CHAPTER 1

802.11 Flavors and System Requirements

1.1 DEFINITION

What is a wireless local area network (WLAN)? A WLAN system, shown in its most general form in Figure 1.1, consists of a network hardware backbone, along with a series of detached components. These detached components may include computer desktops, computer laptops, personal digital assistants (PDAs), cell phones, gaming systems, security cameras, printers, and appliances as clients. Using radio-frequency (RF) technology,[1] the WLAN system would then allow the clients to access local area network resources while physically being detached from this network. At the same time, the clients are capable of communicating with one another (typically indirectly and through access points rather than peer-to-peer networks) while physically being detached from one another. A WLAN system can transmit data, video, and/or audio.

A WLAN system may be deployed as a stand-alone network or in tandem with a wired network. As compared to a wired network, a WLAN system offers several advantages and suffers some disadvantages.

On the positive side, a WLAN system allows mobility and flexibility. For existing infrastructures, especially those with high user density (hotel rooms, apartment complexes, etc.), it offers the lowest cost and most flexible method of connectivity. Whereas it may be inexpensive to install category 5 (CAT5) wiring for new buildings, to do so in an existing building is quite costly and inconvenient. Given the cost of WLAN chipsets at the current time, it would be much more cost effective to install a simple WLAN system than to run wires through such structures. At the same time, even if CAT5 wiring is installed, for example, in every room in a newly constructed home, it is often not exactly "at the right place." Wireless LAN would offer

[1]Of course an alternative wireless technology such as infrared signaling may be used, but the most common WLAN systems today utilize RF technology. As a result the term "WLAN" is almost exclusively utilized to refer to WLAN communications utilizing RF technology.

Wireless LAN Radios: System Definition to Transistor Design. By Arya Behzad

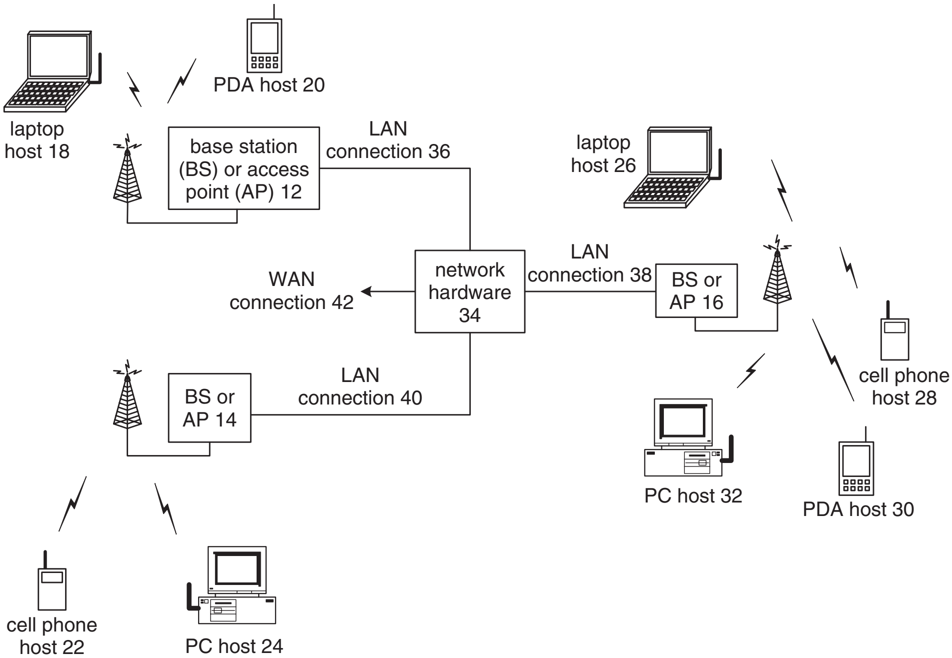


Figure 1.1 Example of WLAN network displaying various associated nodes and backbone network.

the flexibility of connectivity anywhere in the home, without an a priori requirement to determine the precise locations of the network taps.

On the other hand, a WLAN system is typically never as secure as a dedicated (for example, T1) or even shared (for example, cable modem) wire connection. The mere fact that the medium is shared by potentially many users and no physical connection is required to "tap" into the network makes the WLAN network more susceptible to "hacking" and "spoofing." At the same time, various research studies have shown that many WLAN users fail to properly activate the proper encryption options on their access points and thereby make themselves susceptible to hackers.

Recent developments in encryption technology and standards as well as recent software drivers that simplify the installation process of a protected WLAN clients and access points, however, have significantly improved the situation as compared to the early days of WLAN history.

In terms of communication speed, also, WLAN networks are typically a generation or so behind their wired LAN counterparts. This is due to the difficulties associated with the medium of communication (air). For example, in the indoor environment these challenges include propagation losses through the air medium and through walls, multipath caused by reflections from objects and people, and interference due to other wireless communication devices and interferers such as microwave ovens.

It should have become apparent by now that neither a wireless network nor a wired network is capable of providing all the desired characteristics and amenities. Quite often, therefore, an "optimal" network is one that is constructed of a wired LAN "backbone" and is complemented by a WLAN network that would provide flexibility and reconfigurability.

1.2 WLAN MARKET TRENDS

We will spend a few paragraphs discussing the WLAN market trends. The objective here is to put into perspective the phenomenal growth this market has experienced while emphasizing the extremely competitive nature of this market. Thousands of pages of analyst reports are published annually on this subject and we will make no attempt to cover the details that are covered in such reports. Further the WLAN market conditions are quite fluid and change almost quarterly, and therefore the absolute numbers (and possibly even trends) may not hold in the future.[2]

Wireless LAN has been one of the fastest growing segments of the semiconductor market. Despite the slow sales growth (or even decline) of semiconductors for the early 2000 years the WLAN chipset market has grown quite significantly in those years. As seen in Figure 1.2a, the number of WLAN users has grown quite rapidly, especially in the home market. The enterprise has been growing fairly significantly but not nearly as quickly as the home market. The primary reason for this is the concern of the enterprise customer about security. In the early days of WLAN, a major news item about a few University of California—Berkeley Computer Science students breaking the fairly vulnerable 48-bit encrypted WLAN encryption protocol (WEP) did not help the confidence level of the enterprise customers either. By using 128-bit encryption and further enhancements to the security protocols, those issues have been addressed by the standard now (more on this topic later). Of course, the encryption techniques will be continuously updated and strengthened as issues are discovered and as the hackers improve the sophistication of their techniques.

Quality of service (QOS) has also been an issue that has held back the adoption of WLAN by the enterprise as well as certain home users. Certain WLAN applications require a guaranteed maximum latency and would need

[2]Unlike, hopefully, the technical discussions in this book which should hold "forever"!

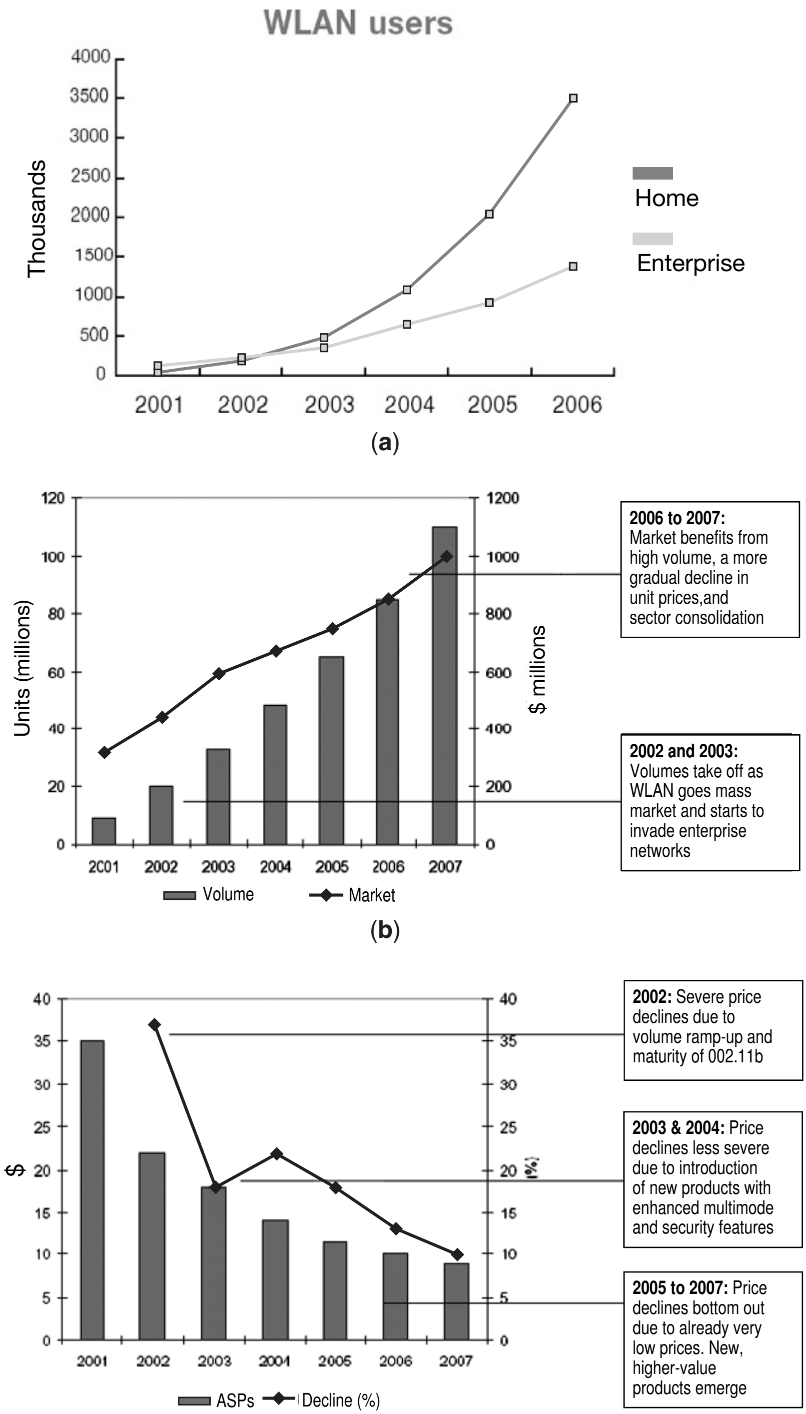


Figure 1.2 (**a**) WLAN growth trend in home and enterprise markets, (**b**) WLAN chipset volume growth chart, and (**c**) historical decline trend in chipset average selling price. (*Sources:* lightreading.com, newsweek.com.)

to be prioritized over other types of network traffic. An example of such latency-sensitive packets is voice-over-Internet protocol (VOIP) packets. VOIP is the standard used to do telephony over an Internet protocol (IP)–based wired or wireless LAN. The resolution and proper implementation of QOS on the WLAN networks would therefore accelerate the adoption and sale of WLAN devices.

Figure 1.2b shows the growth of the 802.11 chipset volumes and the market values extrapolated to the year 2007. The rapid growth of chipset volumes is apparent in this figure and at first may look like an extraordinary business opportunity! However, before trying to put a startup company together to address this market, one needs to review Figure 1.2c. This chart shows the rapid decline in the average selling prices of the chipsets caused by the increase in volume. This steep price drop can be attributed to many factors, such as increase in the selling volumes, very high levels of integration, the numerous players in the market and the resultant competitive nature of the business. In the past few years, the extreme competitive nature of the business has caused many of the smaller and some of the larger players to exit the market segment all together.

Figure 1.2c shows how the average selling prices (ASPs) have dropped very quickly early on as the volumes were ramping up. This period was followed by some price stabilization and then further reduction in prices. The stabilization points correspond to times in the market in which the chipset vendors started offering new features and were therefore able to demand higher prices. This phenomenon temporarily reduces the erosion of price in the WLAN chipset market. For example, in 2003 the steep decline in prices was slowed by the introduction of the 802.11g standard, which allowed for much higher data rates than the traditional 802.11b standard.

Of course, eventually prices will continue their downward trend. It is therefore critical for the chipset industry to keep on innovating and offering newer features. This is necessary in order to be able to offer newer higher margin products as the older ones become commodity items and decline in their profit margins.

A factor that can affect and slow down the reduction in the average selling prices is the addition of new features and new building blocks within the chipsets. So the addition of such blocks into the chips allows the manufacturers of the chips to demand higher prices at the same time the end customer would have a lower bill-of-materials cost.

In summary, this steep price decline and the extreme competitive nature of the WLAN chipset market dictate one of the most important WLAN chipset design requirements: design for low cost. Design for low cost, in turn, translates into design in the lowest possible cost technology, highest levels of integration, smallest possible die size, low packaging and testing cost, and high yields. Since not all of these criteria can be simultaneously satisfied, designers will have to make complex trade-offs to come up with the lowest possible final product cost. Combined with other product requirements such as time to market and system performance, the designers are required to make many difficult choices early on in the design that could quite likely result in a product being successful or a dud.

These trade-offs will be discussed in much more detail in the subsequent chapters.

