RFID Design Principles

Second Edition

For a listing of recent titles in the
Artech House Intergrated Microsystems,
turn to the back of this book.

RFID Design Principles

Second Edition

Harvey Lehpamer

ARTECH HOUSE
BOSTON | LONDON
artechhouse.com

Library of Congress Cataloging-in-Publication Data
A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data
A catalog record for this book is available from the British Library.

ISBN-13: 978-1-60807-470-9

Cover design by Vicki Kane

685 Canton Street
Norwood, MA 02062

10 9 8 7 6 5 4 3 2 1

Contents

CHAPTER 3

CHAPTER 4

CHAPTER 1

Introduction

Radio-frequency identification (RFID) is an emerging technology and one of the most rapidly growing segments of today's automatic identification data collection (AIDC) industry. However, this emerging technology is not new; in fact, it is currently being used, in numerous applications throughout the world. It was originally implemented during World War II to identify and authenticate Allied planes, in an identification system known as *Friend or Foe,* and is still being used today for the same purposes.

RFID usage is steadily increasing, and companies across many industries are now looking at RFID to streamline operations, meet regulatory requirements, and prevent the introduction of counterfeit product into the supply chain to protect both consumer safety and company profitability.

Today, industry experts view RFID not as a competition with but as a complement to bar code technology; in many cases, such as tracking pallets, cartons, and cases in a warehouse, both technologies are used. RFID technology, in fact, overcomes certain limitations found in some bar code applications. Because it is not an optical technology like bar coding, no inherent line of sight is required between the reader and the tagged RFID object. In addition, RFID transmits data wirelessly and is a read/write technology, so it can update or change the data encoded in the tag during the tracking cycle.

For an RFID project to be successful, it is necessary to approach any business problems that may arise and any potential RFID solution by using a systems approach. In a design process, we need to look at all the processes, be forward thinking, and think creatively about how to improve each operation. Implementing an RFID-based system is like implementing any new system: RFID systems should be conceived, designed, and implemented using a systematic development process in which end users and specialists work together to design RFID systems based on the analysis of the business requirements of the organization.

One of the greatest obstacles to the wide adoption of any new technology is a standardization process. The purpose of standardization is to define the most efficient platform on which an industry can operate and advance. For example, standardization would address the question of how to ensure that a tag manufactured and installed in one part of the world will be readable and the product properly identified on another side of the globe.

Several organizations are involved in drafting standards for RFID technology, but in looking at the present status, it seems like it will be some time before all of

the details are agreed on. Because RFID standardization is a very dynamic process, this book discusses only the present standards that were current at the time of this writing in a brief section and then provide readers with directions for pursuing further research.

Despite the considerable technical diversity of RFID technology, much of it is largely transparent to prospective users and much can be done to promote awareness of the technology's attributes without going into considerable technical details. However, some basic technical knowledge is necessary for making an informed choice of products to meet particular application needs and to allow informed discussion among users, suppliers, systems integrators, and consultants.

This book introduces prospective users and system designers to the basics of RFID technology, including applications, benefits, technical characteristics, security and privacy, as well as technical and economic challenges of standardization, design, and implementation. As these technical, policy, and cost challenges are slowly mitigated, RFID will likely become the system of choice for global commerce.

Numerous issues beyond the detailed technical and sheer operational capabilities of RFID technology must be considered. Due to the large number of considerations that must be undertaken, only a few intangible and theoretical considerations, such as security, privacy, social, ethical, and future considerations, are presented in this book.

In addition, this book mentions briefly a wide number of new and exciting topics and concepts; some of them, at this point are only of marginal interest to RFID, with the hope of piquing readers' interest in pursuing these new technologies.

This book should become a valuable resource to a wide spectrum of readers interested in exploring this new and exciting topic.

The first edition of this book was prepared and published almost four years ago. Since then, principles of RFID operation have become widely utilized in many branches of science and engineering, with biotech and medical applications being just some of the examples. Although briefly mentioned in the first edition, an update of the book was required in order to cover these new developments in more details.

I believe that the coverage of medical applications in biomedical field will attract additional audience to this book, including biomedical engineers, program managers, and other professionals involved and/or interested in this new and fast-developing field.

Because medical applications create many new ethical challenges, which at some point, will inevitably have to be discussed, a substantial amount of material has been added to the topic of ethical and moral dilemmas of technology.

CHAPTER 2

Short-Range Communications Systems

2.1 Radio-Frequency Spectrum and Propagation

2.1.1 Theory of Electromagnetism and Maxwell's Equations

James Clerk Maxwell (1831–1879) was a Scottish physicist and mathematician whose major discovery of the ether described the vast sea of space that made possible the transmission of light, heat, and radio waves. Maxwell's nineteenth-century discovery of the ether (or its metaphor) led to many advances in electronic communications. His extension of the electromagnetic theory of light led directly to Heinrich Hertz's discovery of radio waves and to the related advances in science and technology of today. Maxwell's mathematical equations, expressing the behavior of electric and magnetic fields and their interrelated nature, were valid, even though his theory of the ether was not.

Maxwell's calculations were scientific observations resulting in his conclusion that the speed of propagation of an electromagnetic field is approximately that of the speed of light (300 million m/s). His proposal that the phenomenon of light is therefore an electromagnetic phenomenon seemed to fit what he and other scientists could observe of the world around them. Maxwell concluded that visible light forms only a small part of the entire spectrum of possible Electromagnetic Radiation (EMR).

Maxwell's book, *The Complete Laws of Electrodynamics,* defines the relationship between the electric field quantities and the magnetic field quantities, and although a detailed explanation of these laws is beyond the scope of this book, they deserve to be mentioned here at least briefly. Maxwell was the first to correctly assemble the complete laws of electrodynamics in his classic text in 1873. Modern electromagnetism theory is based on the four fundamental equations known as *Maxwell's equations*. Before Maxwell, the laws of electrodynamics, including Gauss's law, Ampere's law of magnetostatics, and Faraday's law, were laws of electrostatics and did not predict waves.

These laws correctly described what is known as the *near field* (i.e., the electrostatic field of an electric charge and the magnetostatic field of a current loop). They described the observable impact of electric charges and magnetic fields close to the source but failed to describe the distant impact of these forces. In the static case, when all electric charges are permanently fixed or if they all move at a steady

state, the electric field and the magnetic field are not interconnected. This allows us to study electricity and magnetism as two distinct and separate phenomena. Up until Maxwell challenged conventional wisdom, the separation of electricity and magnetism was the accepted state of the world.

Maxwell corrected Ampere's law of magnetostatics to become Ampere's law as corrected by Maxwell, so that consistency with the law of conservation of charge now occurred. Maxwell added a term indicating that vortices of magnetic fields can be displacement current density (time-varying electric flux density), as well as conduction current density. The resulting corrected equations define the complete laws of electrodynamics and predict electromagnetic waves. Heinrich Rudolf Hertz confirmed experimentally that these waves exist.

Maxwell showed that any conductor (e.g., an antenna) supplied with an alternating current produces a varying magnetic field (*H-field*), which, in turn, produces electric field lines (*E-field*) in space. This is termed the *near field*. In the near field, both the E- and H-fields are relatively static with no propagation. They only vary in strength as the current varies, with the magnetic flux of the H-field coming out from the antenna, and going back in, and the E-field emanating outward.

Maxwell also proved that beyond this quasi-static near field, both the E-fields and H-fields at a certain distance, detached themselves from the conductor and propagated into free space as a combined wave, moving at the speed of light with a constant ratio of $E/H = 120\pi$ or 377 ohms. (Ohms are used because the E-field is measured in volts per meter and the H-field in amperes per meter.) The point in which this happens is called the *far field*.

The electromagnetic waves are generated by the movement of electrical charges such as in a conductive metal object or antenna. Maxwell's first two equations contain constants of proportionality that dictate the strengths of the fields. These are the permeability of the medium μ in henries per meter and the permittivity of the medium ε in farads per meter. They are normally expressed (2.1) relative to the values in free space:

$$\begin{aligned} \mu &= \mu_0 \mu_r \\ \varepsilon &= \varepsilon_0 \varepsilon_r \end{aligned} \tag{2.1}$$

where $\mu_0 = 4\pi \times 10^{-7}$ [H/m] and $\varepsilon_0 = \dfrac{1}{c^2 \mu_0} \approx \dfrac{10^{-9}}{36\pi} \approx 8.854 \times 10^{-12}$ [F/m] are the values in free space, c is a speed of light in free space, and μ_r and ε_r are the relative values (i.e., $\mu_r = \varepsilon_r = 1$ in free space and greater elsewhere). Free space strictly indicates a vacuum, but the same values can be used as good approximations for dry air at typical temperatures and pressures [1].

By applying the Maxwell's equation to magnetic dipoles, we can identify that the distance $r = \dfrac{\lambda}{2\pi}$ is of significance in determining the nature of the fields surrounding the dipoles. Within this distance, we have the near-field region; beyond this distance, the far-field region starts. Maxwell's equations correctly describe both the energy storage field and the energy propagation field.

It was the physicist Ludwig Boltzmann who said that "there is nothing more practical than a good theory," and that is absolutely true in case of Maxwell's

equations. Although highly theoretical and mathematical in their nature, they provided an unprecedented contribution to our understanding of the world around us, as well as very practical side in the form of the development of radio communications.

2.1.2 RF Propagation and Interference

Radio waves and microwaves are forms of electromagnetic energy we can collectively describe by the term radio frequency (RF). RF emissions and associated phenomena can be discussed in terms of energy, radiation, or fields. We can define *radiation* as the propagation of energy through space in the form of waves or particles. Electromagnetic radiation can best be described as waves of electric and magnetic energy moving together (i.e., radiating) through space as illustrated in Figure 2.1.

For example, the alternating movement of charge (i.e., the current) in an antenna used by a radio or television broadcast station or in a cellular base station antenna generates electromagnetic waves. These waves that radiate away from the transmitting antenna are then intercepted by a receiving antenna such as a rooftop TV antenna, car radio antenna, or an antenna integrated into a handheld device such as a cellular phone or another wireless communications device.

Continuous harmonic waves are typically sinusoidal in nature; thus, they are characterized by frequency, amplitude, and phase. They are also characterized by their three-dimensional shape. The energy radiated by any antenna is contained in a transverse electromagnetic wave (TEM) that is comprised of an electric field and a magnetic field. These fields are always orthogonal to one another and orthogonal to the direction of propagation.

The *power flow density* of the EM wave, vector $\vec{P} = E \times H$ in the direction of propagation, and measured in watts per square meter, is the *propagation* vector

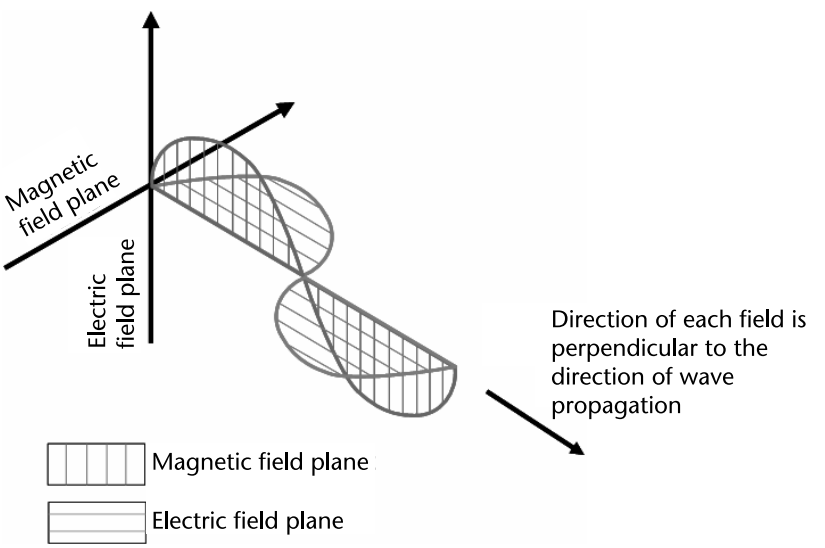


Figure 2.1 Electromagnetic wave.

or *Poynting vector,* named after the English physicist John Henry Poynting, who introduced it in 1884.

