P_{REF} is a reference power. Now human senses have a logarithmic response and the minimum resolution tends to be about 0.1 B. So it is most common for us to use dB — 1 B = 10 dB. Common designations are shown in Table 2.3. Also 1 dBm is a very common power level in RF and microwave power circuits so most powers are in terms of dBm. See, what is very important is obtaining good intuitive definitions. Then when we talk using these common descriptions it is very easy to exchange information with other designers as well as keep complicated information in our heads.

4.1.7 Antenna Gain and Effective Isotropic Radiated Power

We don't want to say much about antennas however there are some concepts that we must introduce now. One of these is the concepts of Antenna Gain and Effective Isotropic Radiated Power. Antennas do not radiate uniformly in all directions and, instead, concentrate power in particular, usually desired directions. So that we can quantify this, the concept of an isotropic radiator or antenna is introduced as a reference. The fictitious isotropic radiator radiates uniformly in all directions. So that if the total radiated power is P_R then the power radiated per unit solid angle, $dP_R/d\Omega$, depends on the characteristics of the antenna. For an isotropic antenna

$$\left(\frac{dP_R}{d\Omega}\right)_{\text{ISOTROPIC}} = \frac{P_R}{4\pi} \tag{3}$$

Defining the efficiency of an antenna as

$$\eta = P_R/P_{\text{IN}} \tag{4}$$

where P_{IN} is the power input to the antenna., (η is very close to one for most antennas) the gain of antenna is defined as

$$G = 4\pi \frac{\text{Power per unit solid angle}}{\text{Total input power to the antenna}}$$

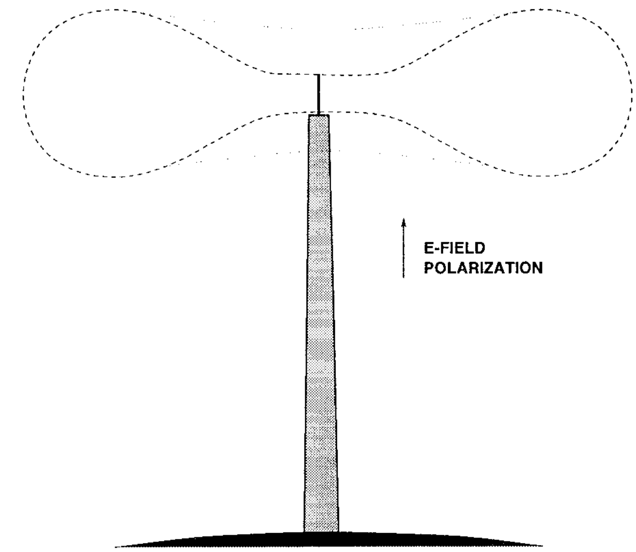

Figure 4.1.13 Base station transmitter pattern.

$$\begin{aligned} &= 4\pi \frac{dP_R/d\Omega}{P_{\text{IN}}} \\ &= \frac{\eta\, dP_R/d\Omega}{(dP_R/d\Omega)_{\text{ISOTROPIC}}} \end{aligned} \tag{5}$$

The gain of a realizable antenna is a function of direction. The peak value is incorporated in another quantity called the effective radiated isotropic power, EIRP:

$$\text{EIRP} = P_{\text{IN}} G_{\text{MAX}} \tag{6}$$

A base station, as shown in Figure 2.13, does not radiate power equally in all directions as it is inefficient to do so. Instead power is concentrated horizontally as shown in Figure 2.13 In the AMPS system the maximum EIRP per channel transmitted at the base-station is 100 W.

4.1.8 RF and Microwave Passive Circuits

RF and microwave circuits process signals and most circuits can be classified as performing one or more of the following functions: Amplify; shift the information bearing part of a signal in frequency (using a mixer); generate a reference signal of a precisely defined or controlled frequency using an oscillator; and reject signals of a certain frequency and pass others.

In this section we will address two of the hurdles to understanding RF and microwave circuits: Scattering Parameters and Smith Chart. We will also touch on another topic: how do you arrive at circuit elements for things that do not look like R, L and Cs. The example we will use is the transmission line.

Scattering (S) parameters are to an RF and microwave engineer as Y parameters are to lower frequency circuits people. The reason for using S parameters is that they provide a good intuitive understanding of how a circuit works. They also map very well on to our objectives in circuit design. Passive microwave elements

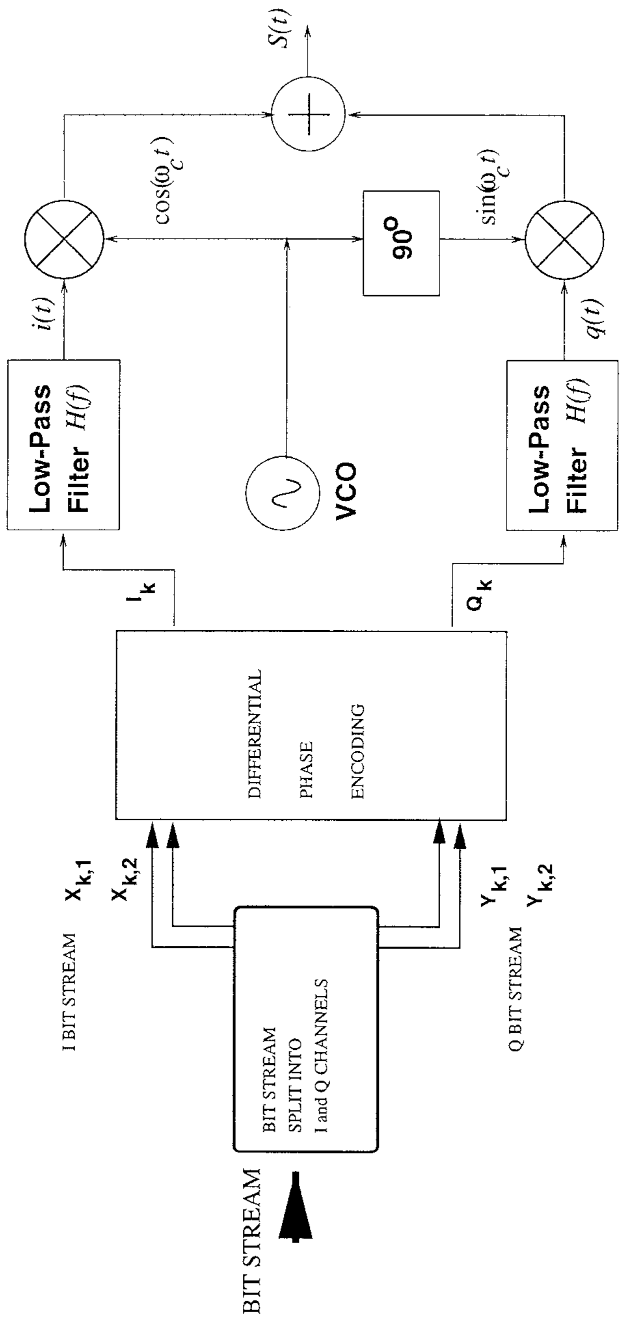

Figure 4.1.14 Modulation schematic for $pi/4$DQPSK modulation.