There are various WLAN standards, such as HyperLAN and the Institute of Electrical and Electronics Engineers (IEEE) 802.11, but at this time, in the United States, Europe, the Far East, as well as elsewhere in world, the 802.11 standard has become the standard of choice for WLAN and will therefore be emphasized in this book.

1.3 HISTORY OF 802.11

In 1990, the IEEE 802 executive committee established the 802.11 working group to create a WLAN standard. The standard specified an operating frequency in the 2.4-GHz ISM (industrial, scientific, and medical) band and began laying the groundwork for a cutting-edge technology. After seven years, in 1997, the group approved IEEE 802.11 as the world's first WLAN standard with data rates of 1 and 2 Mbps. Having great foresight, the executive committee predicted the need for a more robust and faster technology. Therefore, immediately, the committee began work on another 802.11 extension that would satisfy such future demands. Within 24 months, the working group approved two project authorization requests for higher rate physical (PHY) layer extensions to 802.11. The two extensions were designed to work with the existing 802.11 medium access control (MAC) layer, with one being the IEEE 802.11a—5 GHz and the other IEEE 802.11b—2.4 GHz.

The IEEE 802.11 has gained acceptance over competing standards such as HyperLAN and will be the emphasis of this book. The 802.11 is a specific standard that defines the MAC and PHY layers of a WLAN. The original 802.11 standard is a MAC standard plus a low data rate PHY which supports only 1- and 2-Mbps data rates. This first version of the standard operates at the 2.4-GHz ISM band and allows the vendors to choose between a direct sequence spread spectrum (DSSS) and a frequency hopping spread spectrum (FHSS) implementations. As mentioned above, 802.11b is a PHY extension to the original 802.11 standard. It also operates at the 2.40-GHz band and allows for higher data rates of 5.5 and 11 Mbps. It uses a technique known as complementary code keying (CCK).

The 802.11a is another PHY extension to the 802.11 standard. It operates at the 5-GHz unlicensed national infrastructure for information (UNII) band and allows for data rates of 6–54 Mbps. It uses a technique known as orthogonal frequency division multiplexing (OFDM; this technique will be discussed in much more detail in later chapters).

The 802.11g was the next extension to the 802.11 standard. It operates at the 2.4-GHz ISM band and allows for data rates ranging from 1 to 54 Mbps. The 1- and 2-Mbps rates are operated in the DSSS mode whereas the $5\frac{1}{2}$- and 11-Mbps rates are operated in CCK mode. Additionally, rates at 6 to 54 Mbps are operated in OFDM mode. The 802.11g standard borrows the OFDM technique and data rates from the 802.11a standard but operates at the 2.4-GHz ISM band. It can therefore operate at very high data rates while being backward compatible with the 802.11b standard.

In addition to these standards, which have already been approved, the 802.11 committee has “working groups” to evolve and enhance the standard. Here are some examples:

- **802.11e** Tasked to improve QOS. The inclusion of a QOS protocol is essential for tasks that require low latency such as VOIP.
- **802.11i** Tasked to improve encryption. A reliable and hard-to-break encryption technique is essential for the wide adoption of WLAN by the enterprise customer.
- **802.11f** Would allow for an interaccess protocol for easy communication between access points.
- **802.11h** Allows for dynamic frequency selection, and transmit power control. By utilizing dynamic frequency selection, interference between various users would be reduced, and therefore the effective capacity of the cell and therefore the network would increase. Further, by utilizing transmit power control, the minimum required transmit power would be utilized in communication between the access points and the mobile units. This would also reduce cochannel interference and therefore increase the network capacity.
- **802.11n** Allows for multichannel and higher data rate 802.11 in the 2.4- and 5-GHz bands. As of the date of the publication of this book, a “pre-n” standard has been approved by the IEEE, but the final draft has not yet been ratified. The pre-n standard utilizes optional higher order constellations, wider bandwidths, and multi-in, multi-out (MIMO) techniques to dramatically increase the data rate, effective range, and reliability of the WLAN. The 802.11n standard is expected

to be fully backward compatible with the 802.11a and 802.11g standards. We will briefly discuss 802.11n in more detail in Chapter 7.

802.11: b, a, OR g?

The three commonly known versions of the 802.11 PHY are 802.11a, 802.11b, and 802.11g. As described earlier, the 802.11a and 802.11g standards offer much higher speed that 802.11b. However, the advent of 802.11a and g will not necessarily result in the demise of 802.11b in the immediate future. There are applications that would require the lowest power consumption and/or the lowest system cost, and in such cases a stand-alone 802.11b solution may still be the best solution in the immediate future. On the other hand, most system vendors have migrated to 802.11g solutions, which are backward compatible with 802.11b and allow the higher data rates. As the cost of 802.11g solutions drop and their power consumption reduces, this trend will accelerate.

As an alternative to 802.11b and g, if the operator requires a higher data rate, higher user density, and network capacity, he or she would have to choose 802.11a because of the availability of a much wider spectrum at the 5-GHz band and the higher data rates offered by 802.1a.

For longer ranges and higher data rate applications the operator would probably choose 802.11g. The 802.11g offers the added benefit of being backward compatible with 802.11b, which has the largest existing base.

Many applications will probably eventually move to a multiband a/g solution, which would by definition also be backward compatible with 802.11b solutions. This will happen as the cost of multiband solutions drops as a result of further integration and possibly other factors.

Table 1.1 qualitatively shows the advantages and disadvantages of the existing PHY standards. The highlights are listed below.

Currently, there is a much larger existing base for the 802.11b solution. Of course, since 802.11g systems are backward compatible with 802.11b,

Table 1.1 Relative Advantages and Disadvantages of 802.11a, b, and g

Standard	Existing Base	Data Rate	Range	Lack of Interferers	Spectrum Availability	Power Consumption	System Cost
802.11b	++++	+	++++	+	+	++++	++++
802.11a	+	++++	+++	+++	+++	++	++
802.11g	++	++++	++++	+	+	+++	+++

they would be able to take advantage of the 802.11b existing base at lower data rates.

In terms of data rate, the 802.11a and g have an advantage, with rates up to 54 Mbps.

In terms of range of operation, the 802.11b and g have the advantage because they operate at the lower frequency of 2.4 GHz. Since typically propagation losses are lower at lower frequencies, 802.11b and g systems would be able to operate over longer distances as compared to their 802.11a counterpart for a given transmit power and receiver sensitivity. The free-space loss for cases in which the receiver-to-transmitter distance is much larger than the wavelength is given by the relation

$$L = \left(\frac{4\pi d}{\lambda}\right)^2 = \left(\frac{4\pi df}{c}\right)^2$$

where L is the propagation loss, d is the distance between the transmitter and the receiver, λ is the wavelength of the RF signal, f is the frequency of the signal, and c is the speed of light. Antenna gains, absorption losses, reflective losses, and several other factors are not taken into account in the above equation. An indoor environment is much more complex to model or predict than this formula suggests. The interested reader can refer to many publications on this topic.

This simple equation, however, does show the relation between the transmission frequency and the propagation losses. For example, at a distance of 10 m in free space and with the assumptions listed above, a 802.11g system operating at 2.4 GHz would experience 60 dB of propagation attenuation, whereas an 802.11a system operating at 5.8 GHz would experience 68 dB of propagation losses.

The 802.11a has the upper hand when it comes to lack of interferers. This is due to the smaller existing base at the 5-GHz band as well as the wider available spectrum. Additionally, there are far fewer nonwireless LAN systems operating at the 5-GHz band. Such interferers include microwave ovens, security cameras, and cordless phones.

From a spectrum availability point of view, the 802.11a has several hundreds of megahertz of bandwidth available to it (although the exact frequencies would depend on the country of operation). In most countries, on the other hand, there is no more than 100 MHz available for users in the 802.11b or g bands.

From a power consumption point of view, 802.11b would win against the other standards. This is because it utilizes the simplest modulation technique among the three and therefore does not require a high performance radio front end or a sophisticated signal processing baseband. In particular, an 802.11b modulated signal has a small peak to average ratio, and therefore one can use higher efficiency (but lower linearity) power amplifiers on the transmit side.

From a system cost point of view, currently 802.11b offers the lowest system cost. However, the difference in the cost between 802.11g systems and 802.11b systems has been reducing quickly, and today most users are willing to pay the slightly higher cost of an 802.11g system for the significant gains in throughput.

As an interesting marketing point, the number of 802.11g units shipped in The first quarter of 2004 surpassed the shipped 802.11b solutions in that same quarter.

1.5 802.11b STANDARD

As shown in Figure 1.3a, there are a total of 11 designated channels in the 802.11b/g band in the United States. These channels reside in the 2.4-GHz ISM band. However, as shown in Figure 1.3b, there are only three nonoverlapping channels that can operate under the 802.11b/g standard. Within a given cell, if users operate simultaneously on overlapping channels, the interchannel interference would increase, and the overall channel capacity would decrease. The maximum allowed transmit power in the United States for the 802.11b/g standard is 30 dBm or 1 W.[3] This is quite a high transmit power, and most 802.11b/g solutions today operate at significantly lower transmit powers (in the range of 15 to 22 dBm transmit power). This is because the 2.4-GHz ISM band is adjacent to Federal Communications Commission (FCC)–restricted bands. So when operating in the lowest and highest 802.11b/g channels, often the FCC spectral mask requirements associated with these restricted bands is violated before the 802.11b/g mask is violated. Clearly the more stringent of the two masks would set the maximum allowable transmit power.

Worldwide, there are a total of 14 total channels allocated to the 802.11b/g standard operating at the frequency range of 2.40 to 2.58 GHz. The channels are 5 MHz apart. In the United States channels 1, 6, and 11 are typically used to minimize overlap and therefore reduce interference between operating devices. However, as an example, it is possible for a very high power transmitter operating in channel 1 to have an impact on the

[3]Note that this is the average maximum transmit power. Due to potential for large peak-to-average ratio in an OFDM signal, for example, the peak instantaneous power can be significantly more than this.

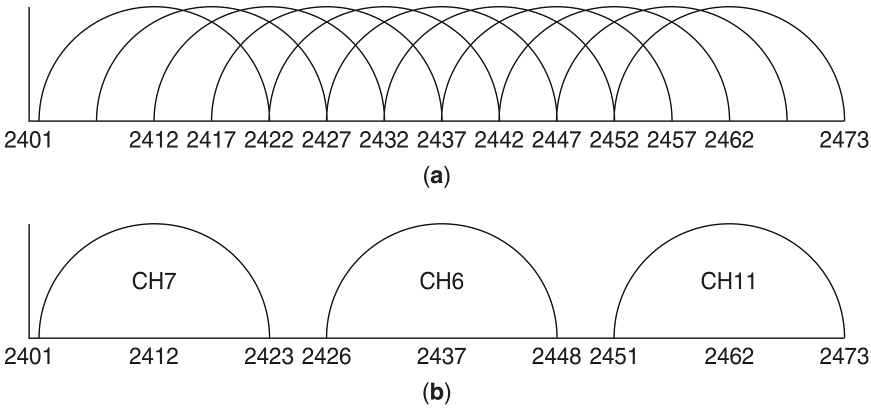


Figure 1.3 IEEE 802.11b/g channel allocations. Note the overlap channels (**a**) as well as the three distinct (nonoverlap) channels (**b**). The x-axis represents frequency in MHz.

throughput of channel 11. Different countries have differing regulations that limit the use of certain channels for 802.11b/g in those countries. For example, in Europe, channels 1 through 13 can be utilized for 802.11b/g operation but at a maximum transmit power of 100 mW. This is done in order to reduce the interference with other ISM band devices.

As mentioned earlier, the original 802.11 standard only allows for 1- and 2-Mbps data rates. In doing so it allows the use of a technique known as DSSS. This technique spreads the data over a wide bandwidth to gain immunity to interferers and multipath reflections. The technique is similar to what is used for the IS-95 cellular code division multiple-access (CDMA) standard.

As an alternative the original standard allows for a FHSS technique. This technique is also designed to improve the immunity of the signal to interferers and multipath channel reflections but, as the name suggests, relies on the carrier frequency to hop around at a pseudorandom center frequency basis. The FHSS technique is similar to what is used in the Bluetooth (BT) standard.

The 802.11b extension to the standard allows for the introduction of higher data rates of 5.5 and 11 Mbps. The 802.11b relies on CCK, a distinct nonsystematic block code which offers both spreading as well as a minimal amount of coding gain. In a sense it can be viewed as a special case of DSSS.

As is typical for any system and any modulation, the signal-to-noise (SNR) requirement for the higher data rates is higher than those for the lower data rates. As such the standard requires a minimum system sensitivity of

–80 dBm for the 1-Mbps data rate and a minimum system sensitivity of –76 dBm for the 11 Mbps. However, today, most systems are capable of delivering much better sensitivity numbers than the standard requires. A state-of-the-art system today can achieve about –98 and –91 dBm "chip sensitivity," respectively, for the 1- and 11-Mbps data rates. The system sensitivity is typically 1 to 2dB worse than the chip sensitivity for the 802.11b operation due to losses of front-end components such as baluns, filters, switches, and board traces at 2.4 GHz.

Table 1.2 summarizes the modulation types and the sensitivity numbers for the various 802.11b data rates.