The term *electromagnetic field* is used to indicate the presence of electromagnetic energy at a given location. The *RF field* can be described in terms of the electric and/or magnetic field strength at that location. Like any wave-related phenomenon, electromagnetic energy can be characterized by a *wavelength* and/or *frequency*.

Frequency, f, is defined as a number of cycles, or periods, per unit of time and is measured in hertz (Hz). The wavelength, λ, of a sinusoidal wave is the spatial period of the wave, that is, the distance over which the wave's shape repeats. In other words, the wavelength is the distance by which the phase of the sinusoidal wave changes by 2π radians, so we can say: $k\lambda = 2\pi$, $\omega = kc$, and $f = \omega/2\pi$. *Angular frequency, ω,* is a rotational unit of angular frequency $2\pi f$ with units in radians per second (rad/s). From here we have:

$$\lambda = \frac{2\pi}{k} = \frac{2\pi c}{\omega} - \frac{c}{f} \tag{2.2}$$

Electromagnetic waves travel through space at the speed of light, and the wavelength and frequency of an electromagnetic wave are inversely related by a simple mathematical formula (2.2) connecting wavelength, speed of light, and frequency.

Variation, or modulation, of the properties of the wave (amplitude, frequency, or phase) then allows information to be carried in the wave between its source (transmitter) and destination (receiver), which is the goal of every type of wireless communications.

Because the speed of light ($c = 3 \times 10^8$ m/s or 1.86×10^5 miles/s) in a given medium or vacuum does not change, we can see from (2.2) that the high-frequency electromagnetic waves will have short wavelengths and the low-frequency waves will have long wavelengths.

The *electromagnetic spectrum*[1] includes all the various forms of electromagnetic energy from extremely low-frequency (ELF) energy, with very long wavelengths, to X-rays and gamma rays, which have very high frequencies and correspondingly short wavelengths (Figure 2.2). Between these extremes are radio waves, microwaves, infrared radiation, visible light, and ultraviolet radiation, in that order. The RF part of the electromagnetic spectrum is generally defined as that part of the spectrum where electromagnetic waves have frequencies in the range of about 3 kHz to 300 GHz.

Different frequencies have different propagation characteristics. All frequencies are attenuated and reflected by materials to a greater or lesser degree, with the higher frequencies being more greatly attenuated than the lower frequencies. Low frequencies, such as the 125-kHz frequency, are attenuated very little as they propagate through materials, allowing them to have significant signal-penetration capabilities through all materials including metal.

1. The word *spectrum* literally means a range of values or a set of related quantities.

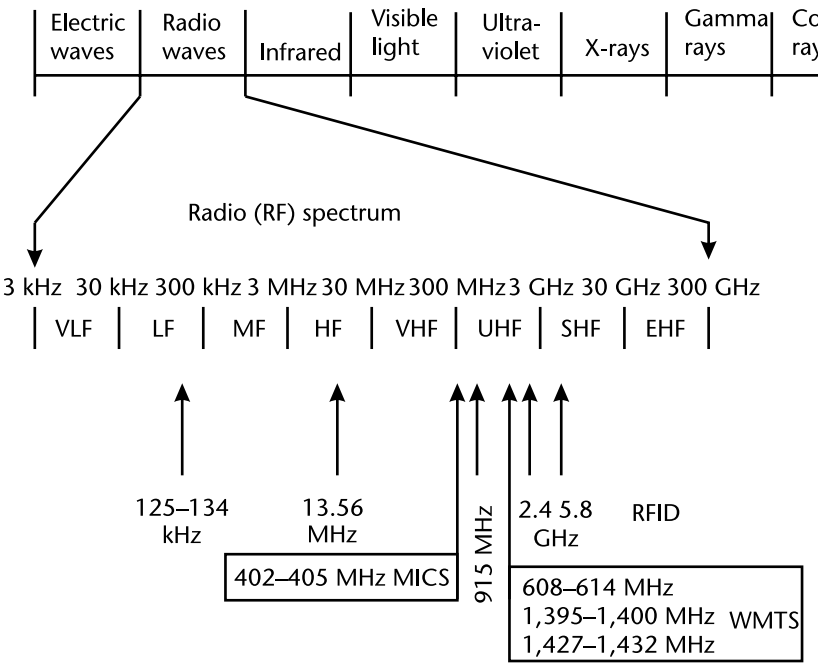


Figure 2.2 Electromagnetic spectrum.

When radiated and used in the far field, these frequencies can also have a significant communication range. For example, we listen to AM radio stations (typically operating between 580 and 1,700 kHz) that were being broadcast hundreds of miles away from us, while FM radio stations, operating typically between 88 and 108 MHz, have a range of about 20 miles.

For the regulations limiting RF emissions, the U.S. Federal Communications Commission (FCC) distinguishes between intentional, unintentional, and incidental radiators:

- *Intentional radiators* are devices that intentionally emit RF energy, such as transmitters.
- *Unintentional radiators* are devices that intentionally generate RF energy for use only within the device or a cable system but not for the purpose of radiation. Examples for unintentional radiators are computer motherboards and receivers with local oscillators.
- *Incidental radiators* are devices that are not designed to generate RF energy at all, but for which RF radiation may occur as an unwanted side effect. Examples of incidental radiators are dc motors and mechanical switches.

As with all waves, electromagnetic waves interact with one another whenever they intersect at a point in space. Depending on the phase, amplitude, and polarization, intersecting waves may either constructively interfere or destructively interfere. This is one of the basic properties of linear waves. The observed wave at a point of intersection is the addition of all of the waves at that point.

Constructive interference increases the amplitude of the detectable wave at that point while *destructive interference* decreases the amplitude of the detectable wave. Both of these cases could be harmful for the communications channel and the quality of the resulting signal.

When two waves that have traversed different length paths intersect at a point, they will be out of phase with one another. The phase difference is due to differences in the time required to traverse the different paths. Most phase differences cause destructive interference and may cause the observed wave at a point to appear to have a different frequency than what was originally transmitted.

Fundamental physics teaches us that at every boundary between two materials, electromagnetic waves incident upon that boundary will be both transmitted from one material to the other and reflected back into the material in which they are traveling. Conducting materials, such as metals, act similar to perfect reflectors for UHF radiation. Materials such as glass, concrete, and cardboard are effectively RF transparent for waves that are incident upon them with an angle of incidence of 90°, but they become less transparent as the angle of incidence becomes more oblique.

Some materials, such as water, act as both good reflectors of electromagnetic waves and good attenuators, or absorbers, of electromagnetic energy. The partial reflection of a wave results in the energy of the wave being separated to traverse multiple paths. The result is that a partial reflection attenuates the partially transmitted wave by the amount of energy reflected at the boundary.

By passing through several materials and being reflected by several more, an electromagnetic wave traverses a path through the environment. In addition to attenuating the wave as it travels through the environment, the environment may impact the polarization of the wave. Two long parallel metal strips separated by a few inches, for example, will filter the UHF waves that are incident upon them by allowing waves that are polarized parallel to the metal strips to pass through the space between the strips, while electromagnetic waves polarized perpendicular to the metal strips will be reflected.

Electromagnetic waves are linear, meaning that the wave experienced at a point in space and time is the sum of the waves that intersect at that point. Because of reflections, attenuation, and different path lengths (multipath) caused by objects in the environment, the waves that arrive at a point in space may have amplitude that sums to zero or nearly zero. The position of nulls may be changed or the nulls may be eliminated by changing the position of the objects in the environment or changing the frequency being radiated by the antenna.

When the environment is static, *standing waves* may result. The common occurrence is when the two waves intersect each other exactly half a wavelength ($\lambda/2$) out of phase and completely cancel out the signals.

2.1.3 Basic Antenna Parameters

Antennas are a very important component of communication systems. By definition, a *transmitting antenna* is a device used to transform an RF signal, traveling on a conductor, into an electromagnetic wave in free space. A *receiving antenna* performs an inverse function, that is, converts electromagnetic field into an electrical signal that can be processed in the receiver.

Antennas demonstrate a property known as *reciprocity*, which means that an antenna will maintain the same characteristics regardless if it is transmitting or receiving.

Most antennas are passive resonant devices, which operate efficiently over a relatively narrow frequency band. An antenna must be tuned to the same frequency band of the radio system to which it is connected; otherwise, the reception and the transmission will be impaired. When a signal is fed into an antenna, the antenna will emit radiation distributed in space in a certain way. A graphical representation of the relative angular distribution of the radiated power in space is called a *radiation pattern*.

2.1.3.1 Antenna Polarization and Polarization Diversity

Polarization is a physical phenomenon of radio signal propagation and refers to the orientation of the electric field vector in the radiated wave. If the vector appears to rotate with time, then the wave is *elliptically polarized*. The ellipse so described may vary in ellipticity from a circle to a straight line, or from circular to linear polarization. So, in the general sense, all polarization may be considered to be elliptical (Figure 2.3).

For linear polarization, the vector remains in one plane as the wave propagates through space. Linear polarization has two subcategories: *vertical polarization* or *horizontal polarization*, and right- or left-handed for circular cases.

The term used to describe the relationship between the magnitudes of the two linearly polarized electric field components in a circularly polarized wave is *axial ratio*. In a pure circularly polarized wave both electric field components have equal magnitude and the axial ratio, *AR*, is 1 or 0 dB (10 logAR). In a pure linearly polarized wave, the axial ratio is equal to infinity.

Generally speaking, in most cases two antennas that form a link with each other must be set for the same polarization; however, intentional exceptions are sometimes made, and we will discuss them here as well. When transmit and receive antennas are both linearly polarized, physical antenna misalignment will result in

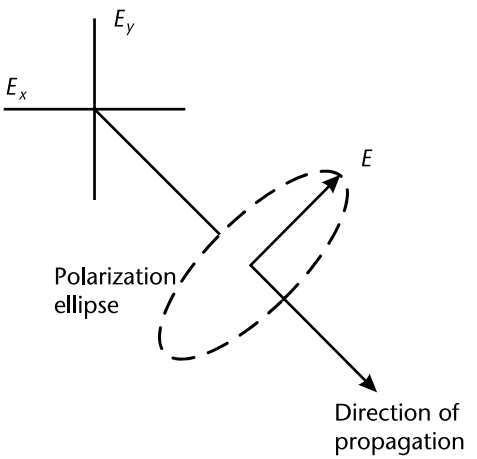


Figure 2.3 Elliptical polarization.

a *polarization mismatch loss* (PML) that can be approximated using the following formula (2.3):

$$PML = 20\log(\cos\theta)\ [\text{dB}] \tag{2.3}$$

where θ is the misalignment angle between the two antennas. For 15° we have a loss of 0.3 dB, for 30° we have 1.25 dB, for 45° we have 3 dB, and for 90° (orthogonal) we, ideally, have an infinite loss, which means no communication at all.

One of the common misconceptions regarding polarization relates to the circumstance where one antenna in a transmit-to-receive circuit is circularly polarized and the other is linearly polarized [2]. It is generally assumed that a 3-dB system loss will result because of the polarization difference between the two antennas. In fact, the polarizations mismatch loss between these two antennas will only be 3 dB when the circularly polarized antenna has an axial ratio of 0 dB. The actual mismatch loss between a circularly polarized antenna and a linearly polarized antenna will vary depending on the axial ratio of the circularly polarized antenna.