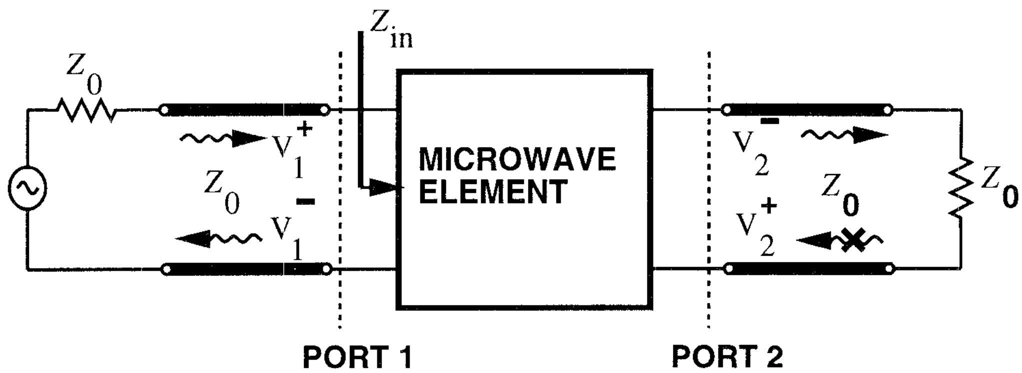

Figure 4.1.15 Incident, reflected and transmitted traveling voltage waves at a two-port microwave element.

are defined in terms of their reflection and transmission properties for an incident wave of electric field or voltage. In Fig. 2.15 a traveling voltage wave with phasor $\mathbf{V}_1^+$ is incident at port 1 of a two-port passive element. A voltage $\mathbf{V}_1^-$ is reflected and $\mathbf{V}_2^-$ is transmitted. In the absence of an incident voltage wave at port 2 (the voltage wave V_2^- is totally absorbed by Z_0) at port 1 the element has a voltage reflection coefficient

$$\Gamma_1 = \mathbf{V}_1^- / \mathbf{V}_1^+ \tag{7}$$

and transmission coefficient

$$\mathrm{T} = \mathbf{V}_2^- / \mathbf{V}_1^+ . \tag{8}$$

More convenient measures of reflection and transmission performance are the return loss and insertion loss as they are relative measures of power in transmitted and reflected signals. In decibels

$$\text{RETURN LOSS} = -20 \log \Gamma (dB) \qquad \text{INSERTION LOSS} = -20 \log \mathrm{T} (dB) \tag{9}$$

The input impedance at port 1, Z_{in}, is related to Γ by

$$Z_{\text{in}} = Z_0 \frac{1 + \Gamma_1}{1 - \Gamma_1} \tag{10}$$

These quantities will change if the loading conditions are changed. For this reason scattering parameters are used which are defined as the reflection and transmission coefficients with a specific load referred to as the reference impedance. Thus

$$S_{11} = \Gamma_1 \qquad S_{21} = \mathrm{T}. \tag{11}$$

S_{22} and S_{12} are similarly defined when a voltage wave is incident at port 2. For a multiport

$$S_{pq} = \mathbf{V}_p^- / \mathbf{V}_q^+ \tag{12}$$

with all of the ports terminated in the reference impedance. Simple formulas relate the S parameters to other network parameters [9][pp. 16-17]. S parameters are the most convenient network parameters to use with distributed circuits as a change in

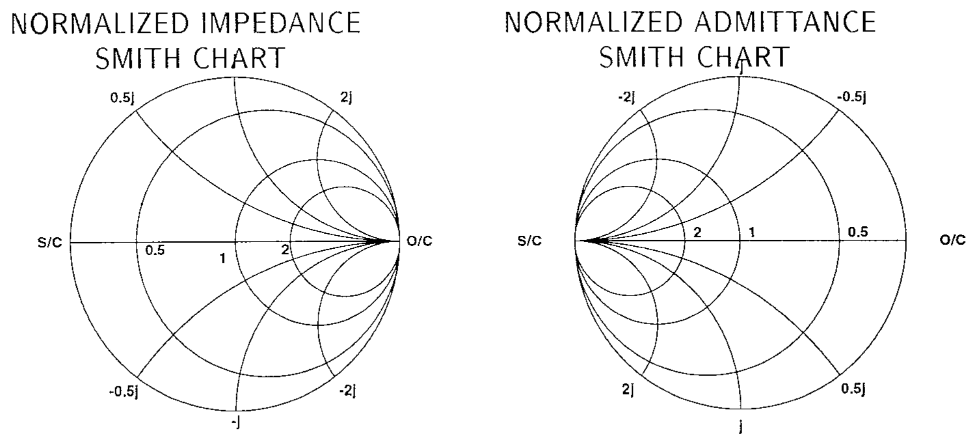

Figure 4.1.16 Smith charts.

line length results in a phase change. As well they are the only network parameters that can be measured directly at microwave and millimeter-wave frequencies.

Most passive devices, with the notable exception of ferrite devices, are reciprocal and so $S_{pq} = S_{qp}$. A loss-less passive device also satisfies the unitary condition: $\sum_p |S_{pq}|^2 = 1$ which is a statement of power conservation indicating that all power is either reflected or transmitted. A passive element is fully defined by its S parameters together with its reference impedance, here Z_0. In general the reference impedance at each port can be different. The Smith Chart is A polar plot of scattering parameters with loci of load impedances is called a Smith Chart, see Figure 2.16, and is the most important graphical tool for the design of microwave circuits. When reflection coefficient, S parameters, and Smith charts are being used the reference impedance must be known. If it is not stated generally we assume 50 Ω.

4.1.8.1 Passive Circuit Components

Wavelengths in air at microwave frequencies range from 1 m at 300 MHz to 1 cm at 30 GHz. These dimensions are comparable to those of fabricated electrical components and this relationship leads to new classes of distributed components that have no analogy at lower frequencies [1]. Components are realized by disturbing the field structure on a transmission line which results in energy storage and thus reactive effects. Microstrip discontinuities, Figure 2.17, are modeled by capacitive elements if the electric field is interrupted and by inductive elements if the magnetic field (or current) is disturbed. Thus a transmission line terminated in a load impedance Z_L has an input impedance

$$Z_{\text{in}} = Z_0 \frac{Z_L + jZ_0 \tanh \gamma d}{Z_0 + jZ_L \tanh \gamma d} \tag{13}$$

where d is the length of the line and γ is the propagation constant of the line. Thus a short section ($\gamma d << 1$) looks like an of short circuited ($Z_L = 0$) transmission line looks like an inductor and a capacitor if it is open circuited ($Z_L = \infty$). When the line is a half wavelength long, an open circuit is presented at the input to the

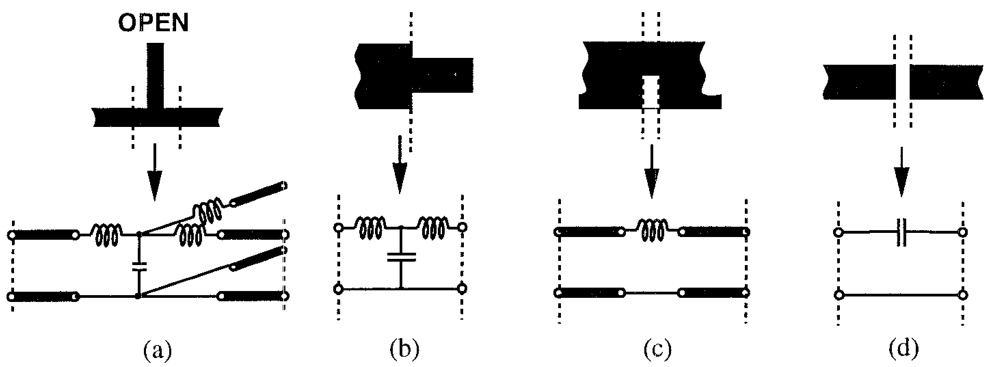

Figure 4.1.17 Microstrip discontinuities: (a) open microstrip stub, (b) step, (c) notch, and (d) gap.

line if the other end is short circuited. The stub shown in Fig. 2.17(a) presents a short circuit to the through transmission line when the length of the stub is a quarter wavelength long. When the stub is electrically short much shorter than a quarter wavelength it introduces a shunt capacitance in the through transmission line. In computer-aided design detailed equivalent circuits (perhaps just scattering parameters) are develop from field solutions are used rather than the simple electrical circuit models shown. However the models in Figure 2.17
are good approximations for manual design and circuit topology formulation.