The 802.11b standard is, in principle and as compared to 802.11g and especially 802.11a, fairly easy to implement. The standard achieves a maximum of 11 Mbps over an equivalent noise bandwidth of 11 to 15 MHz depending on the implementation. This results in a comparatively low spectral efficiency of <1 bit/s/Hz. As a reference, note that a maximum spectral efficiency of > 3 bits/s/Hz is achieved for the 802.11g and 802.11a standards. Of course, in general, wireless communications are limited to much lower spectral efficiencies than those of their wireline counterparts due to the much inferior communication medium (channel). For example, digital subscriber line (DSL) systems, gigabit Ethernet, or cable systems can achieve spectral efficiencies in excess of 10 bits/s/Hz.

Additionally, the 802.1b modulation has a low peak-to-average power ratio (PAPR). This is by no means a constant-envelope modulated signal (like that of Bluetooth, for example), but neither does it have very large PAPR associated with the OFDM coding utilized in the 802.11a and 802.11g standards. The low PAPR characteristic of the 802.11b standard makes the modulation somewhat immune to nonlinearities in the signal path. This characteristic in particular makes the implementation of the power amplifier (PA) in the transmit path much simpler than those required for the 802.11a and g standards.

Table 1.2 IEEE 802.11b/g Allowed Data Rates, Associated Modulation Types, and Required Sensitivities

Data Rate (Mbps)	Modulation	Sensitivity Requirement (dBm)	State-of-the-Art Chip Sensitivity (dBm)
1	D-BPSK	–80	–98
2	D-QPSK		–96
5.5	CCK		–93
11	CCK	–76	–91

Note: Obtained state-of-the-art sensitivity levels are also reported.

1.6 802.11a CHANNEL ALLOCATION

As mentioned earlier the 802.11g channel allocation is identical to that of 802.11b (Fig. 1.3). As such, there are only three nonoverlapping channels available to the users.

One of the advantages of the 802.11a standard as compared to the 802.11g standard becomes apparent in Figure 1.4: There are currently a total of 12 “nonoverlapping” channels available in the United States with proposals at the FCC to open up even more spectrum in the 5-GHz band as part of an expanded unlicensed National Information Infrastructure (NII) spectrum. The large number of channels available in the 802.11a band allow for much higher overall cell and network capacity and less interchannel interference.

As can be seen in Figure 1.5, the statement about the 802.11a channels being nonoverlapping is not completely correct. The spectrum associated with the information content of each channel is designed to be nonoverlapping with its adjacent channels. However, because of imperfect filtering as well as nonlinearities and spectral regrowth in the system, there is a limited amount of spectral leakage from each channel which leaks into its adjacent channels. The magnitude of this leakage is highly regulated by the spectral mask requirements of the standard. The performance of the system in the presence of adjacent channel interferers is also regulated by the standard (more on this later).

In the United States the maximum allowed transmit power for the 802.11a standard is dependent on the subband (Fig. 1.5). In the lower, mid, and higher 802.11a subbands, the maximum transmit power is limited to 16,

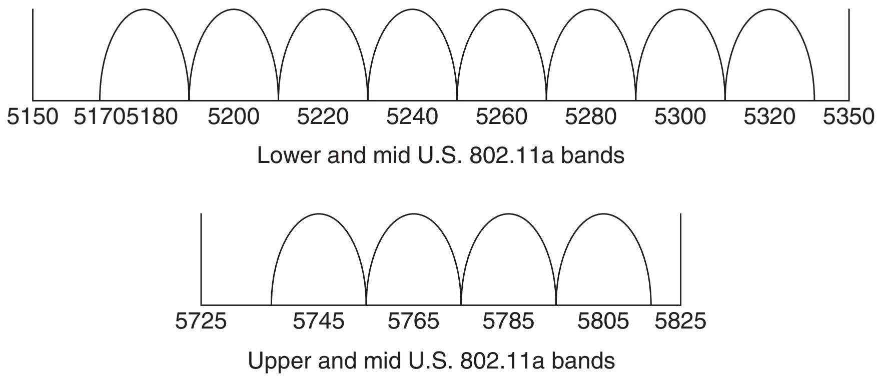


Figure 1.4 Detail of IEEE 802.11a channel allocations in U.S. (total 12 nonoverlapping channels). The lower, mid, and upper bands are shown. Note that no overlapping channels are allowed.

23, and 29 dBm, respectively. The higher subband is primarily intended for long-range outdoor communications.

Various countries allocate different frequency bands for the 802.11a standard. In general, 802.11a systems around the world (non-U.S.) operate in the 4.92- to 5.70-GHz spectrum (Fig. 1.5). Recent proposals have worldwide channels operating as high as 5.845 GHz. For various countries, not only the dedicated frequency channels but also the maximum transmit power per channel as well as various other requirements vary. The interested reader should refer to specific regulations of a given country.

1.7 802.11a AND 802.11g: OFDM MAPPING

The 802.11a and g utilize a technique known as orthogonal frequency division multiplexing, or OFDM. Conceptually, OFDM has been around for a long time. It has been used in a variety of applications for years. These include such applications as digital video broadcasting (DVB) and digital subscriber line (DSL). OFDM does require a significant amount of signal processing horsepower, and such horsepower until recently would consume quite a bit of power consumption. Clearly a high power consumption chipset would not be very suitable for portable applications.

Recent advancements in process technology and also low power design techniques have enabled a dramatic reduction in power consumption of OFDM-based modems. These modems are therefore now suitable for many portable applications such as computer laptops. The push for reducing the power consumption of OFDM-based modems, of course, continues. Further reductions in power consumptions are enabling the integration of WLAN systems in some of the most power-sensitive consumer application gadgets.

OFDM provides a good degree of immunity to multipath fading, which is typically a major problem for high speed wireless communication, especially in an indoor environment. In order to comprehend the concept of multipath fading and its impact on high speed communications in an indoor environment, a brief discussion of the topic is presented in the following section.

1.7.1 Multipath Fading

Multipath propagation, or in short multipath, occurs when signals reflect off of various objects and even people and add constructively or destructively at the receiver antenna. When the signals add destructively, they can significantly impact the quality of the link. This can result in a significant reduction in the throughput of the system. Figure 1.6 depicts multipath when a direct line-of-sight (LOS) path does exist. Figure 1.7 depicts a scenario in

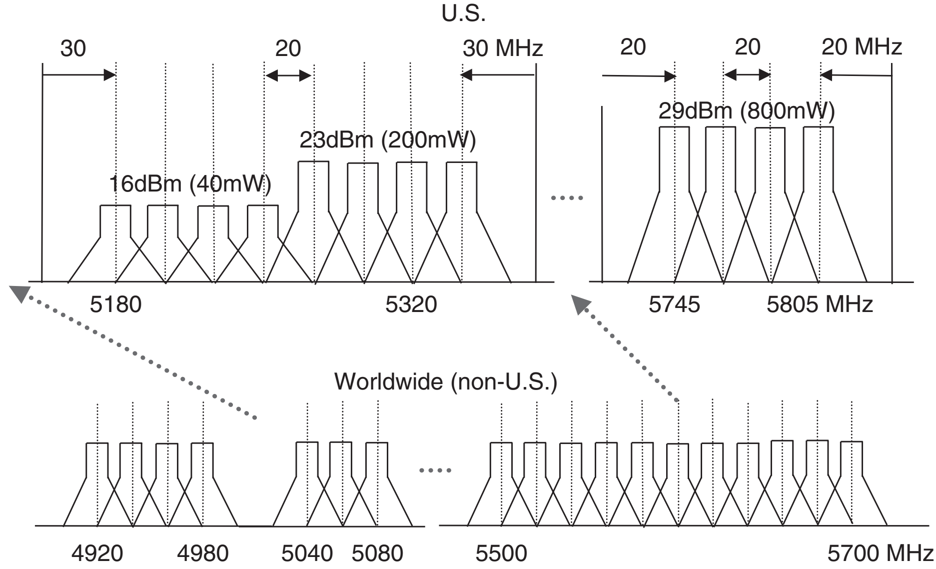


Figure 1.5 Associated power levels for U.S. IEEE 802.11a subbands. The additional worldwide 802.11a subbands are also shown. Note that, although the main channels are nonoverlapping, the channels can interfere with their adjacent channels (as shown) due to inadequate filtering or spectral regrowth.

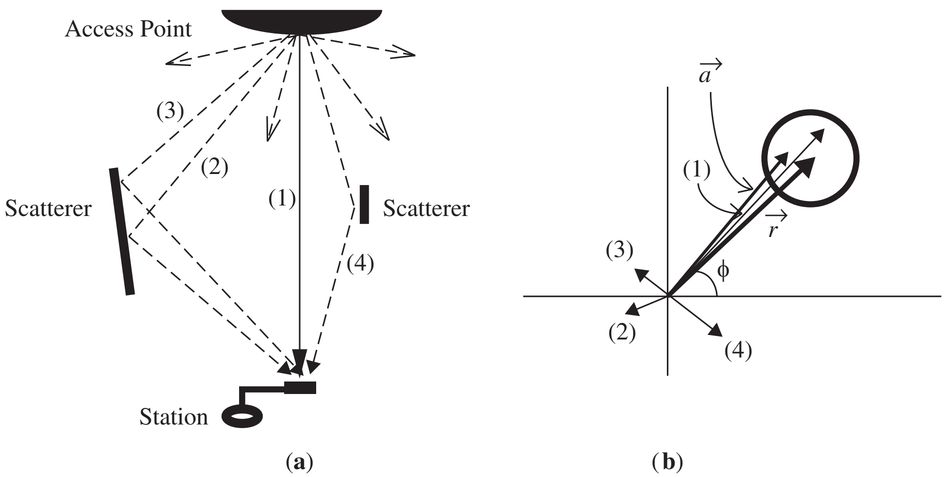


Figure 1.6 (**a**) Multipath in presence of a line-of-sight signal. (**b**) Vector space representation.

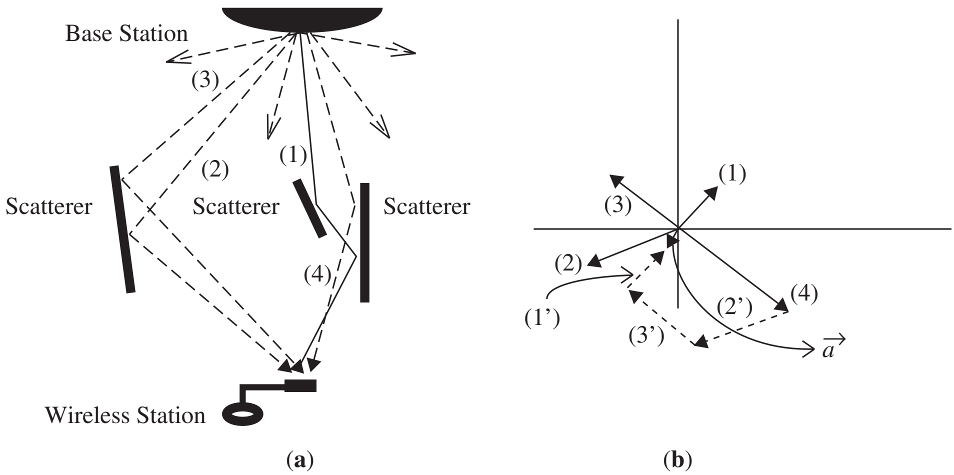


Figure 1.7 (**a**) Multipath response in absence of a LOS signal. (**b**) Vector space representation. Note that the vector magnitudes have been scaled 2 : 1 as compared to Figure 1.6 to simplify visualization.

which a direct LOS does not exist. Clearly, in the latter case, the resultant received signal can be quite small.

Multipath fading is very much environment specific but typically does not exceed about 20 dB in an indoor environment with carrier frequencies in the few GHz range. As described above, multipath is a phenomenon caused by the multiple arrivals of the transmitted signal to the receiver due to reflections off of "scatterers." The gain and phase of these reflections can be modeled as being somewhat random. Multipath is usually much more of a problem if a direct LOS path does not exist between the transmitter and the receiver. In this scenario, the change in the magnitude of the received vector as compared to the mean value of the magnitude of the received vector is small, resulting in a Ricean distribution (Figure 1.6). Figure 1.6b shows the vector space representation of the multipath reception in the presence of a LOS path. The vector represents the resultant vector from the LOS path (1) and the multipath receptions (2), (3), and (4). The magnitude of vector represents the mean value of the possible resultant vectors. The area of the circle indicates the 50% contour for this Ricean distribution.[4] It is clear from this figure that a multipath response may not affect the decision variable significantly in such a scenario.