When the axial ratio of the circularly polarized antenna is greater than 0 dB, this indicates that one of the two linearly polarized components will respond to a linearly polarized signal more so than the other component will. When a linearly polarized wave is aligned with the circularly polarized linear component with the larger magnitude, the polarization mismatch loss will be less than 3 dB. When a linearly polarized wave is aligned with the circularly polarized linear component with the smaller magnitude, the polarization mismatch loss will be greater than 3 dB.

Assuming a 1-dB maximum axial ratio over the main beam, the signal loss, due to polarization mismatch, will be between 2.5 and 3.5 dB. We will see later that in RFID systems 3 dB will be used as an approximation and an average between the minimum and maximum polarization loss for a given axial ratio.

Multipath signals in RF systems arrive at the receiver's antenna via the reflection of the direct signal from nearby objects. If the reflecting objects are oriented such that they are not aligned with the polarization of the incident wave, the reflected wave will experience a polarization shift. The resultant or total signal available to the receiver at either end of the communications link will be the vector summation of the direct signal and all of the multipath signals. In many instances, a number of signals arriving at the receive site will not be aligned with the assumed standard polarization of the system antenna. As the receive antenna rotates from vertical to horizontal, it simply intercepts or receives energy from these multiple signals.

To improve or extend system performance, some system designers use receive polarization diversity techniques in an effort to enhance signal reception. In these systems, a circularly polarized or dual linearly polarized antenna will be used at the receive site to take advantage of the fact that many linearly polarized multipath signals, with different orientation, exist at the receiving site. These circular and dual polarized antennas theoretically have a better chance of receiving more total signal than a single, linearly polarized antenna.

Polarization of the waves in RFID systems becomes very important in tag antenna designs and deployments; antennas may be designed such that they efficiently capture and communicate with energy in one or a few different polarizations. If a reader antenna is linearly polarized and the tag antenna is linearly polarized, then the tag and the reader may communicate only when both antennas are oriented

in the same linear direction. Circularly polarized antennas reduce the orientation requirements, but do not completely eliminate the orientation dependence for optimal performance.

It is difficult to predict the orientation of the electric field in the *near-field region* (i.e., very close to the antenna), because the transmitting antenna cannot be considered as a point source in this region. In the far-field region, the antenna becomes a point source, the electric and magnetic components of the field become orthogonal to the direction of propagation, and their polarization characteristics do not vary with distance.

Most RF and microwave systems operate in far-field region, with the exception of passive point-to-point microwave repeaters. Some RFID systems at lower frequencies operate in the near field and, as we will see later, use coupling methods instead of usual electromagnetic wave propagation.

2.1.3.2 Impedance Matching and Return Loss

Input Impedance

For an efficient transfer of energy, the impedance of the radio, of the antenna, and of the transmission cable connecting them must be the same. Transceivers and their transmission lines are typically designed for 50Ω impedance. If the antenna has impedance different from 50Ω, then there is a mismatch and an impedance matching circuit is required. Of course, other impedances are also common.

Standing Waves and Voltage Standing Wave Ratio (VSWR)

In order for the antenna to operate efficiently, maximum transfer of power must take place between the transmitter and the antenna. Maximum power transfer can take place only when the impedance of the antenna is matched to that of the transmitter. According to the maximum power transfer theorem, maximum power can be transferred only if the impedance of the transmitter is a complex conjugate of the impedance of the antenna under consideration and vice versa [3].

Transmitter and load are usually connected through some type of transmission line; transmitter and the load have to be matched with the transmission line for the optimal transfer of power. If the condition for matching is not satisfied, then some of the power maybe reflected back, leading to the creation of standing waves, which can be characterized by a parameter called the *voltage standing wave ratio* (VSWR), written, for example, as 1.2:1.

Matching the load Z_L to the transmission line is very important, and although for most applications in radio communications, the VSWR is required to be smaller than 2, in some applications (RFID, for example) of interest are also the following special cases (2.4):

$$\begin{aligned} &Z_L = Z_0 \quad VSWR = 1 \text{ Matched Termination} \\ &Z_L = \infty \quad VSWR = \infty \text{ Open Circuit} \\ &Z_L = 0 \quad VSWR = \infty \text{ Short Circuit} \end{aligned} \tag{2.4}$$

Another way to think about VSWR is it is the amount of input power needed to get an equivalent of unity power out. In transmission line theory the accepted definition is that standing waves are created by superposition of two waves traveling in opposite direction.

The reflection coefficient at the load, Γ_L, can be calculated as follows:

$$\Gamma_L = \frac{Z_L - Z_0}{Z_L + Z_0} \tag{2.5}$$

and VSWR can be calculated according to:

$$VSWR = \frac{1 + |\Gamma_L|}{1 - |\Gamma_L|} \tag{2.6}$$

Note that the reflection coefficient defined above is the reflection coefficient at the load. It is possible to define a generalized reflection coefficient along the line and see that the magnitude does not change as we move along the line; only the phase changes.

Because the standing wave ratio is not always calculated from the voltages, the V is sometimes dropped from VSWR, and the term is referred to as standing wave ratio (SWR). The terms VSWR and SWR can be used interchangeably.

Return Loss

The return loss (RL) is another way of expressing impedance mismatch. It is a logarithmic ratio measured in decibels that compares the power reflected by the antenna to the power that is fed into the antenna from the transmission line. For a matched load, VSWR = 1. The relationship between VSWR and return loss is the shown in the (2.7):

$$RL = -20 \log \frac{VSWR - 1}{VSWR + 1} \text{[dB]} \tag{2.7}$$

The reflection coefficient, the return loss, and VSWR are all quantities used for describing impedance matching.

2.1.3.3 Bandwidth

The bandwidth of an antenna refers to the range of frequencies over which the antenna can operate correctly. The antenna's bandwidth is the number of hertz for which the antenna will exhibit a VSWR of less than 2:1.

The bandwidth (BW) can also be described in terms of percentage (2.8) of the center frequency of the band.

$$BW = 100 \cdot \frac{f_H - f_L}{f_c} \ [\%] \tag{2.8}$$

where f_H is the highest frequency in the band, f_L is the lowest frequency in the band, and f_C is the center frequency in the band. In this way, bandwidth is constant relative to frequency. If bandwidth was expressed in absolute units of frequency, it would be different depending upon the center frequency. Different types of antennas have various bandwidth limitations.

2.1.3.4 Directivity and Gain

Directivity is the ability of an antenna to focus energy in a particular direction when transmitting or to receive energy better from a particular direction when receiving. In a static situation, it is possible to use the antenna directivity to concentrate the radiation beam in the given direction.

Gain is not a quantity which can be defined in terms of a physical quantity such as the watt or the ohm, but it is a dimensionless ratio. Gain is given in reference to a standard antenna. The two most common reference antennas are the *isotropic antenna* and the *resonant half-wave dipole antenna.*

The isotropic antenna radiates equally well in all directions. Real isotropic antennas do not exist, but they provide useful and simple theoretical antenna patterns with which to compare real antennas.

Any real antenna will radiate more energy in some directions than in others. Since it cannot create energy (antenna is a passive device), the total power radiated is the same as an isotropic antenna, so in other directions it must radiate less energy. The gain of an antenna in a given direction is the amount of energy radiated in that direction compared to the energy that an isotropic antenna would radiate in the same direction when driven with the same input power. Usually we are only interested in the maximum gain, which is the gain in the direction in which the antenna is radiating most of the power.

An antenna gain of 3 dB compared to an isotropic antenna would be written as 3 dBi. The resonant half-wave dipole can be a useful standard for comparing to other antennas at one frequency or over a very narrow band of frequencies. Comparing the dipole to an antenna over a range of frequencies requires a number of dipoles of different lengths.

An antenna gain of 3 dB compared to a dipole antenna would be written as 3 dBd. Compared to the gain of an ideal isotropic antenna of 1 (0 dB), the gain of a half-wave dipole antenna is 1.64 (2.15 dB).

The relationship between the two units, *dBi* and *dBd,* is therefore expressed as follows:

$$G_{dBi} = G_{dBd} + 2.15 \tag{2.9}$$

2.1.3.5 Radiation Pattern

The radiation pattern of an antenna is a plot of the far-field radiation from the antenna. An antenna's radiation pattern is usually plotted by normalizing the radiation

intensity by its maximum value and plotting the result. To compare radiation patterns, a few different parameters, described next, can be used.

Beamwidth

An antenna's *beamwidth* is usually understood to mean the *half-power beamwidth*. The peak radiation intensity is found and then the points on either side of the peak, which represent half the power of the peak intensity, are located. The angular distance between the half-power points is defined as the beamwidth.

Half the power expressed in decibels is –3 dB, so the half-power beamwidth is sometimes referred to as the 3-dB beamwidth. Both horizontal and vertical beamwidths are usually considered.

Assuming that most of the radiated power is not divided into sidelobes, then the directive gain is inversely proportional to the beamwidth (i.e., as the beamwidth decreases, the directive gain increases).

Sidelobes

No antenna is able to perfectly radiate all the energy in one preferred direction. Some is inevitably radiated in other directions. The peaks are referred to as sidelobes, commonly specified in decibels down from the main lobe.

Nulls

In an antenna radiation pattern, a null is a zone in which the effective radiated power is at a minimum. A null often has a narrow directivity angle compared to that of the main beam. Thus, the null is useful for several purposes, such as suppression of interfering signals in a given direction.

Front-to-Back Ratio

It is useful to know the front-to-back ratio, which is the ratio of the maximum directivity of an antenna to its directivity in the rearward direction. For example, when the principal plane pattern is plotted on a relative decibel scale, the front-to-back ratio is the difference in decibels between the level of the maximum radiation and the level of radiation in an opposite direction (180°).

2.1.3.6 Antenna Modeling

The equations extracted from both the receiving and transmitting antenna models allow us to describe the behavior of a single bidirectional antenna. In both cases, antenna systems can be represented using Thevenin's equivalent circuits.

In the transmit mode, an antenna system can be represented by a Thevenin circuit equivalent, as shown in Figure 2.4. In this figure, the antenna is represented by impedance Z_A given by (2.10):

$$Z_A = R_L + R_r + jX_A \tag{2.10}$$

The radiating element is symbolized by a radiation resistance R_r and an imaginary part X_A. In this case, R_L represents both the conduction and the dielectric losses of the antenna. The source to which the antenna is connected is represented

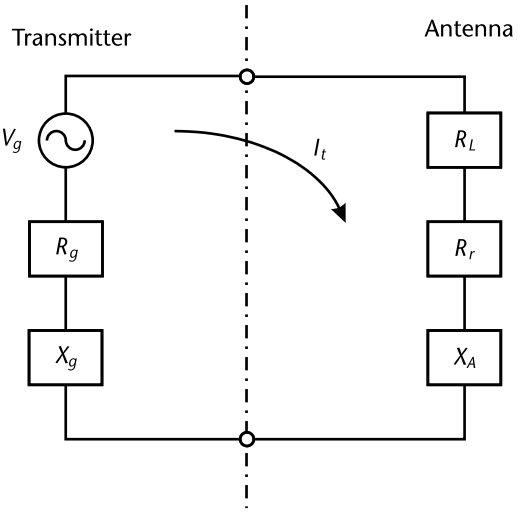


Figure 2.4 Antenna in a transmitting mode.

by an ideal generator V_g having its own internal complex impedance consisting of R_g and jX_g.