Microwave resonators are used in bandpass filters and in establishing precise frequency references. In a lumped element resonant circuit stored energy is transfered between an inductor which stores magnetic energy and a capacitor which stores electric energy, and back again every period. Microwave resonators function the same way exchanging energy stored in electric and magnetic forms but with the energy stored spatially. Resonators are described in terms of their quality factor

$$Q = 2\pi f_0 \left(\frac{\text{Maximum Energy Stored in the Resonator at} f_0}{\text{Power Lost in the Cavity}} \right) \tag{14}$$

where f_0 is the resonant frequency.

Directional Couplers are multi-port components which preferentially route a signal incident at one port to the other ports. This property is called directivity. One type of hybrid is called a directional coupler the schematic of which is shown in Fig. 2.18(a). Here the signal incident at port 1 is coupled to ports 2 and 3 while very little is coupled to port 4. Similarly a signal incident at port 2 is coupled to ports 1 and 4 but very little power appears at port 3. The feature that distinguishes a directional coupler from other types of hybrids is that the power at the output ports are different. The performance of a directional coupler is specified by two parameters: Coupling Factor $= P_1/P_3$ and Directivity $= P_3/P_4$. A microstrip realization of a directional coupler is shown in Fig. 2.18(b) where power is coupled in the backward direction.

Filters and matching networks are combinations of microwave passive elements designed to have a specified impedance and frequency response. Typically a topology of a filter is chosen based on established lumped element filter design

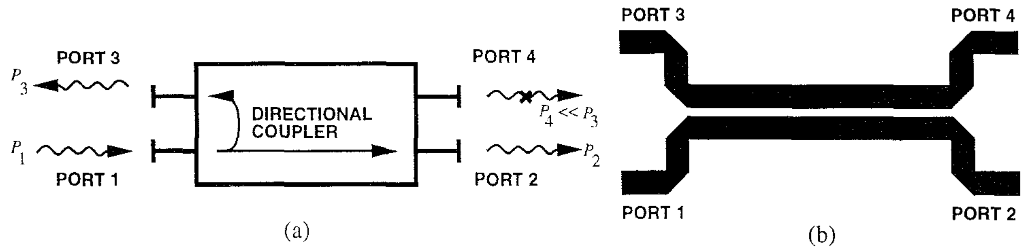

Figure 4.1.18 Directional couplers: (a) schematic, and (b) backward coupling microstrip directional coupler.

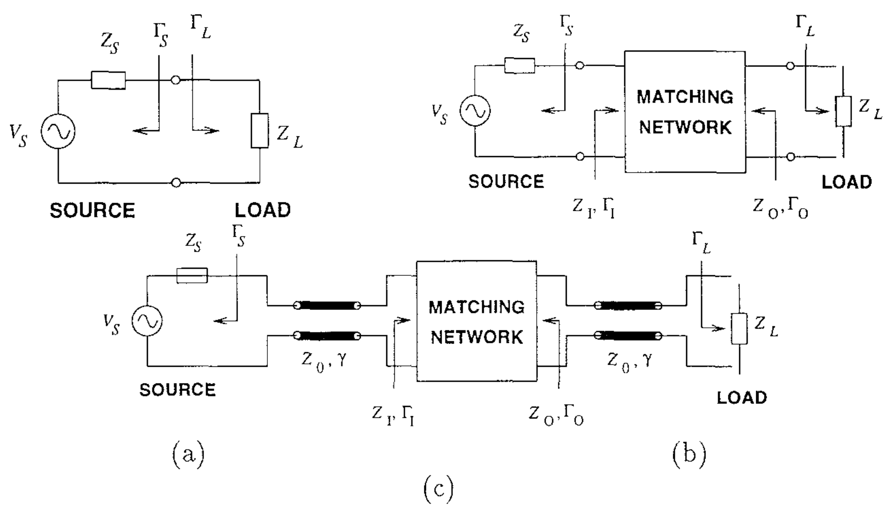

Figure 4.1.19 Matching networks: (a) Load and source network; (b) with matching network; (c) and with a distributed matching network.

theory. The basic design aim is, at in-band frequencies, to provide maximum power transfer to the input of an amplifier or to a load. Then computer aided design techniques are used to optimize the response of the circuit to the desired response. The basic forms of matching networks are shown in Figure 2.19.

4.1.9 Interference and Nonlinear Phenomena

The minimum signal detectable in conventional wireless systems is determined by the signal-to-interference ratio at the input. The noise is due to background noise sources including galactic noise and thermal noise. Placement of "cells" using common frequencies is such that interference between transmitters operating at similar frequencies is large enough for a signal to decay below the noise. In cellular wireless systems the minimum signal detectable is also determined by the signal-to-interference ratio but now the dominant interference is due to other transmitters in the cell and adjacent cells. The noise that is produced in the signal band from other transmitters operating at the same frequency is called co-channel interference. The level of co-channel interference is dependent on cell placement and frequency reuse

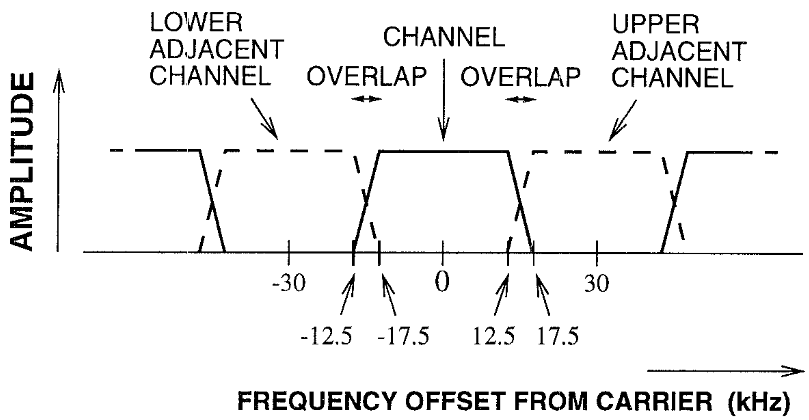

Figure 4.1.20 Adjacent channels and overlap in the AMPS and NADC cellular systems.

pattern. The degree to which co-channel interference can be controlled has a large effect on system capacity. A minimum signal-to-interference (S/I) ratio of 17 dB is generally considered to be acceptable in the North American analog cellular phone system. Digital cellular systems can tolerate higher levels of interference through the use of redundant coding. Control of co-channel interference is achieved by controlling the power levels at the base-station and at the mobile units. Factors affecting the interference are

- The signal power falls off as D^{-4} where D is distance.
- Transmitted power can be reduced to achieve a minimum acceptable signal-to-interference ratio and by so doing reducing interference in neighboring cells.

Adjacent channels also generate interference. This is partly due to the design of the frequency allocation system with overlapping frequency channels, see Figure 2.20, and also to nonlinear effects. Adjacent channel interference is primarily the result of the nonlinear behavior of transmitters and so characterization of this phenomenon is important in RF design. Adjacent channel interference occurs with both digitally modulated and analog modulated RF signals. It turns out that conventional design approaches can be used to control and predict adjacent channel interference for analog modulated signals but there is as yet not a good design practice for digitally modulated signals.