[4]Ricean and Rayleigh fading models are the most common fading models applied to analyze propagation in indoor environments. The names of these fading models are derived from their underlying probability distribution function (PDF) statistics. A Rayleigh fading typically occurs if there are several indirect propagation paths between the transmitter and the re-

Figure 1.7a displays the multipath channel in the absence of an LOS path. Figure 1.7b shows the vector space representation of such a response. Vectors (2), (3), and (4) represent the reflected signals at the receiver. Vector (1) represents the intended LOS signal which has been interrupted and reflected multiple times by the scatterers. Vectors (1′), (2′), (3′), and (4′) represent the vectors used to find the resultant vector, **a**. It is clear that vector **a** is very small in magnitude, resulting in a high probability of error at the slicer. For large number of scatterers, the channel can be modeled to have a Rayleigh distribution, with about 10% probability of a resultant vector with a magnitude less than half the magnitude of the mean. Note that in this case the mean ±25% contour in the vector space is not a circle because of the asymmetry of the Rayleigh density function about its mean value.

In a typical indoor environment (office, home, etc.) root-mean-square (RMS) delay spreads[5] of 50 to 75 ns can be observed. The worst case RMS delay spreads in these environments can be as large as 150 ns. In order to establish a traditional high data rate communication in such an environment, a very high symbol rate corresponding to a short symbol duration would be required. The larger the value of the RMS delay spread as compared to the symbol duration, the more intersymbol interference (ISI) would be generated. ISI can be corrected in the digital domain, but very high speed and typically high power consumption time-domain equalizers would be needed.

With an understanding of multipath, the benefits of OFDM coding can now be discussed in more detail. OFDM coding is a technique that can be quite powerful in reducing the effects of multipath on high speed communications.

With OFDM, the transmitted data are modulated onto multiple subcarriers. This is accomplished by modulating the subcarriers' phase and amplitude. As such, the original high data rate stream is split into multiple lower rate streams and then mapped on to the available subcarriers (which are multiples of a given frequency) and then combined together using an in-

ceiver with none of the paths being a dominant path (i.e., with distinctively larger magnitude than the others). In this situation, the received signal is comprised of the sum of multiple independent random variables and at the limit can be approximated as having a Gaussian distribution function. In reality, Rayleigh fading is really a worst case in which no path dominates. However, since Gaussian PDFs are very well understood and can easily be modeled mathematically, they present a convenient mathematical tool for analyzing the worst case propagation characteristics. On the other hand, Rayleigh fading typically applies if a dominant propagation path (such as a LOS path) between the transmitter and the receiver exists. In this case the PDF is "centered" around the magnitude set by the dominant propagation path.

[5]RMS delay spread is defined from the characteristics of the delay spectrum of a stochastic process. It can be thought of as an indication of the delay between the earliest arriving "rays" and the latest arriving rays.

verse fast Fourier transform (FFT) operation. In creating N parallel transmit streams, the bandwidth of each stream is reduced by a factor N that can be selected in such a way that the RMS delay spread of the channel is much less than the symbol period. This results in a significant reduction in the ISI. A well-designed OFDM system does not therefore require a time-domain equalizer.

The transformations utilized by OFDM are the discrete Fourier transform (DFT) and the inverse discrete Fourier transform (IDFT). The orthogonality of the OFDM signal is obtained through the use of multiples of the subcarrier frequency over an integer cycle which is an inherent property of the DFT and IDFT transformations.

In Figure 1.8 a single subcarrier is displayed in the frequency domain. The OFDM signal is constructed by the summation of multiples of such single subcarriers, as shown in Figure 1.8 (this is an example with five subcarriers). It is clear from this figure that the subcarriers are allowed to have overlap not only with their adjacent subcarrier but also virtually with all of the other subcarriers.

For those familiar with the CDMA technique utilized in many of today's cellular phones, the following analogy may be useful. The construction of an OFDM signal with multiple sinusoidal subcarriers is somewhat similar to the construction of a CDMA signal using orthogonal Walsh codes (Walsh codes are a family of orthogonal codes which are based on "square waves" rather than sinusoids). The main difference between CDMA and OFDM is that in the case of CDMA the orthogonal Walsh codes are primarily used as a means for multiple access,[6] whereas the orthogonal sinusoids in the OFDM coding are primarily used to gain immunity to multipath.

The fact that subcarrier overlaps are allowed enables the spectral efficiency of an OFDM-coded signal to be increased. It is easy to see that with no subcarrier overlap the same number of subcarriers (which is related to the amount of data being communicated) would occupy a much wider spectrum. This would clearly reduce the spectral efficiency. This concept is shown graphically in Figure 1.9 in a simplified diagram.

The obvious question that may arise is the potential interference caused by the overlapping of the subcarriers. However, due to the inherent orthogonality of the subcarriers of the OFDM signal, the peak of each subcarrier occurs at the null of all other subcarriers, as seen in Figure 1.8. Under ideal conditions, this would mean that the subcarriers do not interfere with one another.

Unfortunately, under real-world conditions, various impairments could cause the perfect orthogonality of the subcarriers to be violated. These in-

[6] Note that CDMA also provides immunity to multipath due to the spreading of the signal.

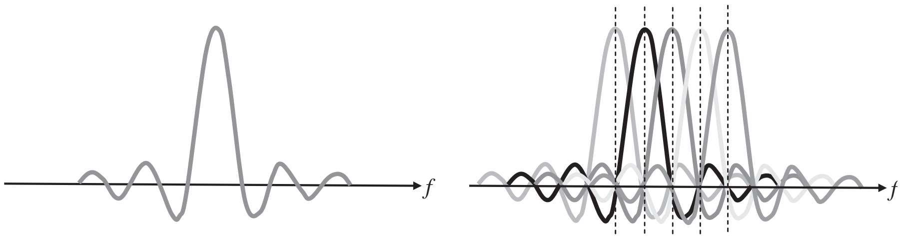


Figure 1.8 Construction of OFDM signal from its individual components (subcarriers). Note the tight "packing" of the subcarriers and the spectral efficiency achieved. Also note that each subcarrier's peak occurs when the other subcarriers are at a null.

clude impairments such as phase noise, quadrature imbalances, distortion, and uncorrected frequency offsets. The location of each subcarrier's peak would shift relative to the other subcarriers in such a way that the peak of one subcarrier would no longer be aligned with the null of the other subcarriers. Such impairments would give rise to "intersubcarrier interference." These impairments and their impact on the OFDM signal and the overall system will be studied in great detail in Chapter 3.

As any good engineer would guess, an OFDM-coded signal could not have all these great properties without some trade-offs.

Probably the biggest "difficulty" with using OFDM-coded data is that it tends to generate very large peak-to-average ratio (PAR) signals. The large

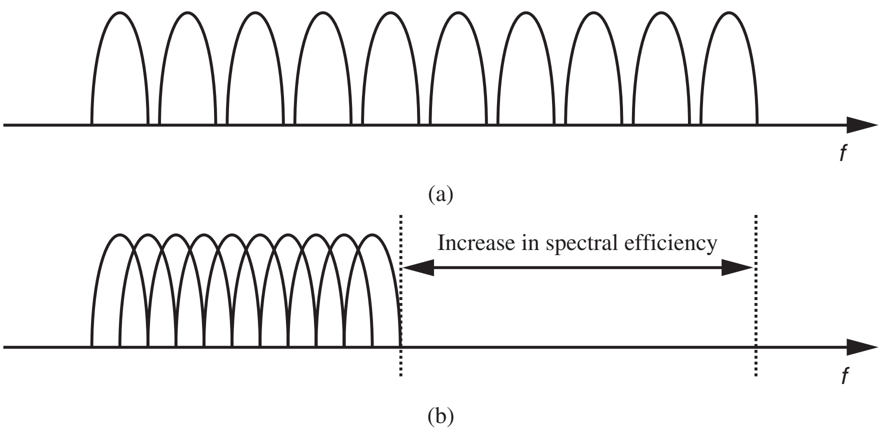


Figure 1.9 Increasing the spectral efficiency of the modulation by using the orthogonal properties of the OFDM signal and packing the subcarriers and their associated data content closer to one another.

PARs significantly complicate the design of the radio and the mixed-signal blocks. The signal path will have to be designed with much more severe linearity constraints than traditional non-OFDM modulations. In particular, on the transmit signal path, the design of the power amplifier becomes quite challenging. Not only is designing high linearity power amplifiers (required by OFDM modulation) quite challenging, but such amplifiers have much worse efficiencies than their nonlinear counterparts.

The topic of the high PAR OFDM-modulated signal and its implications on the power amplifier design will be covered in more detail in Chapter 3.

Now that the general concept of OFDM has been introduced, some of the specifics of 802.11a/g OFDM coding will be discussed.

The 802.11a/g OFDM signal is constructed from 52 total subcarriers, as shown in Figure 1.10. These subcarriers are indexed from –26 to +26, with the zeroth subcarrier eliminated. Out of the 52 subcarriers, 48 are dedicated to carrying the desired data (payload), and 4 of the subcarriers are designated with the task of carrying the "pilot" information.

The subcarrier index numbers for the pilots are –21, –7, 7, and 21. The pilot subcarriers are always modulated in binary phase shift keying (BPSK)[7] format, which is a very simple but robust modulation. The pilot tones are primarily used to help establish a robust "link" before the reception of the desired data (payload) can begin. As such they allow the receiver to set the proper gain, track and correct the carrier frequency offsets, adjust and correct the analog-to-digital conversion (ADC) sampling frequency offsets, and so on. If these tasks are not done properly, the entire packet is likely to be lost, and the effective throughput of the link is significantly reduced. The BPSK modulation, due to its inherent simplicity, is quite robust to various analog and channel impairments such as multipath distortion, phase noise, and quadrature imbalances. This is the reason for transmitting the pilot subcarriers in BPSK format.

The 802.11a/g OFDM subcarriers are spaced 312.5 kHz apart and occupy an overall channel bandwidth of 16.25 MHz,[8] which occupies a baseband bandwidth of –8.125 to +8.125 MHz. The zeroth subcarrier has been eliminated in the 802.11a/g standard and is not used as a pilot or payload subcarrier. This fact has very important implications in the choice and design of the radio architectures used for 802.11a/g solutions. This topic will be discussed in detail later in the book.

The channel-to-channel spacing in the 802.11a standard is 20 MHz. In the 802.11g standard this spacing is set to 25 MHz. The difference between

[7]BPSK is the simplest form of the phase shift keying (PSK) modulation family. It is also the same as the simplest form of a quadrature amplitude modulation or QAM-2.

[8]52 subcarriers × 312.5 kHz/subcarrier = 16.25 MHz.

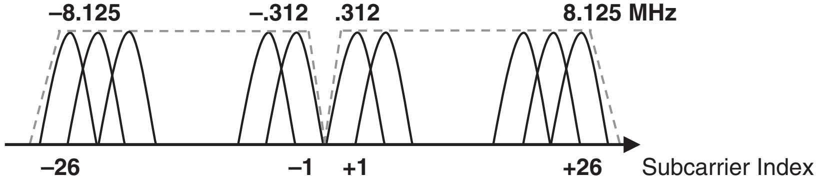


Figure 1.10 Construction of IEEE 802.11a/g OFDM signal from 48 data and 4 pilot subcarriers.

the occupied modulation bandwidth (16.25 MHz) and the channel-to-channel spacing is used to reduce the effects of adjacent channel interference which occur due to imperfections in the transmitter and the receiver.

1.8 802.11a/g: DATA RATES

The various data rates allowed in the 802.11a/g OFDM mode are shown in Table 1.3. As can be seen, the data rates range from 6 to 54 Mbps. The data rates are varied from the highest to the lowest rates by changing one or both of the following modulation-related parameters: (a) modulation order and (b) coding rate.

The modulation order is the primary tool used to adjust the data rate for 802.11a/g. At the higher order modulations, for a given transmit power and with everything else being the same, the spacing between the neighboring constellation points on a constellation diagram is less than those of lower order modulations. This makes the modulation much more susceptible to im-

Table 1.3 802.11a/g Data Rates, Modulation Types, Coding Rates, and Required Sensitivity Levels Set by Standard

Data Rate (Mbps)	Modulation	Code Rate	Sensitivity Requirement (dBm)	State-of-the-Art Chip Sensitivity (dBm)[a]
6	BPSK	$\frac{1}{2}$	–82	–94
9	BPSK	$\frac{3}{4}$	–81	–92
12	QPSK	$\frac{1}{2}$	–79	–90
18	QPSK	$\frac{3}{4}$	–77	–87
24	QAM-16	$\frac{1}{2}$	–74	–84
36	QAM-16	$\frac{3}{4}$	–70	–82
48	QAM-64	$\frac{2}{3}$	–66	–76
54	QAM-64	$\frac{3}{4}$	–65	–74

Note: Representative state-of-the-art sensitivity levels are also specified.
[a]Using hard Viterbi decoding can improve the sensitivity of higher order modulations by as much as 2.5 dB.

pairments such as circuit noise, phase noise, and in-phase/quadrature phase (I/Q) imbalance.

The code rate determines the amount of redundancy and hence robustness built into the modulation. The closer the coding rate to unity, the less the amount of redundancy built in, and the higher the data rate (the data are not "wasted" for the sake of redundancy).

The coding rate is another tool utilized to adjust the data rate. Typically, however, the change in data rates as a result of a change in the coding rate is much smaller than that of changing the modulation order. This is because coding rates much larger than $\frac{5}{6}$ do not provide enough redundancy to be useful and are therefore not typically used in practice. Several examples of changing the data rate by utilizing various coding rates are shown in Table 1.3.