The energy transferred to the antenna and its environs by the reactive power flow is stored mostly in the reactive near field. The energy transferred to the antenna by the resistive power flow either heats up the antenna structure (or objects in the near-field region of the antenna) or else is radiated; most often, combination of both of these phenomena occur.

Thus, we can break down the resistive part, R, of the driving point impedance into the sum of a *loss resistance,* R_L, which gets hot, and a *radiation resistance,* R_r. Because the purpose of an antenna is to radiate energy, it is therefore the radiation resistance R_r that is most interesting.

The radiation power delivered by the antenna is the power collected by resistance R_r and it is given by:

$$P_r = \frac{1}{2}|I_t|^2 R_r \tag{2.11}$$

where I_t is the current through R_r.

The factor 2 arises because the average value of the square of a unit sinusoidal signal over one cycle is just 1/2. In determining the average power dissipated, we intrinsically assume that the average is taken over a whole number of cycles of the ac signal.

The magnitude of I_t can be calculated using the following expression:

$$|I_t| = \frac{|V_g|}{\sqrt{(R_r + R_L + R_g)^2 + (X_A + X_g)^2}} \tag{2.12}$$

Maximum power is delivered to the antenna under conjugate matching, meaning:

$$\begin{aligned} R_L + R_r &= R_g \\ X_A &= -X_g \end{aligned} \tag{2.13}$$

The maximum power at maximum efficiency will be transferred when the impedances are complex conjugate matched throughout the power chain, from the transmitter output, through the transmission line (which may be a balanced pair, a coaxial cable, or a waveguide), to the antenna system, which consists of an impedance matching device and the radiating element(s). For maximum power, $Z_{load} = Z_{source}^*$ (where * indicates the complex conjugate).

If we consider a lossless antenna ($R_L = 0$), the ideal amount of power collected by R_r is calculated by combining (2.12) and (2.13) and given by the resulting equation:

$$P_r = \frac{|V_g|^2}{8R_r} \tag{2.14}$$

It follows that the maximum power transferred to the load will be one-half the total power, while the other half is being lost. If the generator and antenna are mismatched, the transferred power will be even lower.

The case of a receiving antenna is very similar. Power collected induces a voltage V_r on the receiving antenna, which is analogous to V_g of the transmitting antenna model. The Thevenin equivalent circuit of the receiving antenna and its load is shown in Figure 2.5. The load to which the receiving antenna is connected (receiver and transmission line) is represented by the receiver's input complex impedance,

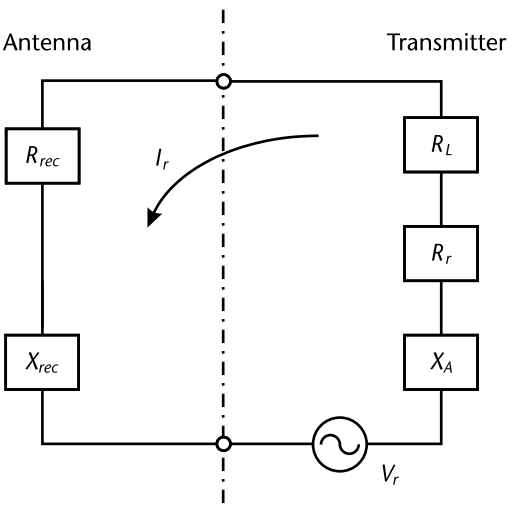


Figure 2.5 Antenna in a receiving mode.

R_{rec} and X_{rec}. As previously shown for the transmitting model, we can derive the functional equations from this receiving antenna model as well.

Typically, the gain in any given direction and the impedance at a given frequency are the same when the antenna is used in transmission or in reception, due to reciprocity.

2.1.4 Range of a Radio Communications System

For any given radio transmitter and receiver, there is a maximum distance (or range) over which communications can work reliably. If the separation between transmitter and receiver in increased beyond this distance, the receiver will no longer be able to correctly recreate the information being transmitted.

It is prudent to operate a radio communications system with a certain margin, that is, not right at the maximum range, because the exact range is likely to vary from moment to moment, and this must be accommodated to achieve reliable performance. The exact range will be affected by a number of factors, but in simple terms there are four:

1. The power contained in the wave transmitted;
2. The sensitivity of the receiving equipment;
3. The environment through which the wave travels;
4. The presence of interference.

These are largely self-explanatory, but of particular importance is the relationship between power and distance. As the radio wave travels away from the transmitting antenna, it disperses in all directions. This means that every time the distance from the transmitter doubles, the proportion of the original wave that is available for reception is quartered (the so-called *inverse square relationship*).

Also note the effect that the environment has on radio communication; when electromagnetic radiation passes through materials, it may be absorbed to a certain extent, depending on the properties of the material and the type of radiation. This absorption results in a reduction of the strength of the radiation, a process known as *attenuation*. This attenuation increases with the thickness of the material.

Visible light is absorbed relatively easily, whereas radio waves are more likely to pass through materials (especially gases in the atmosphere, such as nitrogen and oxygen, and also paper, cardboard, and certain plastics) with only little attenuation of the radiation. Other materials (metal and liquids, for example) have a stronger attenuating effect, although such attenuation varies depending on the frequency of the wave.

In addition to attenuation by absorption, certain frequencies of radio waves are also susceptible to *multipath fading*. This occurs when waves are reflected by objects in the environment, and the reflections interfere with the original waveform, making it much more difficult for the radio wave receiver to determine the original wave. In the worst case there are positions (called nulls) in which reception is not possible even though the transmitter and receiver are relatively close to each other. Similarly, if there are any other radio waves being transmitted on a similar frequency, they will cause interference in the same manner.

2.2 Spread-Spectrum Communications Systems

Human society is entering an era of *ubiquitous computing*,[2] when networks are seamlessly interconnected and information is always accessible when needed. This is the computing environment that all objects and target in physical environment become intelligent and exchange the information by linking each other.

The practical implementation of ubiquitous services requires three levels of connectivity: wide area networks (WAN), typically via the Internet, to remotely connect all types of severs and terminals; local area networks (LAN), typically via Ethernet or Wi-Fi connectivity among all the information and communication appliances in offices and homes; and human area networks (HAN) for connectivity to personal information, media and communication appliances within the much smaller sphere of ordinary daily activities (i.e., the last 1 meter or 3 feet). Short-range wireless technology already plays a key role in scenarios where people are connected anywhere and anytime by different types of communication links.

Short-range radio communications Devices (SRD) have been used for many years to provide low-cost services such as short-range telemetry, voice and video communications, radio LANs, and security systems (Figure 2.6). They are defined by the ITU-R as "radio transmitters which have a low capability of causing interference to other radio equipment." In recent years, there has been rapid growth in the use of SRDs, driven by new technologies and international coordination on specifications.

Designers of SRD wireless systems need to exercise great care in selecting the radio's communication frequency. In most cases, the choice is limited to those portions of the spectrum that allow license-exempt operation given that certain specifications and conditions on usage are met. The international coordination has included attempts to provide common spectrum allocations in the major trading

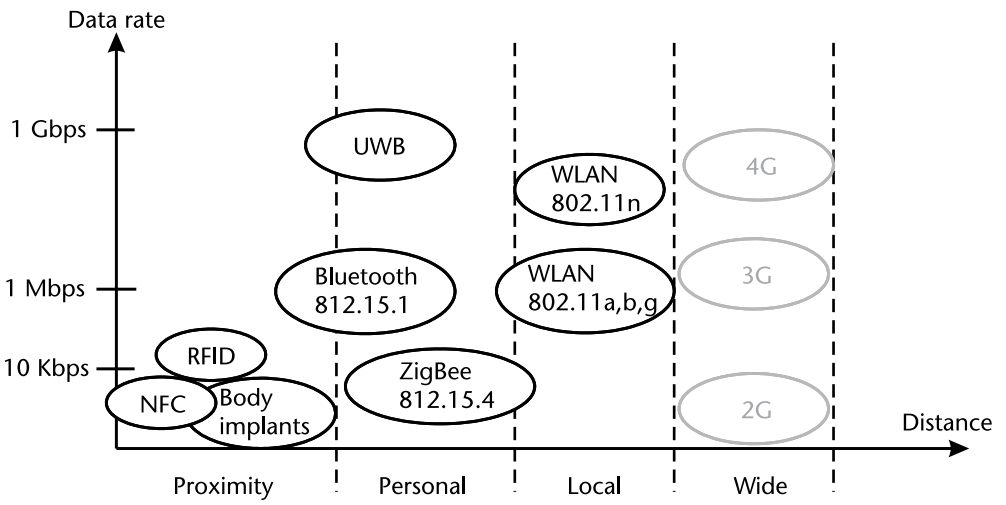


Figure 2.6 Short-range communications systems.

2. The concept called *ubiquitous* (meaning "that is everywhere") was introduced by Xerox's PARC (Palo Alto Research Center) in 1988 for the first time.

regions to ensure mass markets and minimize the chance of nonstandard equipment appearing on the market.

Products typically operating in the frequency range between 300 MHz and 2.5 GHz are often referred to as industrial, scientific, and medical (ISM)-band products in the United States and short-range device (SRD) products in the European Union (EU). Both the U.S. and EU regulatory agencies place limitations on the operating frequencies, output power, spurious emissions, modulation methods, and transmit duty cycles, among other things.

The 2.4-GHz band is widely used by designers who want to build systems that can operate worldwide, and in fact, it has become the frequency band of choice for such standards as Bluetooth, WLAN, and ZigBee. The 5.8-GHz band has also attracted some attention in cordless phones or the 802.11a version of WLAN, for example.

In the United States, the regulations governing *license-exempt* wireless products fall into two broad categories: periodic devices and ISM-band devices. The type of application and the communications range determine the appropriate band and Federal Communications Commission (FCC) classification to use. In the European Union, the regulations governing low-power wireless devices are essentially defined by two separate bodies. One group defines the allocation of frequency bands and their use, and another group defines the test methodologies and general transceiver specifications.

License-exempt systems often (although not always) use *spread-spectrum* principle of operation, originally developed by the military to counter attempts to detect, decode, or block signal transmissions. The most important peacetime characteristics of spread-spectrum systems are that they facilitate radio communications in a manner that minimizes the potential to cause harmful interference to other services and are able to withstand higher levels of interference than other technologies. Hence, spread-spectrum systems have significant potential to share common spectrum with other services. Spread-spectrum devices operate on a license fee-exempt basis, if the technical conditions are met, and type approval is mandatory.

Two main types of spread-spectrum system are commercially available: *direct sequence* and *frequency hopping*. Prior to the latest EN 300 220 specification (1996), however, the U.S. and European bodies took vastly different regulatory approaches. The United States adopted a frequency-hopping approach, while Europe applied duty-cycle limits in each of the subbands as described in the ERC REC-70 document. Both of these implementations are useful in minimizing interference, but manufacturers who were designing systems for both regions needed to completely rewrite the media-access layer (MAC) in the system's communication protocol. Fortunately, the latest European EN 300 220 regulations have extended the frequency bands to allow for frequency-hopping spread-spectrum (FHSS) or direct-sequence spread-spectrum (DSSS), thus making the MAC implementations more similar to those designed for the U.S. market.

Aside from the two most popular spread-spectrum systems, described below, an interesting aspect of the new European regulations is that they provide for other wideband spread-spectrum modulation schemes in addition to FHSS and DSSS. FSK/GFSK (Gaussian frequency-shift-keying) modulation, with an occupied bandwidth greater than 200 kHz, is considered wideband modulation under the European regulations.