Using the frequency limits defined in Figure 2.21, the lower channel ACPR is defined as

$$\mathrm{ACPR}_{\mathrm{ADJ,LOWER}} = \frac{\text{Power in Lower Adjacent Channel}}{\text{Power in Main Channel}} \tag{15}$$

The nonlinear response of active RF and microwave circuits is one of the major issues in RF design. There are several important forms of nonlinear behavior which must be kept in mind during design. These are gain compression; amplitude modulation-to-phase modulation (AM-PM) conversion; and intermodulation distortion. In a radio system the adjacent channel interference is a major design

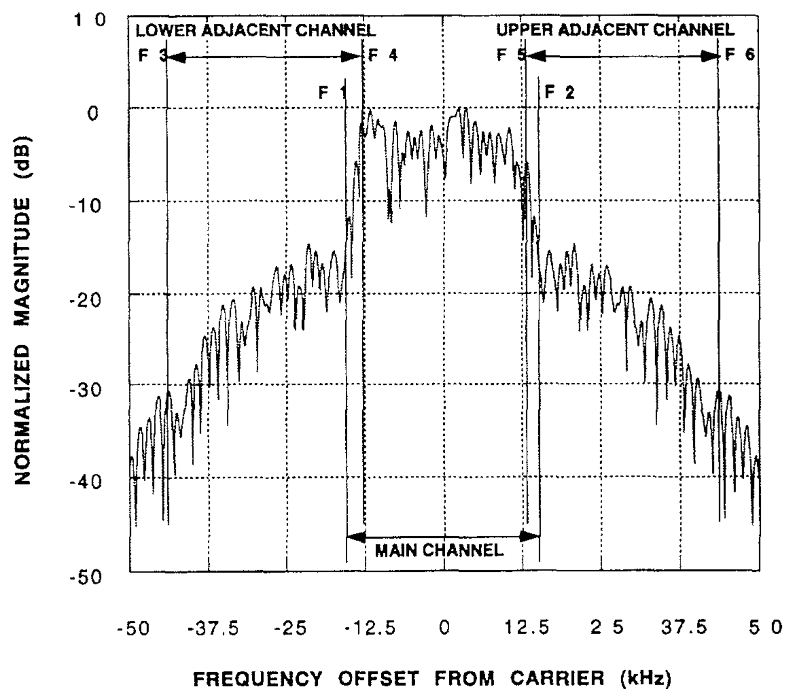

Figure 4.1.21 Definition of adjacent-channel and main channel integration limits using a typical NADC spectrum as an example.

constraint and is related to intermodulation distortion and, for a digitally modulated signal, to AM to PM conversion. AM to PM conversion is such a major problem for digital modulation as modulation includes both phase and amplitude variations. Thus if amplitude modulation appears as phase modulation signal information is lost. Because of these differences in the response to the modulation format, designs vary according to the type of modulation. The principal purpose of this section is to gain an understanding of nonlinear distortion behavior.

4.1.9.1 Nonlinear Phenomena in Analog Circuits

Fourier analysis and the use of phasors is at the heart of linear analog circuit engineering. Nonlinear phenomena in sinusoidally excited nonlinear analog circuits can be examined and classified in the frequency domain. This can provide insight into the nonlinear processes and enable circuits to be designed that suppress, or perhaps enhance, nonlinear characteristics.

Here the term *intermodulation* to describe the process by which power at one frequency, or group of frequencies, is transferred to power at other frequencies. The term intermodulation is also used to describe the production of sum and difference frequency components, or intermodulation frequencies, in the output of a system with multiple input sinewaves. This is a macroscopic definition of intermodulation as the generation of each intermodulation frequency derives from many separate intermodulation processes. In this section we develop a treatment of intermodulation at the microscopic level and use this treatment is used to describe frequency domain manifestations of nonlinearity.

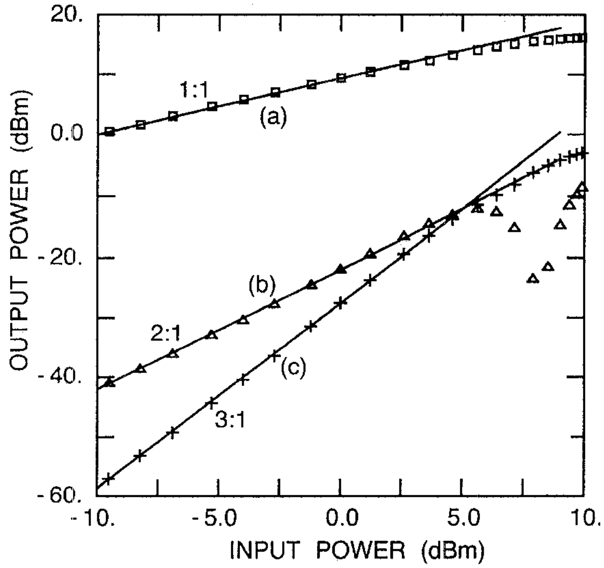

Figure 4.1.22 Measured characteristics of a microwave MESFET amplifier to a one tone input at 2.3 GHz: (a) fundamental, (b) second harmonic, and (c) third harmonic. Measured (markers) and simulated (lines) responses.

Power series expansion analysis of a nonlinear subsystem is straight forward and is a convenient way to introduce and classify nonlinear phenomena in sinusoidally excited nonlinear analog circuits. Nonlinear phenomena in the time-domain manifests itself as saturation or as a nonlinear relation between an input quantity and an output quantity. When a single frequency sinusoidal signal excites a nonlinear circuit the response "usually" includes the original signal and harmonics of the input sinewave. We say usually because if the circuit contains nonlinear reactive elements subharmonics and autonomous oscillation could also be present. The output may not even be approximately periodic if there is chaos. The process is even more complicated when the excitation includes more than one sinusoid as then the circuit response may include all sum and difference frequencies of the original signals.

In the following the term 'intermodulation' is used to describe the process by which power at one frequency, or group of frequencies, is transferred to power at other frequencies. The term intermodulation is also used to describe the production of sum and difference frequency components, or intermodulation frequencies, in the output of a system with multiple input sinewaves. This is a macroscopic definition of intermodulation as the generation of each intermodulation frequency component derives from many separate intermodulation processes. Here we develop a treatment of intermodulation at the microscopic level.

Consider the unilateral nonlinear system of Figure 2.23 described by the power series

$$y(t) = \sum_{l=1}^{\infty} a_l x^l(t) \qquad (16)$$

where $x(t)$ is the input and is the sum of three sinusoids

Figure 4.1.23 A unilateral nonlinear system.

$$x(t) = c_1 \; cos(\omega_1 t + \phi_1) + c_2 \; cos(\omega_2 t + \phi_2) + c_3 \; cos(\omega_3 t + \phi_3) \tag{17}$$

To simplify things let $\alpha_1 = \omega_1 t + \phi_1$, $\alpha_2 = \omega_2 t + \phi_2$ and $\alpha_3 = \omega_3 t + \phi_3$ so that

$$x(t)^l = [c_1 \; cos(\alpha_1) + c_2 \; cos(\alpha_2) + c_3 cos(\alpha_3)]^l \tag{18}$$

$$= \sum_{p=0}^{l} \sum_{k=0}^{p} \binom{p}{k} \binom{l}{p} c_1^k c_2^{p-k} c_3^{l-p} (cos\alpha_1)^k (cos\alpha_2)^{p-k} (cos\alpha_1)^{l-p} \tag{19}$$

This equation includes a large number of components the radian frequencies of which are the sum and differences of ω_1, ω_2 and ω_3. These result from multiplying out the term $(cos\alpha_1)^k (cos\alpha_2)^{p-k} (cos\alpha_1)^{l-p}$. For example

$$(cos\alpha_1)(cos\alpha_2)(cos\alpha_3) = \frac{1}{4}[\quad cos(\alpha_1 + \alpha_2 + \alpha_3) \quad + \tag{20}$$

$$cos(\alpha_1 + \alpha_2 - \alpha_3) \quad + \tag{21}$$

$$cos(\alpha_1 - \alpha_2 + \alpha_3) \quad + \tag{22}$$

$$cos(\alpha_1 - \alpha_2 - \alpha_3) \quad] \tag{23}$$

where the (radian) frequency of the component in (20) is $(\omega_1 + \omega_2 + \omega_3)$, in (21) is $(\omega_1 + \omega_2 - \omega_3)$, in (22) is $(\omega_1 - \omega_2 + \omega_3)$, and in (23) is $(\omega_1 - \omega_2 - \omega_3)$. Thus even though v has only three frequency components, the output z can have many sum and difference frequencies of ω_1, ω_2 and ω_3. In general the frequencies of the components of y are

$$m_1\alpha_1 + m_2\alpha_2 + m_3\alpha_3 \tag{24}$$

with

$$m_1 = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, \infty \tag{25}$$

$$m_2 = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, \infty \tag{26}$$

$$m_3 = -\infty, \ldots, -2, -1, 0, 1, 2, \ldots, \infty \tag{27}$$

If the power index k did not go to infinity then the frequency spectrum of z would be finite but in any case the spectrum invariably needs to be truncated to reduce the number of components to be incorporated in any analysis.