In a real system, the control of the actual data rate selected by the link is done through the media access controller. The goal of MAC is to establish the fastest (but reliable) link possible. As such, it typically starts at the highest data rate and tries to establish a robust link. If it fails to do so, it will drop the rate to a lower rate and retry. It will continue this process until it establishes a link or determines that no link can be established. Detailed discussions of the MAC layer are beyond the scope of this book and the interested reader can refer to the references.

The IEEE 802.11a/g standards require any system that claims compatibility to the standard to be able to maintain certain minimum sensitivity levels (ranging from –65 to –82 dBm for the various data rates). The minimum required sensitivity level by the standard for the various data rates is listed in Table 1.3. Today's systems can significantly outperform the specifications for sensitivity which have been set by the standard. Table 1.3 also shows examples of the capabilities of today's state-of-the-art integrated solutions referred to the input of the chips. In general, the performance of the state-of-the-art solutions is about 10 dB superior to those required by the standard. It is important to note that several assumptions have been made in specifying the sensitivity of the state-of-the-art solutions: (a) the sensitivity numbers specified are referred to the chip input (i.e., the board losses, which can range from 1 to 3 dB are not accounted for); (b) no "external" (nonintegrated) low noise amplifiers (LNAs) are assumed in front of the receiver chip; and (c) hard Viterbi decoding is assumed for the baseband section of the receiver.[9]

It is interesting to note the inverse relationship between the data rates and the minimum sensitivity of the various modes of operation shown in Table

[9]A soft Viterbi decoder would improve the performance numbers specified by as much as 2.5 dB for the higher data rates as compared to the numbers shown in Table 1.3. It will improve the sensitivity for the lower data rates marginally, however, since at the lower data rates the sensitivity is often limited by the problem of "detection" (i.e., whether there is a packet present).

1.3. As the data rates are increased (through increasing modulation order or by using higher coding rates), the minimum sensitivity level suffers. Given the explanation earlier, this should be rather obvious and is related to the larger SNR required by the higher data rates. In other words, as the data rate increases, a higher received power level is required in order to be able to receive the signal (assuming noise levels stay constant). The absolute level of the SNR required for each data rate is dependent on various factors (soft versus hard Viterbi decoding as an example) but is in all cases higher than that of a lower data rate (all else being equal).

Although not shown in Table 1.3, it is a similar situation on the higher end of the power range. The 802.11a/g standards do specify the minimum high end power rate that the receiver should be able to receive (–30 dBm). However, unlike the minimum power level requirements, at the high end the power levels are not specific to each data rate. In reality, though, the higher data rates are much more susceptible to "high power impairments" such as nonlinearities in the receiver (and transmitter). So the receiver would quite likely be able to tolerate much higher receiver power levels for a 6-Mbps link than a 54-Mbps link. This should be obvious by considering the fact that high power impairments such as nonlinearities cause the constellation points on a constellation diagram to deviate from their ideal point and get closer to the neighboring constellation points. Since for a given transmit power the spacing between the constellation points on a high order modulation is larger than that of a low order modulation, the low order modulation would be able to handle much more nonlinearities before it causes an error.

As a side note, given our knowledge of the 802.11a/g and that the duration of each symbol is 4 μs, we should now be able to calculate each one of the data rates listed in Table 1.3. For example, the 54-Mbps data rate can be calculated as follows:

$$48 \text{ (data subcarriers)} \times 6 \text{ (bits/symbol for QAM-64)}$$
$$\times \tfrac{3}{4} \text{ (code rate)} \times \tfrac{1}{4} \text{ } (\mu\text{s}) = 54 \text{ Mbps}$$

For the 6-Mbps data rate

$$48 \text{ (data subcarriers)} \times 1 \text{ (bits/symbol for BPSK)}$$
$$\times \tfrac{1}{2} \text{ (code rate)} \times \tfrac{1}{4} \text{ } (\mu\text{s}) = 6 \text{ Mbps}$$

It is important to make one final point on Table 1.3. For 802.11g, this table only shows the OFDM-related rates. As mentioned earlier, 802.11g is backward compatible with 802.11b and as such is capable of operating at all the lower data rates (11, 5.5, 2, 1 Mbps) at which 802.11b is capable of operating.

1.9 802.11a/g OFDM PACKET CONSTRUCTION

The physical layer convergence protocol (PLCP) layer allows for a common MAC layer to be used for many 802.11 WLAN PHY substandards. The construction of an 802.11a/g OFDM packet is shown in Figure 1.11. As shown in the figure, the packet is comprised of four basic components. The first piece of the symbol is what is known as the "short" preamble. The short preamble is always 8 μs long in duration. During the short preamble, tasks such as automatic gain control and coarse frequency offsets are calculated and adjusted for in the baseband chip. The short preamble is followed by a "long" preamble, which is also 8 μs long.[10] During the long preamble, tasks such as channel estimation, fine frequency offset adjustments, and timing recovery are performed. Again note that the short and long preambles are always communicated in BPSK in order to maintain robustness.

The field that follows the preambles is called the "signal field." It is here that the information relating to the modulation order, coding rate, and packet length is carried. The short and long preambles along with the signal field constitute the PLCP protocol data unit (PPDU).

The actual payload (which carries the desired data to be communicated) follows the signal field. So the PPDU plus the user data together constitute the full packet.

Note that the PLCP preamble is made of 12 OFDM symbols, the signal field is made of 1 OFDM symbol, and the user data are made of a variable number of OFDM symbols.

1.10 802.11 SYSTEM REQUIREMENTS

We will now shift our focus and discuss some of the more important system requirements for the 802.11 standard.

1.10.1 Receiver Sensitivity

We will start by discussing the receiver sensitivity requirements. In general the receiver sensitivity is affected by many factors, including the design of the radio and the baseband PHY layer.[11] Examples of the latter case are the choice of the Viterbi decoder (soft vs. hard), the design of the channel estimator, and the gain control algorithm.

[10]Note that both the short and long preambles are 8 μs in duration. However, the short preamble is constructed of 10 smaller (800-ns) segments—hence the word *short.*

[11]Purist radio designers find the latter to be hard to believe!

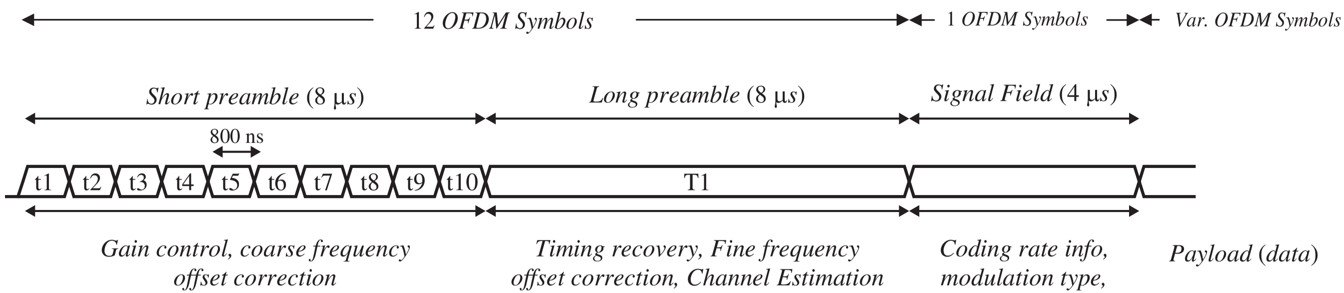


Figure 1.11 Construction of IEEE 802.11a/g packet. The packet is made of the short and long preambles, the signal field, and the data payload.

On the radio side, the factors impacting the minimum sensitivity levels are related to the data rate selected. In general, under low data rate conditions (e.g., 6 Mbps), the sensitivity of the system is primarily set by the receiver noise figure and cochannel interference.[12] For the most part other impairments such as quadrature imbalances would have minimal impact at the sensitivity levels of the lower data rates.

The situation is quite a bit more complex at the higher data rates (such as 54 Mpbs), however. It is clear that noise figure and cochannel interference will still impact the sensitivity levels, but other impairments such as phase noise, quadrature imbalance, transmitter error vector magnitude (EVM), center-frequency inaccuracies, filter corner inaccuracies, multipath, sampling frequency inaccuracies, and gain control inaccuracies will also enter the picture and impact the sensitivity levels. Once again, it is clear that the higher data rates would require a higher SNDR (signal to noise plus distortion level) to be able to operate properly and would therefore have a more limited power range in which they can operate robustly.

The discussions of the previous few paragraphs assume that there are no significant interferers present. Under interference-dominated conditions (conditions in which the desired signal level is significantly smaller than an interfering signal such as a large adjacent channel signal), the linearity of the receiver will also become a factor in determining the sensitivity of the system. Under interference-dominated conditions, a "smart receiver" would need to be able to detect the interference condition (through the use of proper RSSIs, for example) and set the front-end gain control accordingly. This often translates into a trade-off between linearity and noise figure. In general, the gain of the front end would have to be set to the highest level possible (corresponding to the lowest noise figure possible) while avoiding significant nonlinearities in the receiver chain. So, in summary, in interference-dominated conditions, in addition to the linearity of the receiver, all the factors mentioned above for the non-interference-dominated conditions are important and must be considered in the design.

1.10.2 Transmitter Error Vector Magnitude

Another system requirement specified by the 802.11a/g standard is the transmitter EVM, which, in general is a single scalar number that is an indication of the modulation quality of the signal. To calculate the EVM, one needs to compare the actual symbols with their ideal impairment-free sym-

[12]Cochannel interference is caused by users in adjacent cells operating at the same frequency as the user. Since the user data in the adjacent cell is uncorrelated to the user data in the current cell, this would cause a noiselike effect on the user and would degrade the sensitivity levels.

bols on the constellation diagram and compute the error vectors as shown in Figure 1.12. The real symbol will have a different phase and amplitude as compared to the ideal symbol constellation points. Systematic and deterministic errors would simply offset the real constellation points as compared to the ideal ones. Nonsystematic impairments such as noise, however, would cause an "error ball" or "error cloud" of uncertainty in the constellation points about the ideal constellation points.

Mathematically, for a given symbol, EVM is defined as

$$\mathrm{EVM} = \sqrt{\frac{\sum_{i=1}^{M} |Z(i) - R(i)|^2}{\sum_{i=1}^{M} |R(i)|^2}}$$

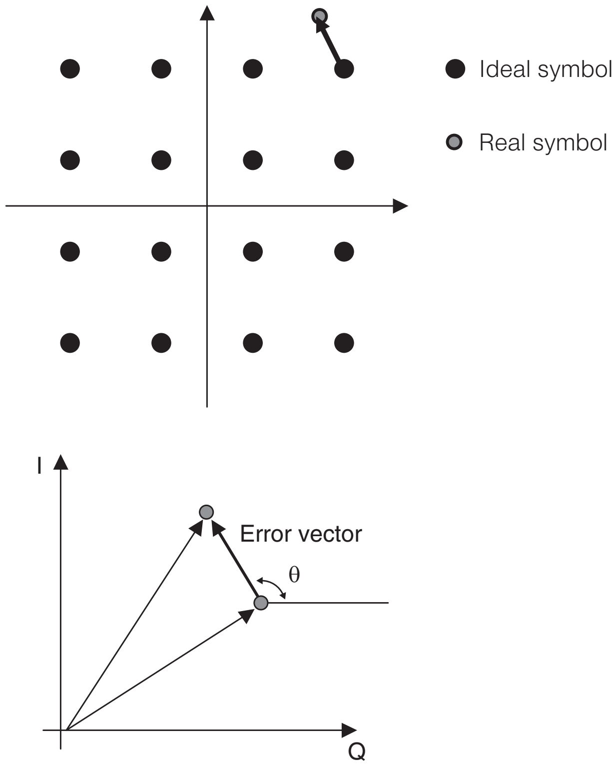


Figure 1.12 Pictorial description of the concept of EVM for 16-QAM constellation diagram. The detail (zoom in) of the EVM calculation applied to a single constellation position is also displayed.

where Z is the measured signal, R is the reference (ideal) signal, M is the number of measurements, and i is the measurement index. Various standards allow for some form of tracking of the symbols (such as timing recovery and frequency offset corrections) before applying this equation to measure EVM. This definition can be extended to all the symbols of a modulation by averaging over all these symbols.

It can be seen from the definition that the EVM is an estimate of the magnitude of the error signal as compared to the magnitude of the ideal signal. As such, it is clear that that the maximum value for the EVM is 1, or 100%, and that the minimum value of the EVM is 0. It is clear that the smaller the value of the EVM in percent, the higher the quality of the modulated signal. The EVM can be expressed in decibels as 20 log(EVM). For example, 1% EVM can be expressed as –40 dB EVM.

However, in some cases different standards specify the EVM definition in slightly different terms. In the case of 802.11a/g OFDM rates the EVM is defined as the RMS of the error vectors over all symbols. In contrast, in the case of 802.11b/g CCK rates, the EVM is calculated using the peak values.