Many countries do not have rules for license-exempt systems like the United States. Instead, they allocate spectrum on a primary or secondary basis and require that all radio transmitters be licensed by the government. Similar to the United States, several other countries do not allocate spectrum specifically for RFID but to categories of service such as *short-range devices*. In 2006, the ITU-R released a Recommendation SM.1538-2 [4] that outlines the spectrum requirements and regulatory approaches applicable to short-range devices in Europe, the United States, the People's Republic of China, Japan, and South Korea.

2.2.1 Frequency-Hopping Spread-Spectrum Systems

The FHSS transmission technology spreads energy in the time domain by dividing the spectrum into a number of channels, switching between them in a pseudo-random sequence or *hopping code,* that is known by both the receiver and transmitter. U.S. and European standards both specify a similar number of hopping channels, and a maximum *dwell time* (the time spent at a particular frequency during any single hop) of 400 ms.

Bandwidths of up to 7 MHz are available once either the listen-before-talk (LBT) or *duty-cycle* limits are met, as compared to the 2-MHz range available previously. Listen-before-talk, a "polite" communication protocol, scans the channel for activity before initiating a transmission. Also called clear-channel-assessment (CCA), systems using it with frequency hopping have no duty-cycle limitations.

2.2.2 Direct-Sequence Spread-Spectrum Systems

Besides FHSS, direct-sequence spread-spectrum (DSSS) is also addressed in the new European regulations. In a DSSS system, a narrowband signal is multiplied by a high-speed pseudorandom number (PRN) sequence to generate a spread signal. Each PRN pulse is called a *chip,* and the rate of the sequence is called the *chip rate*. The extent to which the original narrowband signal is spread is referred to as the *processing gain;* it is the ratio of the chip rate to the narrowband data symbol rate.

At the receiver, the incoming spread-spectrum signal is multiplied with the same PRN code to despread the signal, allowing the original narrowband signal to be extracted. At the same time, any narrowband interferers at the receiver are spread and appear to the demodulator as wideband noise. The allocation of different PRN codes to each user in the system allows isolation between users in the same frequency band. This is known as code-division multiple access (CDMA).

A few examples of systems using DSSS modulation include IEEE 802.15.4 (WPAN), IEEE 802.11 (WLAN), and the global positioning system (GPS). The main advantages of DSSS are:

- *Interference resilience:* The essence of the interference-rejection capability of DSSS is that the useful signal gets multiplied twice (spread and despread) by the PRN code while any interferers are multiplied just once (spread).
- *Low-power spectral density:* Introducing minimal interference with existing narrowband systems.

- *Security:* Very resistant to jamming because of spreading/dispreading.
- Mitigation of multipath effects.

2.3 WLAN

2.3.1 Basics of WLAN

Four standards dominate the WLAN marketplace; 802.11b has been the industry standard for several years. Operating in the license-exempt portion of the 2.4-GHz RF spectrum, it delivers a maximum data rate of 11 Mbps and boasts numerous strengths. 802.11b has broad user acceptance and vendor support since many vendors manufacture compatible devices, and this compatibility is assured through the Wi-Fi certification program. Thousands of enterprise organizations that typically find its speed and performance acceptable for their current applications have deployed 802.11b technology.

In the United States, a number of wireless ISPs have emerged who are offering public access services using IEEE 802.11b equipment operating in the 2.4-GHz band. These providers are targeting public areas in which business travelers may wish to access corporate intranets or the Internet, for instance, in hotels or coffee shops. In some parts of Europe there are now a number of service providers, both mobile operators and ISPs, offering wireless Internet services based on 802.11b technology in the 2.4-GHz band.

Another WLAN standard, IEEE 802.11a, operates in the uncluttered 5-GHz radio frequency spectrum. With a maximum data rate of 54 Mbps, this standard offers a fivefold performance increase over the 802.11b standard. Therefore, it provides greater bandwidth for particularly demanding applications. The IEEE ratified the 802.11a standard in 1999, but the first 802.11a-compliant products did not begin appearing on the market until December 2001. The 802.11a standard delivers a maximum data rate of 54 Mbps and eight nonoverlapping frequency channels, resulting in increased network capacity, improved scalability, and the ability to create microcellular deployments without interference from adjacent cells.

Operating in the license-exempt portion of the 5-GHz radio band, 802.11a is also immune to interference from devices that operate in the 2.4-GHz band, such as microwave ovens, cordless phones, and Bluetooth (a short-range, low-speed, point-to-point, personal-area-network wireless standard). The 802.11a standard is not compatible with existing 802.11b-compliant wireless devices. The 2.4-GHz and 5-GHz equipment can operate in the same physical environment without interference.

IEEE 802.11g is also widely adopted high-performance standard, it delivers the same 54-Mbps maximum data rate as 802.11a, and it operates in the same 2.4-GHz band as 802.11b.

Because this spectrum is license-exempt, even more uses for it are expected to develop in the future. As the band becomes more widely used, radio interference will increase. Bluetooth uses FHSS and is a shorter range and lower bandwidth technology than 802.11b and it uses frequently changing, narrow bands over all channels. Not only do microwave ovens operate within this range, but other RF communications technologies as well, most notable of which is IEEE 802.11b.

802.11n is newer standard for wireless local-area networks with the real data throughput estimated to reach a theoretical 540 Mbps. 802.11n builds upon previous 802.11 standards by adding multiple-input-multiple-output (MIMO).

MIMO uses multiple transmitter and receiver antennas to allow for increased data throughput through spatial multiplexing and increased range by exploiting the spatial diversity and new coding schemes. Channels operating at 40 MHz are another feature incorporated into 802.11n, which doubles the channel width from 20 MHz in previous 802.11 PHYs to transmit data. The IEEE approved the amendment, and it was published in October 2009.

2.3.2 WLAN Components

A WLAN is made up of two key components: an access point (AP), or base station, that is usually, but not necessarily, physically connected to a LAN and a wireless card that is either built into or added to a computer device, be it a handheld (i.e., PDA), tablet, laptop, or desktop computer (Figure 2.7).

With a wireless LAN in place, portable computers can remain connected to the network while on the move. Any device with a wireless adaptor within range of an access point can potentially connect to the WLAN, thus providing greatly increased freedom and flexibility compared to a wired network.

Extending the WLAN to include additional users often only requires that the user has a wireless-enabled computer device and is in range of an access point.

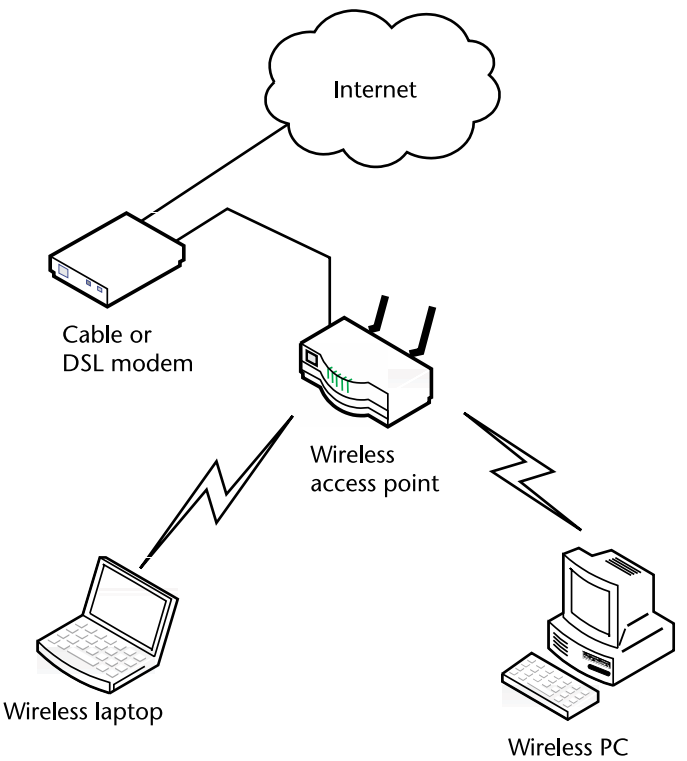


Figure 2.7 WLAN components.

Increasing the overall network coverage of the WLAN can often be achieved by adding further access points.

2.4 Wireless Personal Area Network

Wireless personal area network (WPAN) is defined as a network of a very limited radius, up to 10 feet (3.3m), occupying a very small amount of space. Specification 802.15 is a wireless specification defined by IEEE for WPANs, which has characters such as short range, low power, low cost, small networks and communication of devices within a personal operating space (POS).

As radios decrease their cost and power consumption, it becomes feasible to embed them in more types of electronic devices, which can be used to create smart homes, sensor networks, and other exciting new applications. Two radio technologies have emerged to support this trend, Bluetooth (IEEE 802.15.1) and ZigBee (IEEE 802.15.4). WPANs come in another variation that targets high data rates for image and multimedia applications, known as high-rate wireless personal area network, described by the IEEE 802.15.3 standard.

2.4.1 Bluetooth

Bluetooth[3] radios provide short-range connections between wireless devices along with rudimentary networking capabilities. Bluetooth is mainly used for short-range communications, for example, from a laptop to a nearby printer or from a cell phone to a wireless headset. Its normal range of operation is 10m (at 1-mW transmit power), and this range can be increased to 100m by increasing the transmit power to 100 mW.

The Bluetooth standard is based on a tiny microchip incorporating a radio transceiver that is built into digital devices. The transceiver takes the place of a connecting cable for devices such as cell phones, laptop and palmtop computers, portable printers and projectors, and network access points. The system operates in the license-exempt 2.4-GHz frequency band; hence, it can be used worldwide without any licensing issues.

For carrier frequencies, 79 RF channels are available in the 2.4-GHz frequency band and the following relation is valid:

$$f = 2,402 + k[MHz], \quad \text{where} \quad k = 0,1,2,\ldots,78 \tag{2.15}$$

The RF channel bandwidth is 1.0 MHz. In order to comply with out-of-band regulations, the guard bands are defined. The lower guard-band bandwidth is 2.0 MHz, and the upper guard-band bandwidth is 3.5 MHz.

The Bluetooth standard provides one asynchronous data channel at 723.2 kbps. In this mode, also known as asynchronous connectionless (ACL) there is a

3. The Bluetooth standard is named after Harald I Bluetooth, the king of Denmark between 940 and 985 AD, who united Denmark and Norway. Bluetooth technology proposes to unite devices via radio connections, hence the inspiration for its name. The Bluetooth standard was developed jointly by 3Com, Ericsson, Intel, IBM, Lucent, Microsoft, Motorola, Nokia, and Toshiba.

reverse channel with a data rate of 57.6 kbps. The specification also allows up to three synchronous channels, each at a rate of 64 kbps. This mode, also known as Synchronous Connection Oriented (SCO), is mainly used for voice applications such as headsets, but can also be used for data. These different modes result in an aggregate bit rate of approximately 1 Mbps.

Routing of the asynchronous data is done via a packet switching protocol based on frequency hopping at 1,600 hops per second. There is also a circuit-switching protocol for the synchronous data. Bluetooth uses frequency hopping for multiple access with a carrier spacing of 1.0 MHz and typically, up to 80 different frequencies are used, for a total bandwidth of 80 MHz. At any given time, the bandwidth available is 1 MHz, with a maximum of eight devices sharing the bandwidth. Different logical channels (different hopping sequences) can simultaneously share the same 80-MHz bandwidth. Collisions will occur when devices in different *piconets,* on different logical channels, happen to use the same hop frequency at the same time. As the number of piconets in an area increases, the number of collisions increases, and performance degrades.