4.1.9.2 Frequency-Domain Expansions

The expansion in (19) is in the time domain but frequency-domain forms can be developed. Consider a two tone input

$$x(t) = | X_1 | \cos(\omega_1 t + \phi_1) + | X_2 | \cos(\omega_2 t + \phi_2)$$

which can be written using phasor notation as

$$x(t) = \frac{1}{2}\left[X_1 e^{j\omega_1 t} + X_1^* e^{-j\omega_1 t} + X_2 e^{j\omega_2 t} + X_2^* e^{-j\omega_2 t}\right].$$

The first three powers of x can be easily expanded manually, *e.g.* expanding x^2 gives

$$\begin{aligned} x^2(t) &= \left(\frac{1}{2}\right)^2 \Big[X_1^2 e^{j2\omega_1 t} + 2X_1X_1^* + 2X_1X_2 e^{j(\omega_1+\omega_2)t} + 2X_1X_2^* e^{j(\omega_1-\omega_2)t} \\ &+ (X_1^*)^2 e^{-j2\omega_1 t} + 2X_1^*X_2 e^{j(\omega_2-\omega_1)t} + 2X_1^*X_2^* e^{-j(\omega_1+\omega_2)t} + X_2^2 e^{j2\omega_2 t} \\ &+ 2X_2X_2^* + (X_2^*)^2 e^{-j2\omega_2 t}\Big], \end{aligned} \tag{28}$$

and similarly expanding x^3 yields

$$\begin{aligned} x^3(t) &= \left(\frac{1}{2}\right)^3 \Big[X_1^3 e^{j3\omega_1 t} + 3X_1^2X_1^* e^{j\omega_1 t} + 3X_1^2X_2 e^{j(2\omega_1+\omega_2)t} \\ &+ 3X_1^2X_2^* e^{j(2\omega_1-\omega_2)t} + 3X_1(X_1^*)^2 e^{-j\omega_1 t} + 6X_1X_1^*X_2 e^{j\omega_2 t} \\ &+ 6X_1X_1^*X_2^* e^{-j\omega_2 t} + 3X_1X_2^2 e^{j(\omega_1+2\omega_2)t} \\ &+ 6X_1X_2X_2^* e^{j\omega_1 t} + 3X_1(X_2^*)^2 e^{j(\omega_1-2\omega_2)t} + (X_1^*)^3 e^{-j3\omega_1 t} \\ &+ 3(X_1^*)^2X_2 e^{j(\omega_2-2\omega_1)t} + 3(X_1^*)^2X_2^* e^{-j(2\omega_1+\omega_2)t} + 3X_1^*X_2^2 e^{j(2\omega_2-\omega_1)t} \\ &+ 3X_1^*(X_2^*)^2 e^{-j(\omega_1+2\omega_2)t} + X_2^3 e^{j3\omega_2 t} + 6X_1^*X_2^*X_2 e^{-j\omega_1 t} + 3X_2^2X_2^* e^{j\omega_2 t} \\ &+ 3X_2(X_2^*)^2 e^{-j\omega_2 t} + (X_2^*)^3 e^{-j3\omega_2 t}\Big], \end{aligned} \tag{29}$$

so that the output of the cubic equation

$$y(t) = a_0 + a_1 x(t) + a_2 x^2(t) + a_3 x^3(t)$$

can be calculated for the two tone input. Table 2.4 lists the various intermodulation products resulting from x, x^2, and x^3 and groups them according to frequency. (Only the terms for the positive frequencies are listed.) To calculate the output at a specific frequency, the appropriate terms are summed. For example, the phasor output at ω_1 is given by the sum of three intermodulation products

$$Y_{\omega_1} = a_1\left(\frac{1}{2}\right)X_1 + 3a_3\left(\frac{1}{2}\right)^3 X_1^2X_1^* + 6a_3\left(\frac{1}{2}\right)^3 X_1X_2X_2^*. \tag{30}$$

The algebraic formulas presented in the preceding section generalize this procedure for N input frequencies and for the more complicated generalized power series descriptions of the nonlinear elements.

Multi-tone Expansions

When a sum of sinusoids is input to a nonlinear element additional frequency components are generated. In our previous example there were two distinct input tones and there are components of the output due to the x^2 term at $2\omega_1$, $2\omega_2$, $(\omega_1+\omega_2)$, $(\omega_2-\omega_1)$, and DC in addition to the original frequencies ω_1 and ω_2. If there are many tones a similar analysis as was done could be repeated. This can be tedious but fortunately there are formulas that enable us to calculate the response at individual frequencies not matter how many tones are in the input

Table 1.4 The intermodulation products resulting from x, x^2, and x^3 where x is a two tone signal, showing only the positive frequencies.

Intermodulation Product	Frequency	Order
$\frac{1}{2}X_1X_1^*$	0	2
$\frac{1}{2}X_2X_2^*$	0	2
$\frac{1}{2}X_1$	ω_1	1
$(\frac{1}{2})^3 3X_1^2X_1^*$	ω_1	3
$(\frac{1}{2})^3 6X_1X_2X_2^*$	ω_1	3
$\frac{1}{2}X_2$	ω_2	1
$(\frac{1}{2})^3 3X_2^2X_2^*$	ω_2	3
$(\frac{1}{2})^3 6X_1X_1^*X_2$	ω_2	3
$(\frac{1}{2})^2X_1^2$	$2\omega_1$	2
$(\frac{1}{2})^2X_2^2$	$2\omega_2$	2
$(\frac{1}{2})^3X_1^3$	$3\omega_1$	3
$(\frac{1}{2})^3X_2^3$	$3\omega_2$	3
$\frac{1}{2}X_1X_2$	$\omega_1+\omega_2$	2
$\frac{1}{2}X_1X_2^*$	$\omega_1-\omega_2$	2
$(\frac{1}{2})^3 3X_1^2X_2$	$2\omega_1+\omega_2$	3
$(\frac{1}{2})^3 3X_1^2X_2^*$	$2\omega_1-\omega_2$	3
$(\frac{1}{2})^3 3X_1X_2^2$	$\omega_1+2\omega_2$	3
$(\frac{1}{2})^3 3X_1^*X_2^2$	$2\omega_2-\omega_1$	3