The EVM is affected by a variety of impairments. These include nonlinearities, phase noise, quadrature imbalances, and filter shapes and bandwidth. The reason that the EVM is so commonly used as a measure of the quality of the transmitter is that it is essentially impacted by many such impairments. Of course, one should note that there are impairments that may not impact the EVM as measured by a vector signal analyzer (VSA), and conversely there are impairments that may impact the EVM as measured by a VSA but may not have a significant impact on the packet error rate as measured by the actual receiver. More on this point will be discussed later.

For the 802.11a and g standards, at 54 Mbps, the transmitter EVM (or transmitter quality) is required to be a minimum of –25 dB. On the other hand, there are no specifications given for a receiver EVM. The receiver is completely specified by the specification of the sensitivity for a given data rate. If there are impairments that would affect the receiver EVM, they will quite likely be evident in the sensitivity measurements of the receiver, especially at the higher data rates, and this will impact only that receiver. On the other hand, a problem in the quality of the transmitted signal can cause another user (with potentially a perfectly good receiver) to be unable to operate at the proper data rate. This is the reason that the standard enforces the EVM on the transmitter but not on the receiver.

1.10.3 Transmitter Spectral Mask

Another 802.11 system requirement on the transmitter side is the satisfactory passing of the spectral mask. To facilitate the discussion of spectral mask, four observations need to be made here:

1. A linear system is one that faithfully replicates the input signal by a constant multiplier (gain or attenuation) without the introduction of any frequencies (harmonics, intermodulations, etc.) that did not exist in the incoming signal. Of course, a constant value can be added to the output. The relation $V_{out}(t) = kV_{in}(t) + c$ describes the transfer function of a linear system, where k and c are constants. If a sinusoid of frequency f_1 is applied as $V_{in}(t)$, the output spectrum will only have components at f_1 and possibly at the direct current (DC). No other frequency components will be present.
2. A nonlinear system, in contrast, is capable of generating frequencies in the output spectrum that did not exist in the incoming signal. A nonlinear system may, for example, be represented by the following transfer function: $V_{out}(t) = kV_{in}^2(t) + c$. If a sinusoid of frequency f_1 is applied as $V_{in}(t)$, the output spectrum will have components at f_1 and $2f_1$ and possibly at DC (this can be seen by applying basic trigonometric identities). In this example, if two sinusoids are applied at the input, one with frequency f_1 and another with frequency f_2, tones at $2f_1$, $2f_2$, $f_1 - f_2$, and $f_1 + f_2$ will be present at the output. It is clear that frequencies at the output which did not exist at the input have been generated.
3. A constant-envelope modulation is a modulation that has no peaks and valleys observed in the transient waveform of the modulated signal. In other words, the envelope of the carrier does not change with a change in the modulated signal. This class of modulations includes modulations such as FM (frequency modulation) and FSK (frequency shift keying).
4. A nonconstant-envelope modulation is one that has peaks and valleys in the transient waveform of the modulated signal. In this class of modulations, the amplitude of the envelope of the modulated signal can vary as a function of time. This class of modulations includes modulations such as AM (amplitude modulation) and QAM (quadrature amplitude modulation).

In general, all real systems are at least weakly nonlinear (i.e., perfectly linear systems do not exist in the real world). Also, in general, constant-envelope modulations are significantly less spectrally efficient than their nonconstant-envelope counterparts. On the other hand, as will become apparent by the argument below, nonconstant-envelope modulations are much more forgiving of nonlinearities in the system.

All forms of modulations used by the 802.11 standard possess a relatively high spectral efficiency and have a high PAR.

Now that we have some of the basics out of the way, we can discuss the concept of spectral regrowth in more detail. When a modulated signal of bandwidth W (constant envelope or not) is passed through a system with no nonlinearities, a signal with the same W bandwidth is obtained at the output. For example, assuming a "brick-wall"-shaped signal was applied at the input of such a system, the signal at the output would maintain the same brick-wall shape and bandwidth (and would of course have no harmonic outputs).

If the modulated signal is of the constant-envelope type and it is passed through even a nonlinear system, it maintains its bandwidth.[13]

In reality, however, nonlinearities are ever so present. Therefore when a non-constant-envelope signal (especially one with large PARs) is passed through such a system, even- and odd-order harmonics of the input signal are generated, creating harmonic distortion (HD) at the output. These components are typically out of band and can be filtered out. However, if there is more than a single CW tone present at the input, intermodulation (IM) distortion terms will also be generated. Specifically, the odd-ordered intermodulation terms (IM3, IM5, IM7, etc.) are the ones that result in spectral regrowth.

The concept of intermodulation distortion (IMD) will be discussed in more detail in Chapter 3, but for now it suffices to note that IM terms can generate spectral components that fall very close to the frequencies of the input signal. For example, for a two-CW tone input with frequencies f_1 and f_2, the problematic third-order IM products (IM3) fall at $2f_1 - f_2$ and $2f_2 - f_1$. Since f_1 and f_2 are presumably very close in frequency, the $2f_1 - f_2$ and $2f_2 - f_1$ terms will also fall fairly close to the requency range and will be difficult or even impossible to filter out. It is easy to imagine with an input signal with many CW input tones how the intermodulation terms would generate a spectral intermodulation floor within the band of interest and its vicinity.[14] A modulated signal with a nonconstant envelope for the purposes of this discussion can be considered as a multitone CW signal with frequency components across the modulated bandwidth.

Although these IMD terms are present *within* the bandwidth of the modulated signal, their amplitudes are significantly smaller than the amplitude of the desired spectral component of the modulated signal and can therefore typically not be observed by looking at a simple power spectral density plot

[13]Note that if a modulated signal with abrupt phase transitions that has constant envelope is passed through a band-limiting operation (e.g., filtering), it will no longer have a constant envelope and will therefore require somewhat linear amplification in order to avoid the generation of spectral regrowth.

[14]This intermodulation floor is sometimes referred to as the spectrum "grass" or FFT grass. This is because, in looking at the FFT results of such a multitone simulation, the FFT bins of the area immediately outside the bandwidth of interest will show FFT components that have grown above their normal levels.

or a spectrum analyzer output (they do certainly take a hit on EVM, however). These intermodulation components, however, will result in the creation of undesirable spectral components (as observed on a spectrum analyzer or a power spectral density plot) in the immediate vicinity of the modulated signal. These undesirable spectral components generated by the passing of a nonconstant-envelope modulated signal through a nonlinear system is commonly referred to as spectral regrowth.

With the description provided in the previous paragraphs it is easy to understand why large PAR signals are more susceptible to IMD and hence can create large spectral regrowth components as a result of passing through a nonlinear system.

To summarize, when a modulated signal is passed through a nonlinear system, its bandwidth is broadened by odd-order nonlinearities. This is caused by the creation of mixing products between the individual frequency components of the modulated signal spectrum.

In a typical system, spectral regrowth is typically of concern in the transmitter side. Within the transmitter, spectral regrowth is typically caused by the most nonlinear component of the transmit chain. This component is almost always the power amplifier.

Spectral regrowth can cause several problems in a system. In a full-duplex system, for example, a client's transmitter can generate enough out-of-band power due to spectral regrowth to saturate the client's own receiver (note that a transmitter in a typical full-duplex system can be transmitting power levels that are orders of magnitude larger than what the receiver is trying to receive, and therefore it is fairly easy to saturate this receiver). Of course, WLAN systems, in particular 802.11, are not based on a half-duplex operation, and therefore this particular potential impairment caused by spectral regrowth is not an issue for a WLAN system.

A similar issue can be observed when spectral regrowth of one transmitter causes the desensitization of the receiver of a different system in the same communication appliance. For example, the transmitter of a WLAN system operating at 2.4 GHz can desensitize the receiver of a wireless CDMA cellular receiver in a cell phone hand set.

Another problem caused by spectral regrowth is to create interference with adjacent channels. Typically, the spectrum is a very precious commodity. The channels are therefore typically packed together in order to maximize the efficiency of the spectrum usage. Excessive spectral regrowth can therefore cause interference with the adjacent channels (this is the primary concern with spectral regrowth in stand-alone 802.11 applications such as computer laptops). This is the primary reason why the standards and the FCC require conformance to a spectral mask.

Take the case of the 802.11a spectral mask requirements as an example. As discussed earlier, the 802.11a modulated signal is comprised of 52 total subcarriers with modulated data around each of these subcarriers. Also, as stated earlier, such a signal possesses a very high PAR. As this signal is passed through a nonlinear system, these subcarriers can nonlinearly interact with other subcarriers, creating intermodulation components. The data content of each subcarrier can even nonlinearly interact with itself, causing IMD. All of this can cause spectral regrowth. So an ideal brick-wall filtered OFDM-modulated signal (with a bandwidth of 16.25 MHz) could look like what is shown in Figure 1.13 after it goes through a real transmitter with a much wider bandwidth. As shown in Figure 1.13, the 802.11a spectral mask requires that the spectrum of the transmitted signal be more than 20, 28, and 40 dBc below the peak of the modulated signal at offset frequencies of 11, 20, and 30 MHz, respectively, away from the center of the band.

Finally, it is important to note, that the spectral regrowth due to the nonlinearities (of typically the power amplifier) *should* be the limiting factor in achieving the (close-in) spectral mask requirements of a well-designed system which has a fairly high transmit power requirements (such as the 802.11). However, many other factors can cause spectral mask violations. These include insufficient baseband analog or digital filtering of the digitally generated modulated signal, the quantization noise levels of the digital-to-analog converters, and the phase noise of the RF phase-locked loop (PLL).

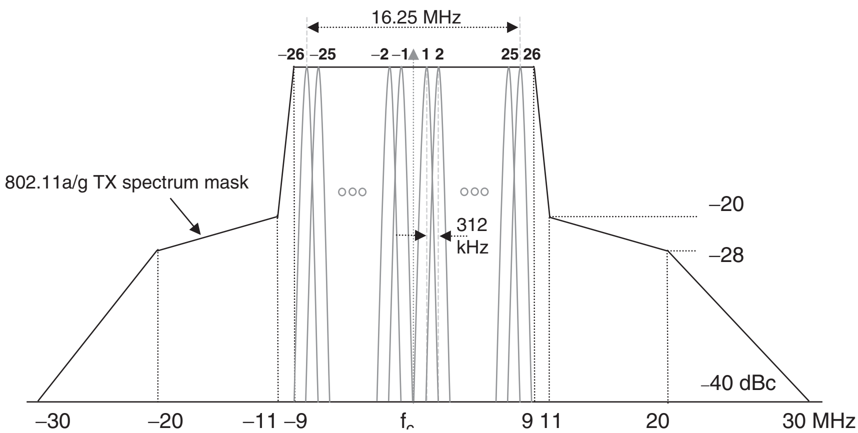


Figure 1.13 IEEE 802.11a/g channel construction from OFDM subcarriers and required transmitter spectral mask.

The subject of spurious emissions is also closely related to the issue of spectral masks. For example, the 802.11 standard specifies that the transmitter local oscillator (LO) feedthrough[15] should be limited to less than 15 dB below the level of the desired transmitted signal power for the 802.11a and 802.11g standards. For the 802.11b standard, the level of LO feedthrough needs to be at least 25dB below the desired signal level. It is important to note that quite often LO feedthrough can exceed the standard specified limits without affecting the transmitter error vector magnitude (see next section). However, excessive LO feedthrough can cause problems on the receiver side, especially for direct-conversion receivers (by generating large-baseband DC offsets).

Another type of spurious emissions is synthesizer-generated and LO-generator-type spurs. For example, the synthesizer can have an excessively large reference feedthrough which can easily violate the spectral mask of the system and/or cause interference with other users that are operating at the adjacent channels. Another example would be the case of the first or second LO of a superheterodyne transmitter leaking to the output and violating an FCC mask for a restricted band. As a final example, quite often in direct conversion systems, the voltage-controlled oscillator (VCO) frequency is selected to be different than the output frequency of the transmitter in order to reduce the pulling effects on the VCO. In such cases the VCO frequency can leak to the output and potentially violate the FCC limit in a restricted band.

The FCC requirements can be divided into two main categories: conductive requirements and radiative requirements. Conductive measurements are performed by connecting a cable to the antenna port of the device under test and measuring the power spectral density of the transmitted signal at various frequencies out of the 802.11 bands. On the other hand, radiative requirements are conducted by measuring the received power at various frequencies at certain defined distances and with specified antennas. The interested reader can refer to FCC documents for more details.

1.11 VECTOR SIGNAL ANALYSIS

A very powerful tool in analyzing a digitally modulated signal is a VSA. A VSA is even more useful in analyzing digitally modulated signals that are

[15]The subject of LO feedthrough will be discussed in more detail in Chapter 4. For now, it is sufficient to note that LO feedthrough is the continuous wave (CW) tone observed in the middle of the transmitted band as observed at the RF output. The two most common sources of LO feedthrough are DC offsets in the baseband analog circuitry of the transmitter and, in the case of direct-conversion transmitters, direct LO coupling to the output of the transmitter.

further mapped by OFDM. This is because in many cases the impact of impairments on an OFDM-modulated signal would be quite different (and often more complicated) than a similar impairment's impact on a "single-carrier"-based (i.e., non-OFDM) modulated signal. Further, typically the impact of various impairments on single-carrier-based modulated signals is more intuitively clear. We will elaborate on this issue in much more detail in Chapter 4.