Bluetooth technology allows for the replacement of the many proprietary cables that connect one device to another with one universal short-range radio link. But beyond replacing the cables, Bluetooth radio technology provides a universal bridge to existing data networks, a peripheral interface, and a mechanism to form small, private, ad hoc groupings of connected devices away from fixed network infrastructures. Designed to operate in a noisy RF environment, the Bluetooth radio uses a fast acknowledgment and frequency-hopping scheme to make the link robust. Bluetooth radio modules avoid interference from other signals by hopping to a new frequency after transmitting or receiving a packet.

Compared with other systems operating in the same frequency band, its radio typically hops faster and uses shorter packets, thus making the Bluetooth radio more robust than other systems. Short packages and fast hopping also limit the impact of domestic and professional microwave ovens. Use of forward error correction (FEC) limits the impact of random noise on long-distance links. The encoding is optimized for an uncoordinated environment since Bluetooth radios operate in the license-exempt ISM band at 2.4 GHz.

Bluetooth can support an asynchronous data channel, up to three simultaneous synchronous voice channels, or a channel that simultaneously supports asynchronous data and synchronous voice. Each voice channel supports 64-kbps synchronous (voice) link. The asynchronous channel can support an asymmetric link of maximally 721 kbps in either direction while permitting 57.6 kbps in the return direction or a 432.6-kbps symmetric link.

The Bluetooth system supports both point-to-point and point-to-multipoint connections; a piconet is an ad hoc computer network of devices using Bluetooth technology protocols to allow one master device to interconnect with up to seven active slave devices (3-bit address space limits the number of devices in any piconet to eight). Up to 255 further slave devices can be inactive, or parked, which the master device can bring into active status at any time. Several piconets can be established and linked together in an ad hoc manner, in which each piconet is identified by a different frequency-hopping sequence, and all users participating on the same piconet are synchronized to this hopping sequence. The topology can best be described as a *multiple piconet structure.*

Voice channels use the continuous variable slope delta modulation (CVSD) voice coding scheme and never retransmit voice packets. The CVSD method was chosen for its robustness in handling dropped and damaged voice samples. Rising interference levels are experienced as increased background noise: even at bit-error rates (BER) up to 4%, the CVSD coded voice is quite audible. The Bluetooth air interface is based on a nominal antenna power of 0 dBm. The air interface complies with the FCC rules for the ISM band at power levels up to 0 dBm. Spectrum spreading has been added to facilitate optional operation at power levels up to 100 mW (20 dBm) worldwide. Spectrum spreading is accomplished by frequency hopping in 79 hops displaced by 1 MHz, between 2.402 GHz and 2.480 GHz. Due to local regulations, the bandwidth is reduced in Japan, France, and Spain. This is handled by an internal software switch.

Different master/slave pairs of the same piconet can use different link types, and the link type may change arbitrarily during a session. Each link type supports up to 16 different packet types. Four of these are control packets and are common for both SCO and ACL links. Both link types use a *time division duplex* (TDD) scheme for full-duplex transmissions. There are three error-correction schemes defined for Bluetooth baseband controllers:

- 1/3 rate FEC;
- 2/3 rate FEC;
- Automatic repeat request (ARQ) scheme for data.

The purpose of the FEC scheme on the data payload is to reduce the number of retransmissions. However, in a reasonably error-free environment, FEC creates unnecessary overhead that reduces the throughput. Therefore, the packet definitions have been kept flexible as to whether or not to use FEC in the payload. The packet header is always protected by a 1/3 rate FEC; it contains valuable link information and should survive bit errors. An unnumbered ARQ scheme is applied in which data transmitted in one slot is directly acknowledged by the recipient in the next slot. For a data transmission to be acknowledged both the header error check and the cyclic redundancy check must be okay; otherwise, a *negative acknowledge* is returned.

In April 2009, the Bluetooth Special Interest Group (SIG) formally adopted Bluetooth Core Specification Version 3 High Speed (HS), or Bluetooth 3. Bluetooth 3 gets its speed from the 802.11 radio protocol. The inclusion of the 802.11 Protocol Adaptation Layer (PAL) provides increased throughput of data transfers at the approximate rate of 24 Mbps. In addition, mobile devices, including Bluetooth 3, will realize increased power savings due to enhanced power control being built in.

In July 2010, the Bluetooth SIG announced the formal adoption of Bluetooth Core Specification Version 4.0 with the important new feature, Bluetooth low-energy technology. Low-energy technology is an evolution in Bluetooth wireless technology that will enable many new applications, including health care [5]. Similar to previous versions of Bluetooth, Bluetooth low-energy technology will operate using a simple protocol stack, and will be used primarily in short-range communications systems.

For example, a product that Bluetooth makes possible is a wireless electrocardiogram. Each patient lead can be designed as a separate battery-powered Bluetooth device that communicates with a battery-powered, Bluetooth-enabled patient monitor. That patient monitor, which also communicates with the hospital's 802.11b network (WLAN), is continuously sending the electrocardiogram data to the network, and the doctor can monitor this data from anywhere in the hospital using his or her handheld PDA.

2.4.2 ZigBee

The ZigBee[4] radio specification is designed for lower cost and power consumption than Bluetooth. ZigBee can be defined as a low tier, ad hoc, terrestrial, and wireless standard, and in some ways it is similar to Bluetooth. It is promoted by ZigBee Alliance and described in the IEEE 802.15.4 standard, but ZigBee has some features in addition to those of 802.15.4. Another protocol was needed because other existing short-range protocols such as 802.11 and 802.15 use too much power and the protocols are too complex, and thus too expensive, to be embedded in virtually every kind of device imaginable.

Potential applications are sensors, interactive toys, and remote controls. In addition, ZigBee technology makes it possible to control home networks, such as those used for controlling electrical appliances, checking temperature and humidity, and sending mobile messages to alarm in cases of trespass.

The goal of ZigBee is to provide radio operation for months or years without recharging, thereby targeting applications such as sensor networks and inventory tags. The IEEE 802.15 WPAN Task Group 4 (TG4) was chartered to investigate a low data rate solution with a multimonth-to-multiyear battery life and very low complexity, and the new standard has been published as IEEE 802.15.4-2006, which is the latest version at the time of this writing. It operates in the 2.4-GHz (ISM) radio band, the same band as 802.11b standard, Bluetooth, and microwaves. It is capable of connecting 255 devices per network [6].

The data rate of ZigBee is 250 kbps at 2.4 GHz, 40 kbps at 915 MHz, and 20 kbps at 868 MHz, whereas that of Bluetooth is 1 Mbps. ZigBee's data rates are slower than 802.11b (11 Mbps) and Bluetooth (1 Mbps), but it consumes significantly less power. The transmit power levels for 802.15.4 radios are very low, typically –3 dBm (0.5 mW). The receive sensitivity is –80 to –100 dBm, depending on the 802.15.4 radio.

ZigBee allows small, low-cost devices to quickly transmit small amounts of data such as temperature readings for thermostats, on/off requests for light switches, or keystrokes for a wireless keyboard. ZigBee devices, typically battery-powered, can actually transmit information much farther than 20m (60 feet) because each device within listening distance passes the message along to any other device within range, and only the intended device acts upon the message.

4. ZigBee takes its name from the dance that honey bees use to communicate information about newly found food sources to other members of the colony. The ZigBee Alliance, which totals nearly 100 companies, including Honeywell, Mitsubishi, Motorola, Philips, and Samsung, is hoping to penetrate the home-automation market overwhelmed by a variety of other proprietary technologies.

Although a ZigBee-compatible device for could be used for medically implanted wireless telemetry as well, there could be a few drawbacks. One is represented by the overheads included in the standardized devices that have to meet very broad market requirements. Furthermore, the IEEE 802.15.4 standard allocates most of the communication channels (16 out of 27) in the 2.45-GHz band. This frequency range is close to the resonance frequency of water molecules; therefore, this may severely affect the transmission efficiency from an implanted device, especially if placed in areas such as the stomach and the colon that are completely surrounded by living tissues. In addition, the power required to establish a reliable connection with an external unit may exceed the international regulatory levels for human safety.

In many ZigBee applications, the total time the wireless device is engaged in any type of activity is very limited; the device spends most of its time in a power-saving mode, also known as a *sleep mode*. ZigBee networks are primarily intended for low-duty-cycle sensor networks (<1%). A new network node may be recognized and associated in about 30 ms. Waking up a sleeping node takes about 15 ms, as does accessing a channel and transmitting data. ZigBee applications benefit from the ability to quickly attach information, detach, and go into a deep sleep, which results in low power consumption and extended battery life. As a result, ZigBee-enabled devices are capable of being operational for several years before their batteries need to be replaced [7].

One Zigbee application is in-home patient monitoring in which a patient's blood pressure and heart rate, for example, can be measured by wearable devices [8].

2.5 Wireless Body Area Networks

2.5.1 About Wireless Body Area Networks

So far we talked about WLANs and WPANs, and now we will focus on the wireless networks with the range of less than 1 foot, the wireless body area networks (WBANs). People with chronic disease and the elderly need effective personalized health monitoring and delivery and represents the primary motivation for the development of BANs, sometimes also referred to as body sensor networks (BSNs).

Devices such as implantable biomedical systems will become really miniature computers that employ sensitive, low-voltage, low-power application-specific integrated circuits (ASICs) to measure, monitor, and regulate physiological parameters and control the delivery of electrical impulses to different organs in the human body (Figure 2.8). The implanted medical devices and on-body sensors are mainly connected with monitoring tools to provide patient health data in real time using BANs.

Implantable medical devices (IMDs) are already very successful in the treatment of many diseases. By definition, an active implantable medical device (AIMD) is any active medical device that is intended to be totally or partially introduced, surgically or medically, into the human body or by medical intervention into a natural orifice, in which it is intended to remain after the procedure.

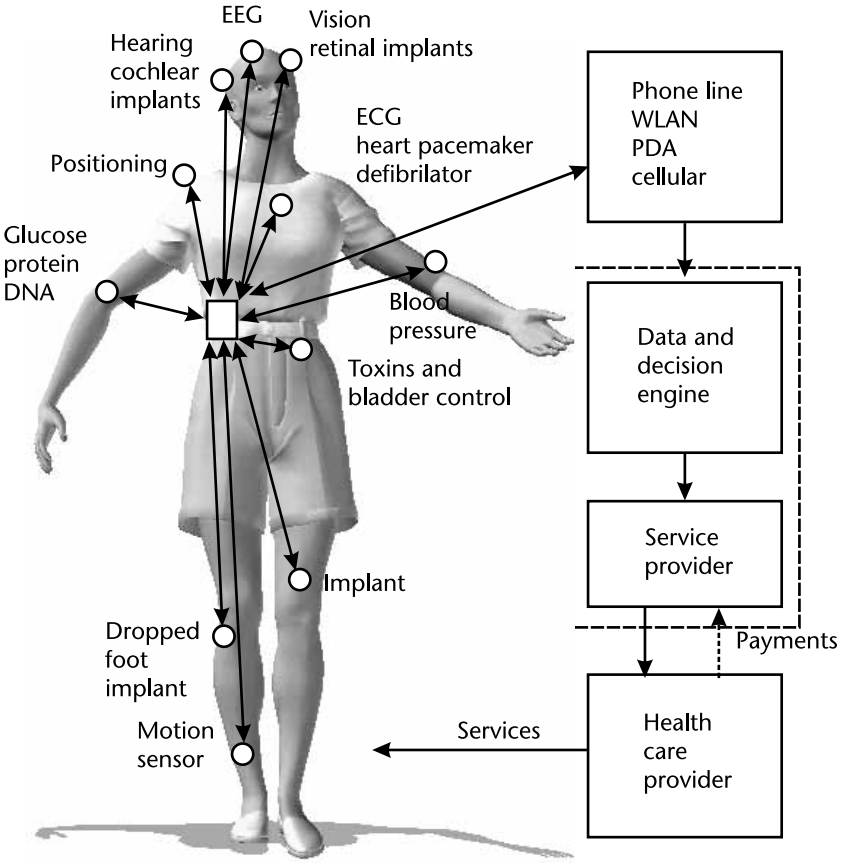


Figure 2.8 Wireless body area network.