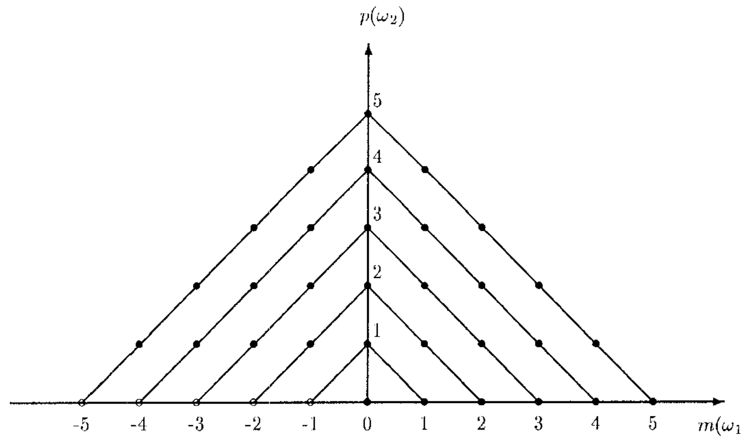

Figure 4.1.24 Possible combinations of two input frequencies including fifth order products.

signal [10]. When components resulting from the nonlinear process are input to the nonlinear element even more frequencies are generated. A similar situation occurs for the other terms in the power series. In order to make the analysis tractable, the number of frequency components considered must be limited. For the two tone input just discussed, the frequencies generated are integer combinations of the two inputs, *e.g.* $\omega = m\omega_1 + p\omega_2$ where m and p are integers. One way of limiting the number of frequencies is to consider only the combinations of m and p such that

$$| m | + | p | \leq n_{max}$$

assuming that all products of order greater than n_{max} are negligible. This approach is illustrated in Figure 2.24. For the more general case in which N components are considered as inputs, a set of integers, denoted n_k are used to specify the frequency,

$$\omega = \sum_{k=1}^{N} n_k \omega_k$$

where the order of intermodulation is given by

$$n = \sum_{k=1}^{N} | n_k | .$$

This set of integers (n_k) is called the intermodulation product description (IPD).

To further illustrate the concept of intermodulation products and IPD's, consider the simple spectrum for a mixer circuit shown in Figure 2.25. This spectrum retains only those components integral to the operation of the mixer: f_2 is the LO; f_3 is the RF; f_4 is DC; and $f_1 = f_2 - f_3$ is the IF, all of which are considered as

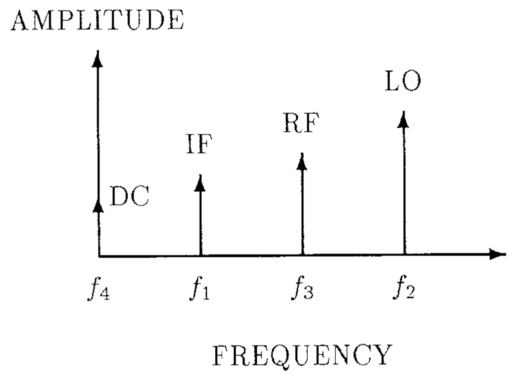

Figure 4.1.25 Simplified spectrum for a mixer circuit.

Table 1.5 Partial listing of IPD's when DC is not an input to the algebraic formula.

Output Frequency	n	n_1	n_2	n_3
f_1, IF	1	1	0	0
	2	0	1	-1
	4	2	-1	1
	5	-1	2	-2
	7	3	-2	2
	8	-2	3	-3
f_2, LO	1	0	1	0
	2	1	0	1
	4	-1	2	-1
	5	2	-1	2
	7	-2	3	-2
	8	3	-2	3
f_3, RF	1	0	0	1
	2	-1	1	0
	4	1	-1	2
	5	-2	2	-1
	7	2	-2	3
	8	-3	3	-2
f_4, DC	0	0	0	0
	3	1	-1	1
	6	2	-2	2

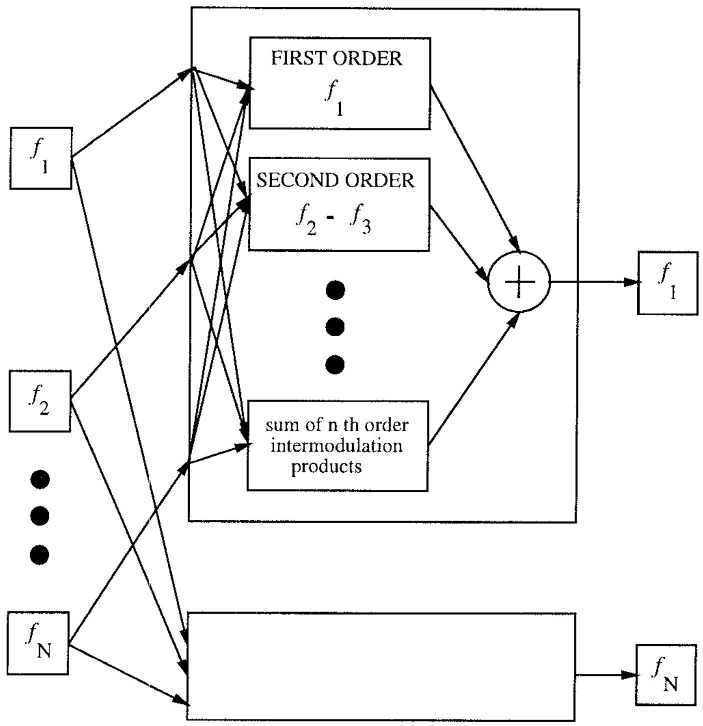

Figure 4.1.26 Illustration of nonlinear process in the frequency domain.

inputs when evaluating the algebraic formula. A partial listing of the intermodulation products is given in Table 2.5. For example, Table 2.5 lists the fourth order intermodulation product description $f_1 = 2f_1 - f_2 + f_3$ which yields a component at f_1 and corresponds to an intermodulation product of the form $X_1^2 X_2^* X_3$. The evaluation of the algebraic formula for all IPD's is illustrated in Figure 2.26. This shows that each intermodulation product is calculated independently and summed to give the output at a particular frequency.

4.1.9.3 Nonlinear System Response to Single-Frequency Excitation

With one-tone excitation, the spectra of the input and output of the nonlinear system are shown in Figure 2.27. Here intermodulation converts power at f_1 to power at DC (this intermodulation is commonly referred to as rectification), and to power at the harmonics ($2f_1$, $3f_1$, ...), as well as to power at f_1. At least this is the conventional view of the nonlinear response to a single frequency sinusoidal excitation. However, if the nonlinearity is reactive subharmonics can also be generated.

4.1.9.4 Nonlinear System Response to Multifrequency Excitation

The nonlinear response to single-frequency sinewave excitation of a nonlinear circuit is simple to describe. However it is much more complicated to describe the nonlinear response to multi-frequency sinusoidal excitation. If the excitation of an analog circuit is sinusoidal then specifications of circuit performance are generally in terms of frequency-domain phenomena, e.g. intermodulation levels, gain, and the

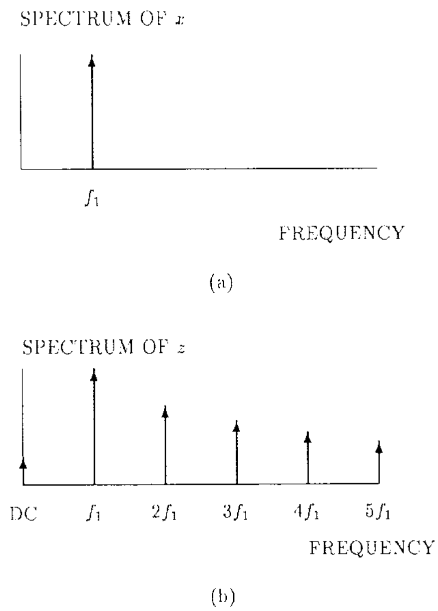

Figure 4.1.27 Spectrum of x with one tone excitation (a), and spectrum of z for a general nonlinearity (b).

1 dB gain compression point. In the time-domain the nonlinear behavior is evident as saturation or clipping of a waveform so that a sinusoidal waveform is distorted to a perturbed periodic waveform. However with multi-frequency excitation by signals that are not harmonicly related, the waveforms in the circuit are not periodic. Here we first look at the nonlinear response of a circuit to multi-frequency excitation and then classify the nonlinear phenomena.