Vector signal analysis is a tool to relate analog impairments to system requirements. In its simplest form one would look at the constellation diagram of the digitally modulated signal as shown in Figure 1.14. This figure shows a very high quality 64-QAM (an array of 8 × 8 blue constellation points) 802.11a signal, with an EVM of approximately –40 dB.[16] The high quality of the modulation is apparent in the tightly packed dots at the intersection of the circles. As described earlier in discussing the concept of the EVM, these tightly packed dots are an indication that the actual received symbols are quite close to their ideal values. As the signal quality degrades, the dots get larger and "fuzzier" and turn into "balls," as seen in Figure 1.15. At the extreme the various constellation balls start intruding on the adjacent neighbor constellation balls, causing packet errors and degrading the link quality. Note that the black dots in Figure 1.14 in the center left and the center right are associated with the 802.11a pilot tones, which are always transmitted in BPSK format.[17] In particular, these constellation points are quite useful in identifying certain kinds of impairments in the system. More details on this topic will follow later in this book.

So, by looking at a constellation diagram, one can recognize a "good"-quality signal such as that shown in Figure 1.14. But what if the constellation diagram shows a signal with a relatively poor quality (e.g., that of Fig. 1.15)? How would the designer go about finding the problem and remedying it? How would one determine if the problem is due to spurs, phase noise, quadrature imbalance, and so on? This is where, especially in the case of an OFDM-mapped signal, further signal analysis tools and diagrams would be useful.

Figure 1.16 shows an example of an 802.11a-modulated signal (the same signal of the constellation diagram of Fig. 1.14) viewed on a VSA.

[16]Note that the 802.11a/g standard requires an EVM of only –25 dB for a a/g 64-QAM transmitted signal; therefore the constellation shown here is 15 dB better than what the standard requires.

[17]The pilot tones are always transmitted in BPSK format. The pilot tones are used to establish the initial link (establish frequency offsets, set packet gain levels, etc.). It is therefore imperative to obtain the highest amount of immunity to impairments present in the system and in the channel and establish a robust link for the payload to be properly decoded.

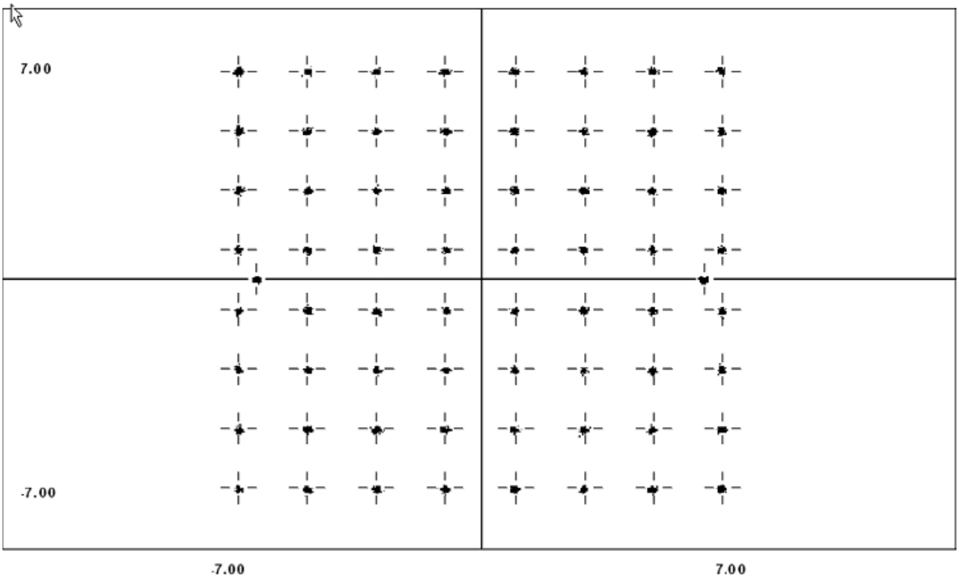


Figure 1.14 Constellation diagram for 802.11a signal with very low (good) EVM (~ –45 dB). This constellation diagram is obtained by feeding the output of a laboratory-class transmitter to a VSA.

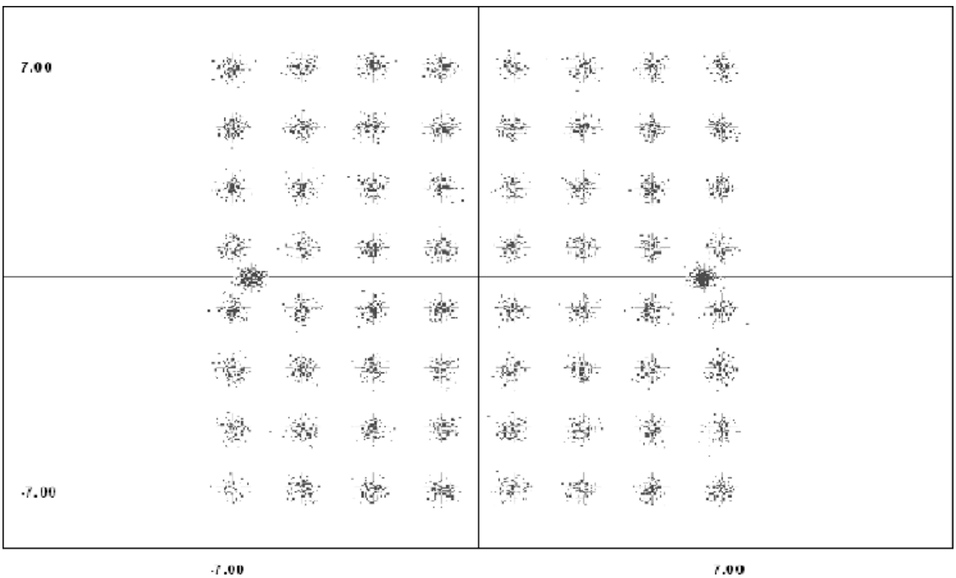


Figure 1.15 Constellation diagram for 802.11a signal when quality of signal is marginally acceptable for 54-Mbps transmission. The EVM of this constellation is ~ –27 dB. Note the "fuzzy" balls that have replaced the well-defined constellation points of Figure 1.14.

In this case, however, the VSA is set up to display the EVM in decibels versus the subcarrier index.[18] In other words, this is an alternative way to view the same signal. As will be shown, this method of observing the signal will shed more insight into certain impairments that may be affecting the system.

The key insight is that a quick glance at Figure 1.16 can provide significant amount of information about the existence (or lack of) analog impairments in the system. Figure 1.16 shows a signal with excellent quality (very low EVM) across all of the subcarriers. Some examples of signals that are impaired in various ways are given in the following figures. The spectrum flatness and group delay associated with this near ideal signal is shown in Figure 1.17.

Figures 1.18a and 1.19b show a signal which has a significant CW spur present at subcarrier 13 (frequency offset of +4 MHz). As can be seen, the EVM for this subcarrier is significantly degraded as compared to other subcarriers. Also note that, by looking at the constellation diagram alone, it would not be possible to pinpoint the reason for the degraded performance. On the other hand, by looking at the plot of the EVM versus the subcarrier, it is quite clear that a large narrowband interference is the source of the degraded EVM. Many sources can contribute to large spur levels. These include harmonics of the crystal oscillator, reference spurs of the PLL, and harmonics of the master clock frequency used in the digital domain of the chip.

Figure 1.20 shows a signal which has fairly significant impairments on the lower index subcarriers.[19] This situation, for example, can arise from excessive flicker noise in the baseband circuitry or as a result of cutting off the low frequency subcarriers by a high pass filter with too high of a corner frequency. The EVM hit can come from the magnitude attenuation due to filtering at the low index subcarriers and/or the group delay variations due to the pole(s) associated with the high pass filter. This situation is not uncommon on 802.11a direct-conversion receivers that use some form of high pass filtering to reject the DC offsets.

Figure 1.21 shows a signal which has significant impairments on the high order subcarriers. This situation is the dual of that described in the previous paragraph but is caused by low pass filters with the poles placed

[18]Recall that the 802.11a signal is constructed of 52 subcarriers ranging from index –26 to +26 with subcarrier index 0 eliminated. The subcarriers are spaced 312.5 KHz apart.

[19]Note that the EVM floor in this case is different that the previous examples due to the fact that a different device under test (DUT) was used for this measurement.

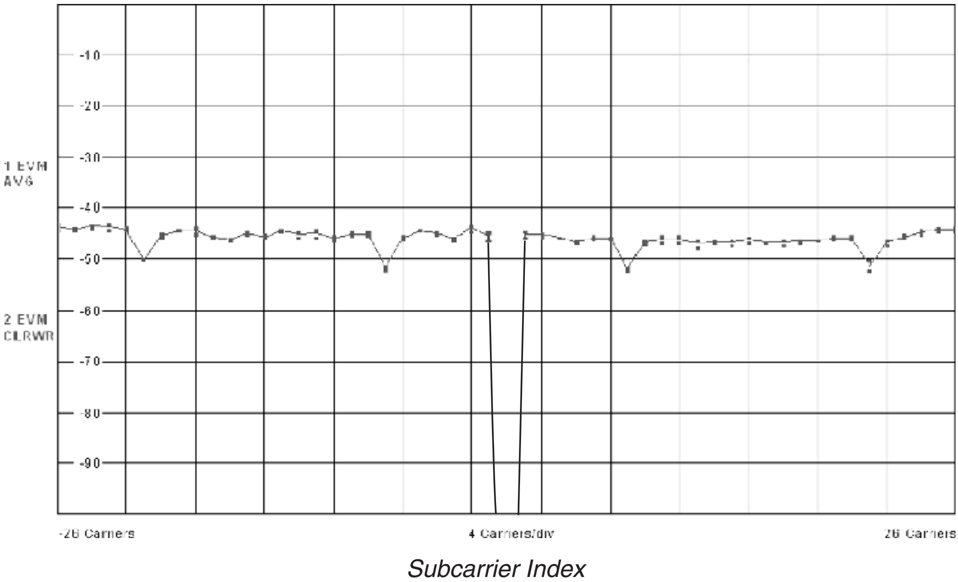


Figure 1.16 EVM versus subcarrier index for constellation plot of Figure 1.14 obtained on a VSA.

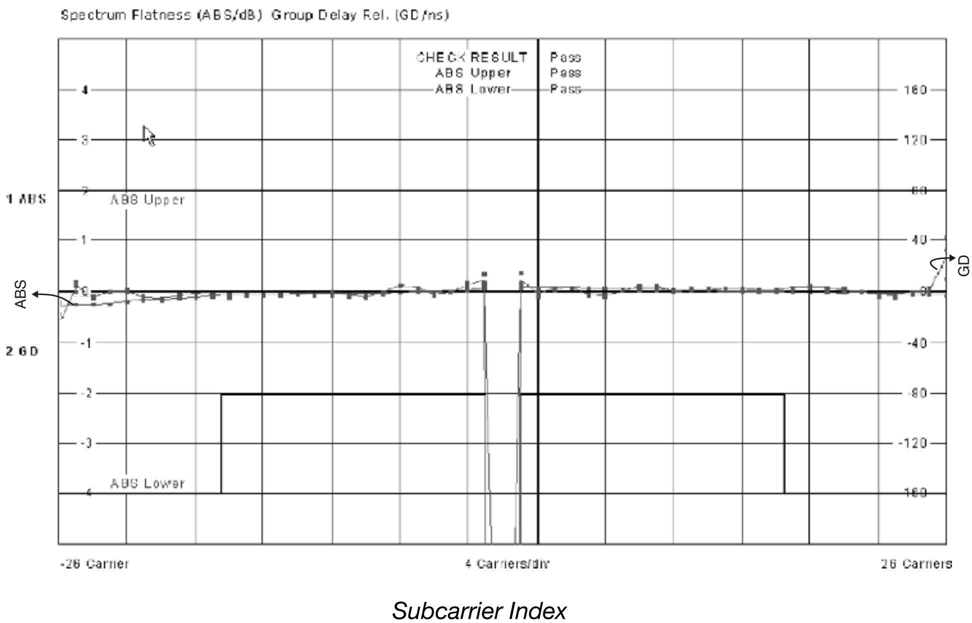


Figure 1.17 Spectrum flatness (ABS, dB) on the left y-axis and group delay variation (GD, ns) on the right y-axis of signal of Figure 1.14.