These days, the new ultralow-power RF technologies are part of the development of innovative medical tools, from endoscopic camera capsules that are swallowed to implanted devices that wirelessly transmit patient health data. Radio links between external programming devices (or base stations) and medical implants are critical to the success of IMDs and have been researched extensively [9].

The network can deliver services to the person using the WBAN, including the management of chronic disease, medical diagnostics, home monitoring, biometrics, and sports and fitness tracking [10]. A health care provider provides a service that automatically collects data from the patient, integrates the data into the patient's medical record, processes the information, and issues recommendations, if necessary. The patient's electronic medical record is updated in real time.

The exact location of the patient can be determined using GPS, and medical professionals can monitor the activity of the patient and issue new guidance based on the new information. Standard IEEE 11073 describes and regulates communication between medical, health care, and wellness devices and external computer systems.[5]

5. The ISO/IEEE 11073 Medical/Health Device Communication Standards is a group of ISO, IEEE, and CEN joint standards that address the interoperability of medical devices, enabling communication between medical, health care, and wellness devices and with external computer systems.

Continuous monitoring the life signs of patients and analyzing the signal patterns enable early detection of the dangerous medical conditions and lead to more effective treatment and shorter hospital stays. In addition, long-term life-sign data could improve the quality of diagnoses when a person becomes ill. Many other applications of on-body sensor have been explored by research groups, and up-to-date information on the new products and applications can be found on the Medical Connectivity Web site [11].

The body network is composed of tiny portable devices equipped with a variety of sensors (such as heart rate, heart rhythm, temperature, pulse oximeter,[6] and accelerometer sensors) and performs biophysical monitoring, patient identification, location detection, and other desired tasks [12]. The energy consumption of these miniature sensors is also optimized so that the battery does not need to be changed regularly; they may use kinetic recharging.

Actuators notify the wearer of important messages from an external entity. For example, an actuator can remind an early-stage Alzheimer's patient to check the oven because sensors detect an abnormally high temperature, or a tone may indicate that it is time for the patient to take medication. A node in the body network is designated as the gateway to the emplaced sensor network. Due to size and energy constraints, nodes in this network have small processing and storage capabilities.

Today a number of other implants in use and in development, for example, brain pacemakers for the treatment of Parkinson's disease, implantable drug pumps, cochlea implants, artificial eyes, muscle stimulators and nerve-signal recorders for use with robotic prostheses, bladder control implants, pain control implants,[7] and so forth. All of these implants need some kind of data transfer, either in one or two directions. Neither inductive nor RF is the best for all of them because the power requirements, range, and speed are different from application to application.

In medical applications, the radio-communication link enables a clinician to reprogram therapy and obtain useful diagnostic information. The low-frequency *inductive links* (introduced in the early 1970s) used to be the most prevalent method of communication, typically operating in the tens to hundreds of kilohertz range, and with data rates of 1–30 kbps. These low-power systems can accommodate a small coiled antenna in the IMD, and have proven to be robust and sufficiently reliable. Unfortunately, antenna size and power limitations in implants result in a very low magnetic field strength for an IMD communication with an external programmer, and therefore they have a short range.

Using functional electrical stimulus (FES), implants can stimulate muscles or nerves in response to movement detected by sensors elsewhere on the body, allowing a paralyzed patient to walk again. Similarly, a radio-controlled valve for the urinary tract is in development that will be operated on-demand to restore bladder control.

6. A pulse oximeter is a medical device that indirectly monitors the oxygen saturation of a patient's blood and changes in blood volume in the skin.
7. Hollywood comedian Jerry Lewis has suffered from chronic back pain (and addiction to pain killers) for many years until April 2001, when he received an implant. The pain pacemaker delivers low-voltage stimulation to his spinal cord to block the pain messages from reaching his brain, so he no longer feels pain; his back problem still exists, but he is made to think it doesn't.

In case of ventricular fibrillation, a common type of cardiac arrest, the impulses travel through the heart as little wavelets, causing unsynchronized tightening and releasing of the muscle fibers in the ventricles, the two lower chambers of the heart. Without proper synchronization, blood flow ceases; starved of oxygen, other organs rapidly begin to fail, and within 10 minutes, the victim will almost certainly die.

In 2008, EU approved the Biotronik home monitoring system (the late Professor Dr. Max Schaldach, the company founder, developed the first German cardiac pacemaker in 1963) that lets physicians keep tab on patients at home. As with its other systems, this one relies on the wireless network to send vital information to the patient's physician. The system will work with a whole range of different implants, including pacemakers, implantable defibrillators, and cardiac resynchronization therapy devices [13].

In November 2010 the West Wireless Health Institute unveiled its new portable cardiotocography (electronic fetal monitoring) system. The device, which communicates either over cellular or a wireless LAN connection, should allow a physician to monitor fetal heartbeat and uterine contractions remotely [14].

In 2009, FDA approved a totally implanted brain stimulator intended to suppress symptoms associated with obsessive-compulsive disorder (OCD)[8] that are not adequately controlled with medications and/or other therapies [15]. an implanted pulse generator (IPG) is connected with a lead extension to a lead with four electrodes. The electrodes contact the patient at a specific location within the brain.

The pulse generator is implanted under the skin of either the abdomen or under the clavicle, and sends programmable electrical stimulation pulses to a selected combination of output electrodes within the brain. Two of these device systems may be implanted to stimulate both sides of the brain in order to relieve symptoms or one device with two lead outputs.

2.5.2 Technical Challenges of Body Area Networks

Integrated communications from different in-body implants and on-body sensors will also allow hearing for the deaf, sight for the blind, and mobility for the disabled. The objective is to design and manufacture intelligent medical devices that have communication capability and utilize the full range of advanced technologies in design, materials, processes and manufacturing [16].

To promote its widespread use, there are a number of technical challenges that need to be tackled, like the need for better sensor design, MEMS integration, biocompatibility, power source miniaturization, low-power wireless transmission, context awareness, secure data transfer, and integration with therapeutic systems. As radios decrease their cost and power consumption, it becomes feasible to embed

8. Obsessive-compulsive disorder is a type of mental disorder in which an individual experiences obsessions, compulsions, or both. Either the obsessive thought or the compulsive act may occur alone, or both may appear in sequence. Obsessions are recurring or persistent thoughts, images, or impulses that invade a person's consciousness despite attempts to ignore, suppress, or control them. Compulsions are urges (or impulses) to commit meaningless repetitive acts. Should the sufferer be forcibly or externally prevented from performing the compulsive act, he or she may experience an overwhelming anxiety.

them in more types of electronic devices, which can be used to create smart homes, sensor networks, and other new medical applications [17].

Microsensor devices are exposed to varied constraints that determine their size and capability (Table 2.1). For example, a body network device need not be especially small, but a medical implant often has size constraints that must be met. In addition, a wireless system designed to operate in close proximity to the human body must not exceed regulated power limits, usually at the expense of the range, especially if it is an implant. The constraints of the implanted wireless technology and the difficulties of the antenna design are described in [18].

In another example of implants described in [19], microfluidic systems allow analytical tools/sensors to complete assays more rapidly due to the relationship between size, diffusion, and time.

With the increased sophistication of medical implants, there is a growing need for flexible high-speed communication with the implant from outside the body. Up until recently, communication was achieved strictly using an inductive link operating at a low carrier frequency.

For example, a small coil is placed inside the case of the pacemaker, and a larger coil is placed upon the chest of the patient, directly on top of the pacemaker. The inductive coupling between these two coils is then used to transfer data to and from the pacemaker. The link is usually working at half-duplex meaning that transmission is in only one direction at any given time. The speed is typically low, a few hundred bits per second, and although the higher speeds are achievable, the low-carrier frequency limits the available data bandwidth severely.

Extended range and communication speed are possible to achieve by increasing the carrier frequency and the bandwidth; even the 2.45-GHz ISM band is a possibility, but has the drawback of being heavily used by other applications, such as wireless computer networks and microwave ovens. A number of advantages accrue if the communication with the implant can be moved to a higher carrier frequency. The first one is an increase in bandwidth, which makes it possible to achieve a higher bit rate. The second one is that a higher-frequency gives rise to a propagating electromagnetic wave, which makes the system usable at longer ranges. A longer communication range makes a number of new user scenarios possible.

A key element of an RF-linked implant is the in-body antenna, which must meet stringent biocompatible and size-limit requirements. An implanted transceiver

Table 2.1 Summary of Technical Challenges of BANs

Features	*Technical Challenges*
Size/packaging	Constrained space, miniaturization, operates in harsh environment, conforms to medical standards
Power	Size, replacement issues, power scavenging still premature and unreliable
Lifetime	Devices must operate over extended periods of time
Electronics/sensors	Miniaturization, partitioning into system nodes, choice of different principles of operation
Wireless	Conformity to standards, complex radio environment, tissue absorption
Cost	Customized and therefore expensive
Security	Encryption should protect user's privacy

also faces numerous RF challenges. Unlike free-air performance, the human body is often an unpredictable and hostile environment for a wireless signal. Improving therapy and diagnoses, an implanted pacemaker will regularly transmit performance data and the patient's condition to a doctor's office. If the pacemaker detects a cardiac arrest, the device could signal to a base station to alert an emergency response team.

2.5.3 Principle of Inductive Coupling

2.5.3.1 Brief Description and Applications

The inductive link uses the phenomenon of mutual inductance of two coupled coils for energy transfer and data transmission utilizing the inductive near field [20]. In the past years, an electromagnetic link is the most common link used for communication and to supply power to the implanted medical devices. Since the first pacemaker was implanted in the human body in 1958, all communication takes place through electromagnetic link. With the amplitude modulation their use can be expanded, for example, to the transmission of pulse trains for deep brain stimulation (DBS).

Implants have a small coil outside the small housing, while the external source has another coil and is placed on the skin of the patient directly above the implant. The inductive coupling between these two coils serves as the communication channel but with the range of only a few inches. External circuitry consists of a modulator and a transmitter. Transmitter is usually designed using Class D or E power amplifier to provide maximum power with high efficiency. The internal (implanted) circuitry contains voltage regulator and a demodulator [21]. In addition, previous studies have shown that RF energy between 1 and 10 MHz penetrates the body with minimum energy loss.

Systems with two coils (one internally and one externally) effectively behave as the primary and secondary coils of a transformer. Detailed design and experimentation have shown that such a system can communicate effectively over the required range (typically few inches) consuming less electrical power than a more conventional radio system [22]. Electromagnetic induction is an attractive option, as it not only reduces the power constraint on the device but also removes the need for batteries.

The maximum power density permitted near the human body is in the region of 10 W/sq m but varies with frequency and differs from country to country [23].

2.5.3.2 Theory of Operation

In general, inductive coupling is the transfer of energy from one circuit to another through a shared magnetic field. An electrical current passing through the coil of a primary conductor creates a magnetic field that induces an electrical current in the coil of a secondary conductor exposed to the magnetic field. A complete inductive powering system consists of two major parts; a drive coil and a pickup coil, and their associated impedance matching networks.

The *drive coil* (reader) is designed to maximize the magnetic field within the desired enclosure at the drive frequency. Similarly, the *pickup coil* (tag) is designed

to maximize the amount of magnetic flux density converted to power for the implant, while minimizing its own dimensions. The following is description of the reader/tag operation that can be used in the telemetry system for implanted devices in humans or animals.