Consider the nonlinear response of the system of Figure 2.23 to the two-tone excitation shown in Figure 2.28(a). The frequencies f_1 and f_2 are, in general, non-harmonicly related and components at all sum and difference frequencies (mf_1+nf_2, $m, n = -\infty, \ldots, -1, 0, 1, \ldots, \infty$) of f_1 and f_2 will appear at the output of the system. If the nonlinear system has a quadratic nonlinearity (i.e. the maximum value of l in (16) is 2) the spectrum of the output of the system is that of Figure 2.28(b). With a general nonlinearity, so that k can be large, the spectrum of the output will contain a very large number of components. An approximate output spectrum is given in Figure 2.28(c). Also shown is a truncated spectrum which will be used in the following discussion. Most of the frequency components in the truncated spectrum of Figure 2.28(c) have been named: DC results from rectification; f_3, f_4, f_5, f_8, f_9, f_{10} and intermodulation frequency components; f_4, f_5 are commonly called image frequencies as well; f_1, f_2 are the input frequencies; and f_6, f_7 are harmonics.

All of the frequencies in the steady-state output of the nonlinear system result from intermodulation — the process of frequency mixing. In other words, each frequency component is the summation of intermodulation products. However it is

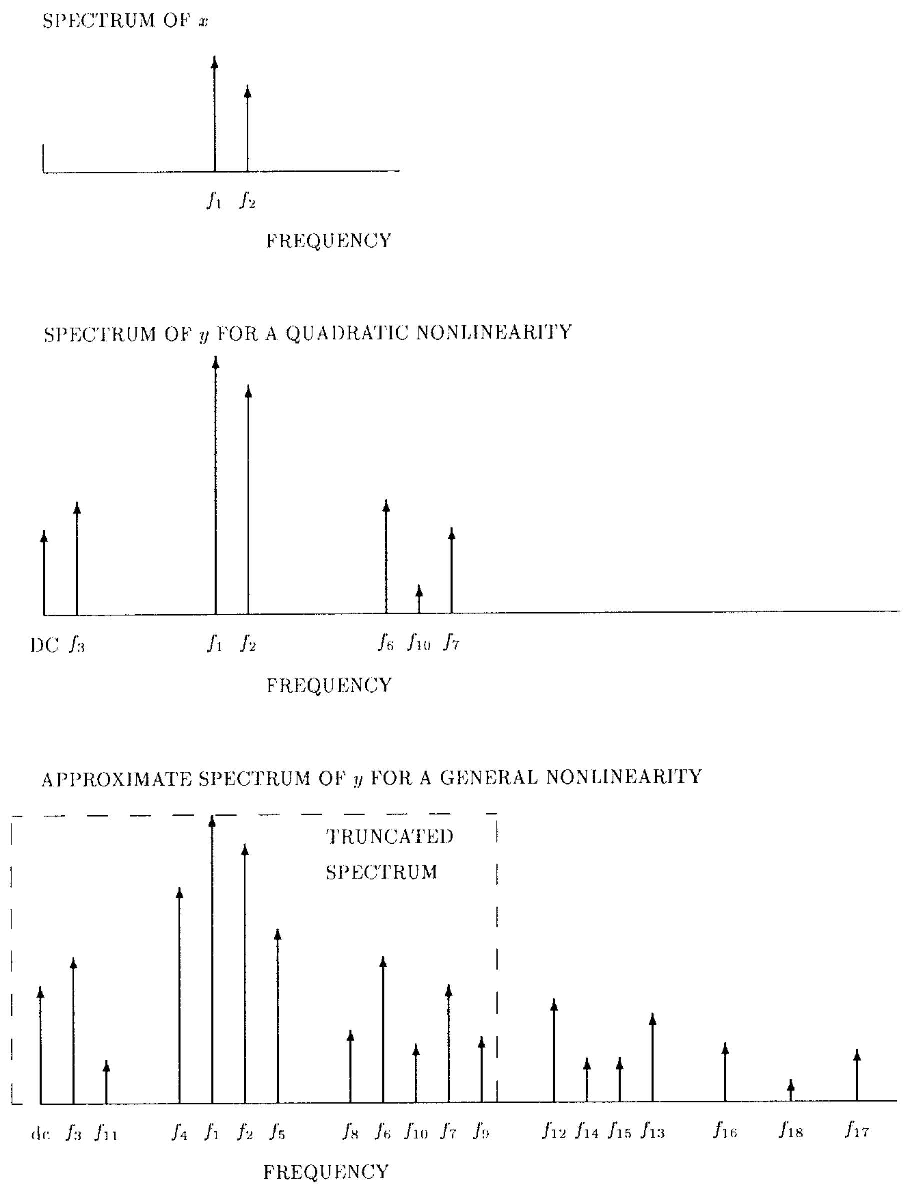

Figure 4.1.28 Spectrum of x with two tone excitation (a), complete spectrum of y for a quadratic nonlinearity (b), and approximate spectrum of y for a general nonlinearity (c), for the unilateral system of Figure 2.23.

usual to refer those frequencies additional to the input and output frequencies, and their harmonics, as intermodulation frequencies. It is unfortunate that the term intermodulation is used in two related but slightly different contexts. The terms 'intermodulation frequency' or 'intermodulation component' refer to the undesired frequencies generated in the mixing process as described here. The term 'intermodulation product' (IP) refers to the entire nonlinear process of sinusoids (or even a single frequency component) mixing to produce components at any frequency.

Gain Compression/Enhancement

Gain compression can be conveniently described in the time-domain or in the frequency-domain. Time-domain descriptions refer to limited power availability or to limitations on voltage or current swings. At low signal levels moderately nonlinear devices such as class A amplifiers behave linearly so that there is one dominant IP. As signal levels increase other IPs become important as harmonic levels increase. Depending on the harmonic loading condition, these IPs could be in phase with the original IP contributing to gain enhancement or out of phase contributing to gain compression. As well saturation can be larger. Which effect — saturation or additional IPs — dominates depends on the particular design although the two effects can be used to balance each other and so extend the power at which departures from linearity become significant. In the truncated spectrum only the fundamental and the second harmonic are present so few low order IPs contribute to gain saturation/enhancement. In a class A MESFET amplifier gain compression results in the classic saturation characteristic shown as curve (a) in Figure 2.22.

Desensitization

Desensitization is the variation of the amplitude of one of the desired components due to the presence of another noncommensurable signal. This phenomenon is depicted in Figure 2.29 where the amplitude of the signal at f_1 is affected by the amplitude of the signal at f_2. Desensitization is the result of saturation and the generation of IPs is not involved. When signals are small saturation is negligible but it becomes larger as the signal levels increase. Thus if there are two noncommensurable signals, one small and the other large, at the input of an amplifier, the gain of the small signal can be affected by just the presence of the large signal.

Harmonic Generation

Harmonic generation is the most obvious result of nonlinear distortion. Simply squaring a sinusoidal signal will give rise to a second harmonic component. In the truncated spectrum only the second harmonic is present so that few IPs are involved. The first order IP corresponds to the generation of, for example, current at the second harmonic due to voltage at the second harmonic. This IP, of course, does not give rise to the second harmonic. Second and higher order mixing leads to a large number of IPs. At small signals only the second order IPs are significant leading to the classic slope of 2 characteristic of the second harmonic power as a function of the input fundamental power, see Figure 2.22. As the input power increases higher order IPs become significant. If these are out-of-phase with the dominant second order IP then destructive cancelation can occur at higher input powers as can be seen in Figure 2.22 at an input of approximately 8 dBm.