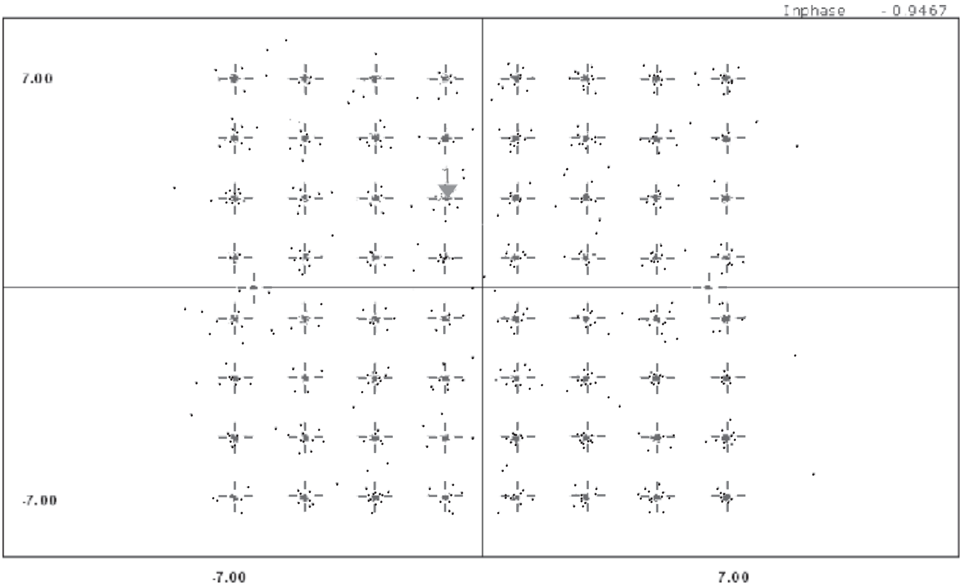
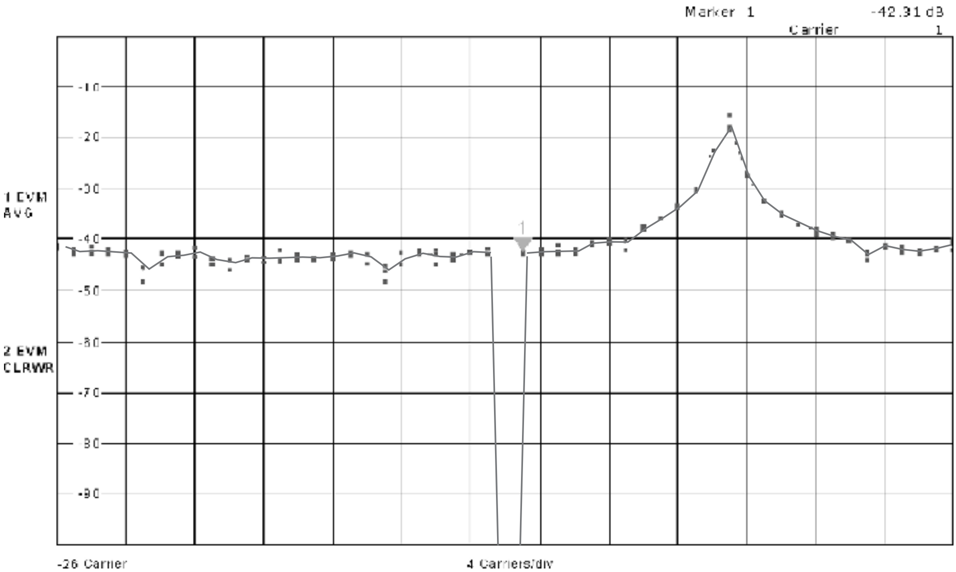


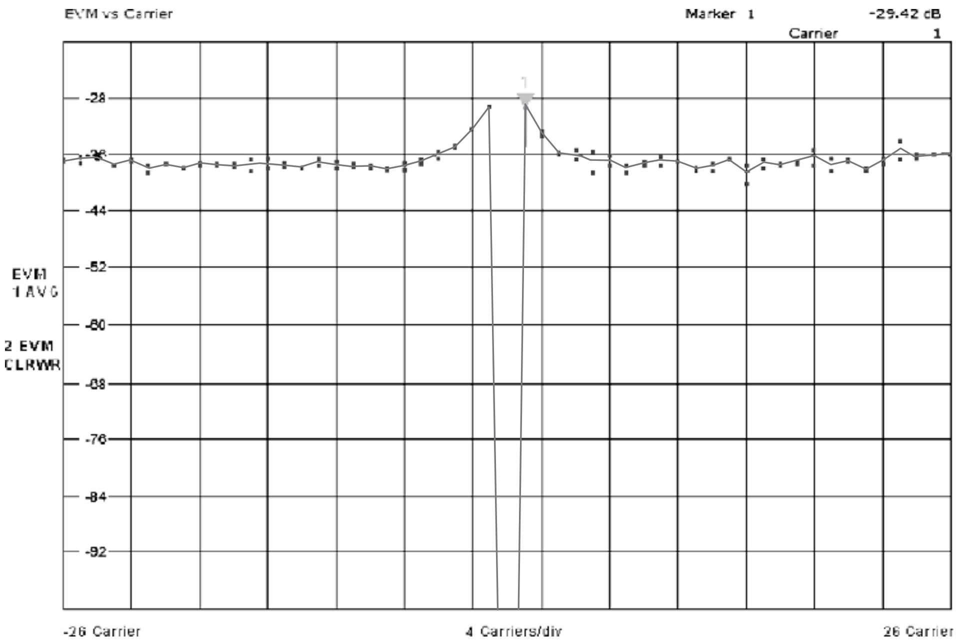
Figure 1.18 Constellation diagram of otherwise excellent quality 802.11a signal with large CW spur. In this case the burst power (of the OFDM signal) is –5 dBm and is centered at 5.24 GHz, and the CW spur has a power level of –35 dBm and is at 5.244 GHz. Measured EVM is –31.6 dB. Notice the out-of-place constellation points on the constellation diagram.

too low. Again, the EVM hit can be due to the magnitude attenuation of the higher order subcarriers (and associated SNR degradation) or due to excessive group delay variations associated with the poles of the high pass filter being placed too low in frequency. A similar situation can arise if a constant group delay variation exists between the *I* and the *Q* channels of the received signal. This results in a subcarrier-dependent quadrature imbalance which, if uncorrected, would result in worse EVM at higher subcarriers. The plot of Figure 1.21 represents such a case where a significant delay difference exists between the *I* and the *Q* channels. As explained further in Chapter 3, this condition may arise due to mismatches in the analog baseband sections of a receiver. A plot of the amplitude variation of this signal over the subcarrier as well as the group delay variation is shown in Figure 1.22. The constellation diagram for this signal is shown in Figure 1.23. It is again clear that, by looking at the constellation diagram alone, it would be quite difficult, if not impossible, to determine the type of impairment impacting the system.



Subcarrier Index

Figure 1.19 EVM versus subcarrier for constellation diagram of Figure 1.18. Note the large degradation in the EVM level at the frequency of the spur. Also notice the "leakage" of the EVM degradation effect on the adjacent subcarriers.



Subcarrier Index

Figure 1.20 EVM versus subcarrier plot of 802.11a signal that shows fairly significant impairments on lower index subcarriers.

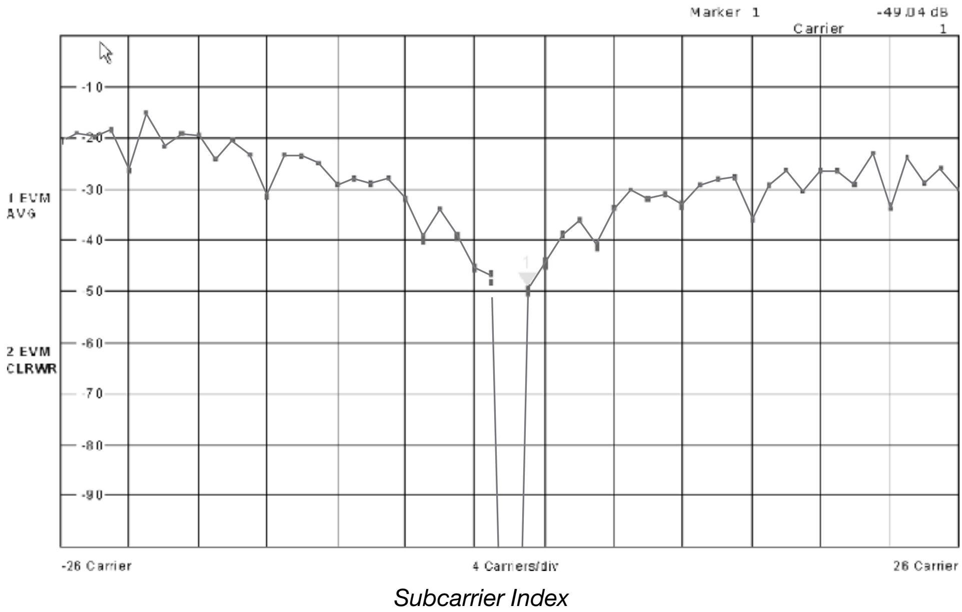


Figure 1.21 Plot of EVM versus subcarrier index for 802.11a signal subject to significant group delay mismatch between *I* and *Q* channels.

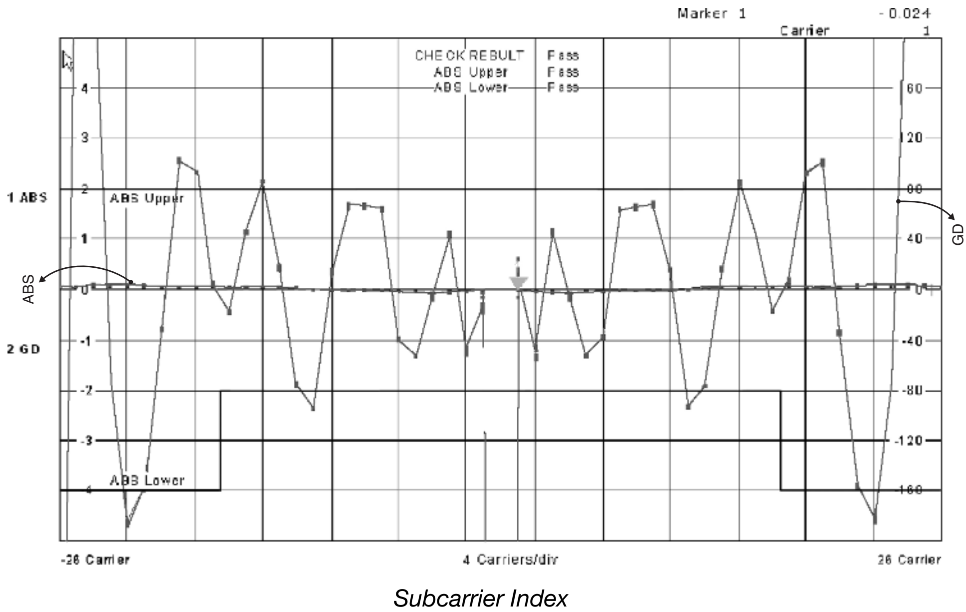


Figure 1.22 Plot of spectrum flatness (ABS, dB, left y-axis), and group delay (GD, ns, right y-axis), versus subcarrier index for 802.11a signal of Figure 1.21.

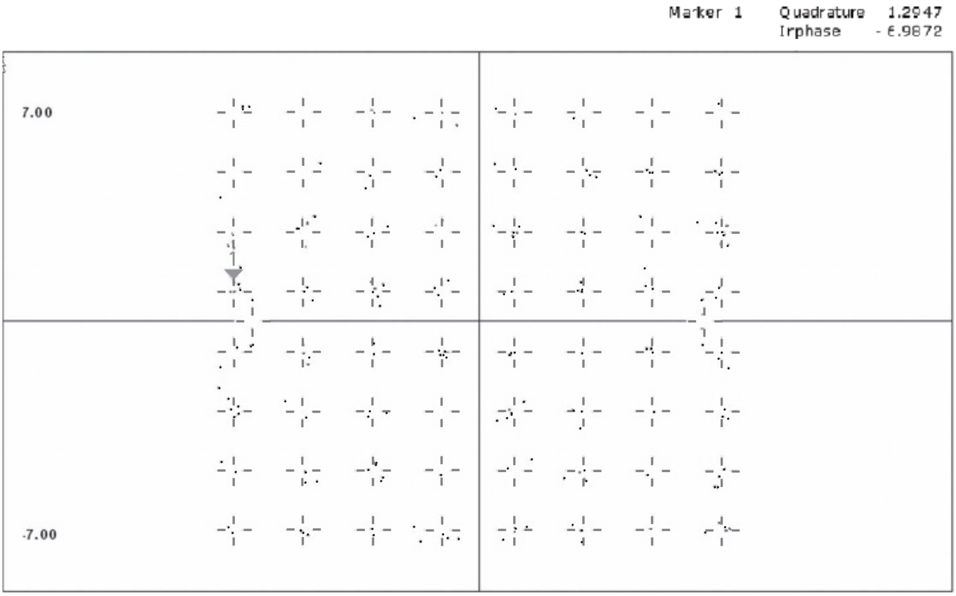


Figure 1.23 Constellation diagram of signal of Figure 1.21.

The examples mentioned above should provide a good perspective on the capabilities of a VSA. Often, the problems could be even further analyzed to very accurately pinpoint the source of the problem. At the very least, this kind of analysis provides the proper hints that the engineers can then use to debug their system. Several more examples of various impairments and methods of debugging them using a VSA will be presented throughout the book.