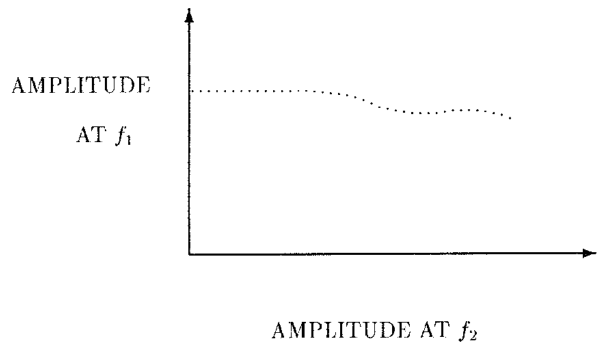

Figure 4.1.29 Depiction of desensitization.

Intermodulation

Intermodulation is the generation of spurious frequency components at the sum and difference frequencies of the input frequencies. In the truncated spectrum f_3, f_4, f_5, f_8, f_9, f_{10} and f_{11} are intermodulation frequencies. There are a large number of low order IPs contributing to this effect. For the reasons given above the first order response is usually large, but does not describe the origins of this phenomenon. The lowest order mixing that gives rise to intermodulation is a second order process and we see that for the truncated spectrum there are four second order IPs. These can destructively and constructively interfere so that the power of an intermodulation frequency component can vary erraticly at high input powers.

Cross-modulation

Cross-modulation is modulation of one component by another noncommensurable component. Here it would be modulation of f_1 by f_2 or modulation of f_2 by f_1. For discrete tone excitation cross-modulation can not be easily distinguished from desensitization. However with cross-modulation, information contained in the sidebands of one noncommensurable can be transferred to the other noncommensurable tone. This transfer of information does not occur with desensitization.

Detuning

Detuning is the generation of DC charge or DC current resulting in change of an active device's operating point. The generation of DC current with a large signal is commonly referred to as rectification. In a rectifier a sinusoidal signal is rectified to produce an output that contains a large DC component. The effect of rectification. can often be reduced by biasing using voltage and current sources. However DC charge generation in nonlinear reactances is more troublesome as it can neither be detected nor effectively reduced.

AM-PM Conversion

The conversion of amplitude modulation to phase modulation (AM-PM conversion) is a troublesome nonlinear phenomenon in high frequency analog circuits and results from the amplitude of a signal affecting the delay through a system. The most convenient way of modeling this effect at the subsystem level is to introduce order dependent time delays in the power series description of the transfer function. With these, the formula for an intermodulation component of the output contains a phase shifting factor. As a result the various IPs at a particular frequency can be out of phase. Since the magnitude of each IP is signal level dependent the frequency component, which is the sum of the IPs, will also have a level-dependent phase shift. No specific IPs can be assigned to this affect but frequency-domain power series expansion analysis leads to expressions describing AM-PM conversion.

Sub-Harmonic Generation and Chaos

In systems with memory effects, i.e. with reactive elements, sub-harmonic generation is possible. The intermodulation products for sub-harmonics can not be expressed in terms of the input non-commensurable components. Sub-harmonics are initiated by noise, possibly a turn-on transient, and so in a simulation must be explicitly incorporated into the assumed set of steady-state frequency components. The lowest common denominators of the sub-harmonic frequencies then becomes the basis noncommensurable component.

Chaotic behavior can only be simulated in the time-domain. The nonlinear frequency-domain methods as well as the conventional harmonic balance methods simplify a nonlinear problem by imposing an assumed steady-state on the nonlinear circuit solution problem. Chaotic behavior is not periodic and so the simplification is not valid in this case. Together with the ability to simulate transient behavior, the capability to simulate chaotic behavior is the unrivaled realm of time-domain methods.

4.1.9.5 Other Sources of Nonlinear Interference

Many components produce very low levels of nonlinearity. This is principally due to the low levels semiconductor action when dissimilar types of materials are used to form an electrical connection. A good example of this is the intermodulation that results at cable connections [5]. The most troublesome interference in intermodulation distortion – the generation of frequency tones (the intermod frequencies) from the sum and difference of tones at other frequencies. The most common figure-of-merit is third order intermodulation level — IMP 3. This is generally quoted as the ratio of the third order intermodulation frequency to the carrier level. The nonlinear performance of a cable connection is presented in Figure 2.30. Here the third order intermod was measured for the GSM system with two equal amplitude tones at 936 MHz and 958 MHz at 20 W (43 dBm) output power. The third order intermods occur at 914 MHz and 980 MHz. The 914 MHz tone is within the mobile-to-base receive band and so could cause a false received signal at the base station receiver. The data shows the effect of that varying the number of contact points can have. The push-in cable has the fewest contact points and so has the highest level of third order intermodulation. The best is the soldered connection which effectively distributes the contact points over a large area. Note also that flexing the cable, intentionally or because the cable is improperly pointed so that it can be moved by wind, can significantly degrade connector performance. In practice this distortion does not have as significant impact as it may seem as channel assignment minimizes the effect of intermods. So here, if the transmit channels included the 936 MHz and 958 MHz were transmit channels then the cell site would not use the 914 MHz channel as a base-station receive channel.

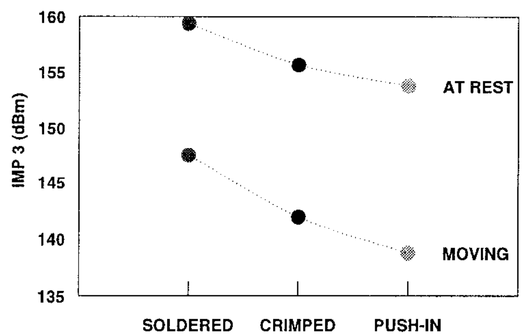

Figure 4.1.30 Third order intermodulation performance of moving and stationary cables for various connection technologies.

REFERENCES

[1] M. B. Steer, "Passive Microwave Devices," in *The Electrical Engineering Handbook*, CRC Press, 1993, pp. 882-891.

[2] M. B. Steer, C. R. Chang and G. W. Rhyne, "Computer Aided Analysis of Nonlinear Microwave Circuits using Frequency Domain Spectral Balance Techniques: the State of the Art," (invited), *International Journal on Microwave and Millimeter Wave Computer Aided Engineering*, Vol. 1, April 1991, pp. 181-200.

[3] R. Gilmore and M. B. Steer, "Nonlinear Circuit Analysis Using the Method of Harmonic Balance — a Review of the Art: Part I, Introductory Concepts," *International Journal on Microwave and Millimeter Wave Computer Aided Engineering*, Vol. 1, January 1991, pp. 22-37.

[4] R. Gilmore and M. B. Steer, "Nonlinear Circuit Analysis Using the Method of Harmonic Balance — a Review of the Art: Part II, Advanced Concepts" *International Journal on Microwave and Millimeter Wave Computer Aided Engineering*, Vol. 1, April, 1991, pp. 159-180.

[5] M. Lang, "The Intermodulation Problem in Mobile Communications," *Microwave Journal*, Vol. 38, No. 5, May 1995, pp. 20–28.

[6] J. E. Padgett, C. G. Gunther, and T. Hattori, "Overview of Wireless Personal Communications," *IEEE Communications Magazine*, vol. 33, no.1, pp. 28–41, January 1995.

[7] R. M. Clarke, "RF and IC Designers: Two Professions Separated by an Uncommon Language," *Microwave Journal*, June 1995.

[8] A. Mehrotra, *Cellular Radio: Analog and Digital Systems*, Artech House, 1995.

[9] G. D. Vendelin, A. M. Pavio, U. L. Rohde, "Microwave Circuit Design Using Linear and Nonlinear Techniques,"

[10] M. B. Steer and P. J. Khan, "An algebraic formula for the complex output of a system with multi-frequency excitation," *Proc. IEEE*, pp. 177–179, January 1983.