Without getting into too many details, we can say that a transformer is a couple of coils of wire which transfer power from one to the other by a changing magnetic field. By having different numbers of windings, or turns of wire, a transformer can step up or step down an ac voltage. According to (2.16), the proportion of energy captured by the secondary coil can be represented by the coupling coefficient, k:

$$k = \frac{M}{\sqrt{L_1 L_2}} \tag{2.16}$$

where k is the coefficient of coupling and $0 \le k \le 1$, L_1 is the inductance of the first coil, L_2 is the inductance of the second coil, and M is the *mutual inductance*.

Mutual inductance depends only on the geometry of the two coils and is independent of the current in the coil. The coefficient of coupling is always between 0 and 1, and is an important factor in the operation of any inductively coupled system; $k = 1$ means that all of the magnetic flux produced by one coil passes through the other coil, which is practically never the case. In other words, this coefficient describes how closely linked the two inductors are magnetically; the better these two inductors are magnetically coupled, the more efficient the energy transfer between them should be.

The basic principles behind transferring power and data through an inductive link are the same as those used in transformer circuits; the major difference here is that in this case the two coils are fairly weakly coupled (i.e., through the air). Typical values for k in inductively powered system, in a very close proximity of a few millimeters, are between 0.01 and 0.1.

The coupling coefficient between the two coils, where the radius of the reader coil is much larger than radius of transponder (implanted) coil, can be determined empirically for the air-coupling case:

$$k = \frac{a_{\text{Implant}}^2 \cdot a_{\text{Reader}}^2}{\sqrt{a_{\text{Implant}} \cdot a_{\text{Reader}}}\left(\sqrt{r^2 + a_{\text{Reader}}^2}\right)^3} \cdot \cos\theta \tag{2.17}$$

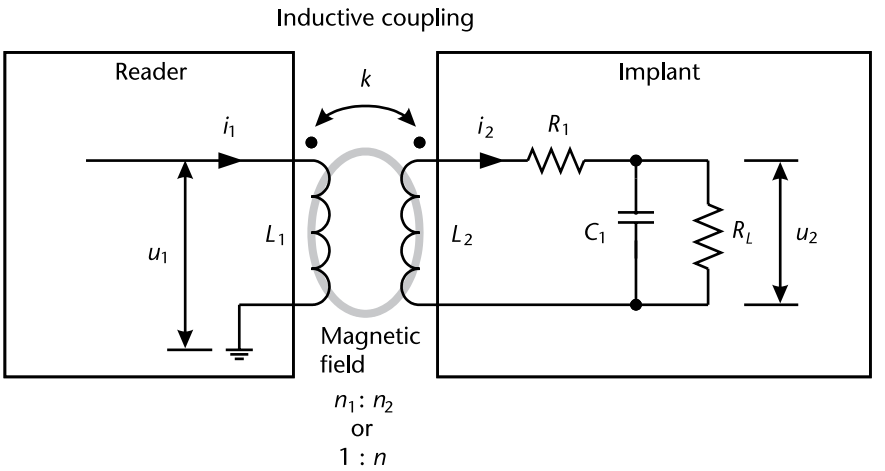
Equation (2.17) is based on the radii of the two coils ($a_{Implant}$ and a_{Reader}) and the distance r between them (see sample calculation in the Table 2.2). We will assume that the two coils are parallel ($\theta = 0° \Rightarrow \theta = 1$) and center aligned, with only air between the two coils. While not as accurate as finite element modeling, this still provides a good approximation of the system coupling coefficient [24].

This value can then be used in a simplified model of the complete inductively coupled system shown in Figure 2.9.

The two windings of the transformer were labeled rather generically 1 and 2. It is also quite common to refer to the two windings as *primary* and *secondary*. This convention is often used when a generator is connected to a primary winding, and

Table 2.2 Coupling Coefficient Values

Implant Radius (inches)	*Reader Radius (inches)*	*Distance (inches)*	*Coupling Coefficient k*
0.3	5.0	1.0	0.0139
0.3	7.0	1.0	0.0086
0.4	5.0	1.0	0.0213
0.4	7.0	1.0	0.0133
1.0	5.0	2.0	0.0716
1.0	7.0	2.0	0.0480



Note: The dot on each of the two coils indicates both voltages have same polarity.

Figure 2.9 Simplified model of the inductively coupled system.

a load is connected to a secondary winding. In that case, the energy flow is into the primary and out of the secondary; however, all transformers are bidirectional, so there is nothing inherently primary about either of the two windings.

The transformer turns ratio was indicated as n_1:n_2 or 1:n. Because the ideal transformer is a model for a real transformer, the numbers n_1 and n_2 may be the actual physical turns count of a transformer, such as 363:33. For circuit analysis purposes it is equivalent to give the turns ratio as 11:1, or as $n = 1/11 = 0.0909$. It might be noted that the turns ratio of a physical transformer is always a rational number, that is, the ratio of two integers. Therefore, a turns ratio of $\sqrt{3}:1$ is not possible, although 173 turns and 100 turns on a physical transformer would do a good job of approximating it.

The left side of this model represents the outside components of the system (reader), while the right side includes a basic model of the implanted system. Here, R_1 represents the parasitic resistance in the coil, C_1 is the tuning capacitance used to raise the coil voltage, and R_L is the load on the system. The weakly coupled transformer is used here to represent the two discrete coils, L_1 and L_2. The primary (reader's) coil, L_1, is driven by an RF amplifier supplying current at frequency ω. In the real system, R_L is time varying and complex, while in this model it is represented as a real resistor.

In circuit analysis it is quite common to use the concept of an *ideal transformer,* which does not generate, dissipate, or store energy (Figure 2.10). Therefore, the instantaneous power leaving the transformer is the same as that entering. In other words, if one were to draw a box around an ideal transformer and sum the power flows in and out of the box, the result is zero at every moment in time.

According to the basic equations governing ideal transformer (i.e., the magnetizing current is negligibly small) behavior, the current out of the transformer, i_2, is shown in here:

$$\begin{aligned} i_1 n_1 &= i_2 n_2 \\ i_2 &= \frac{n_1}{n_2} i_1 \\ i_2 &= \frac{i_1}{n} \end{aligned} \tag{2.18}$$

The ideal transformer has the voltage relationship shown in:

$$\frac{u_1}{u_2} = \frac{n_1}{n_2} \tag{2.19}$$

The voltage across the secondary of the transformer (the output windings) can be calculated as follows:

$$u_2 = \frac{n_2}{n_1} u_1 = n \cdot u_1 \tag{2.20}$$

Equation (2.21) illustrates the *impedance scaling* property of the ideal transformer.

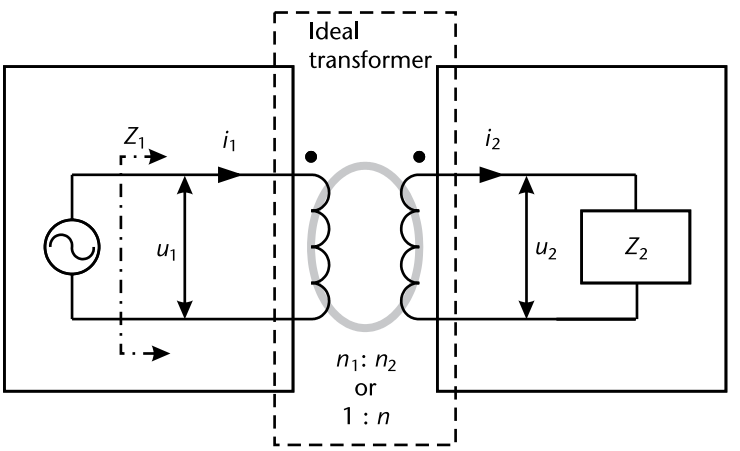


Figure 2.10 Ideal transformer.

$$Z_1 = \frac{u_1}{i_1} = \frac{u_2}{n^2 i_2} = \frac{Z_2}{n^2} \tag{2.21}$$

Impedance scaling property means that if a given impedance is connected to one winding of an ideal transformer, it will appear the same at the other winding scaled in magnitude by the turns ratio squared (n^2). The impedance appears greater at the winding having the greater number of turns and smaller at the winding having fewer turns.

The impedance scaling property of an ideal transformer allows circuit elements to be moved from one winding to another by scaling their impedances according to the square of the transformer turns ratio.

Now, the weakly coupled transformer can be replaced by an approximation (Figure 2.11). In this model, the new ratio n' (actually, a transfer function) is approximated by:

$$n' \approx k \times \sqrt{\frac{L_2}{L_1}} \tag{2.22}$$

Both the data signal and other external spurious magnetic fields are transformed by the primary coil. The magnitude of these transformed voltages depends on the coil geometry. Generally speaking, increasing primary inductance L_1 will cause an increasing influence of the external fields. With a low coupling coefficient k, the impedance seen by current i_2 is approximately equal to that of an inductor with value k^2L_2 (the impedance of this inductor is much lower than that of the other inductor in the circuit, therefore, nearly all of the current will flow through this inductor).

Using the impedance equation of the inductor at a known frequency, the voltage induced by this current is:

$$u_T = jX_L \cdot i_2 = j\omega L_2 k^2 \frac{1}{n'} i_1 = j\omega k \sqrt{L_1 L_2}\, i_1 \tag{2.23}$$

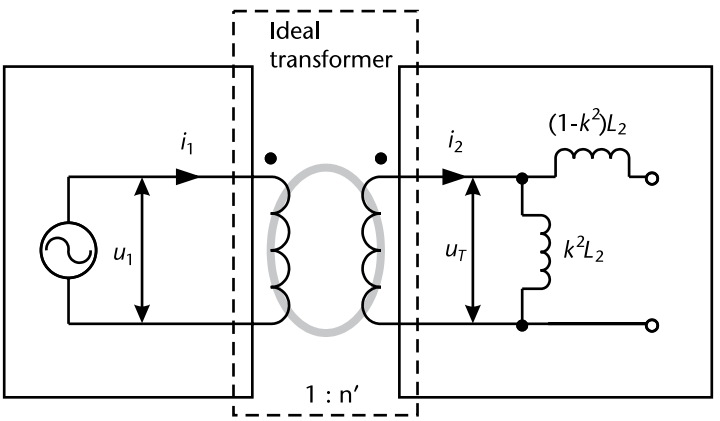


Figure 2.11 Model of a weakly coupled transformer.

Here u_T is the voltage induced by the current i_1 across the L_2 component of the transformer. We can now replace the weakly coupled transformer with a voltage source, u_T, in series with an inductor. With a small k value, we can approximate the value of this inductor by L_2 (Figure 2.12).

In this simplified case, a basic equation for the voltage across the load, u_2, is given by:

$$u_2 = \frac{u_T}{j\left(\frac{\omega L_2}{R_L} + \omega R_1 C_1\right) + \left(1 - \omega^2 L_2 C_1 + \frac{R_1}{R_L}\right)} \tag{2.24}$$

Equation (2.24) makes use of the impedances of the various components at a known frequency of operation (UHF band). Substituting in the equation for the transformer voltage and solving for the real part of the solution leads to a final answer with respect to the known parameters of the system:

$$u_2 = \frac{\omega k \sqrt{L_1 L_2}\, i_1}{\sqrt{\left(\frac{\omega L_2}{R_L} + \omega R_1 C_1\right)^2 + \left(1 - \omega^2 L_2 C_1 + \frac{R_1}{R_L}\right)^2}} \tag{2.25}$$

This creates a linear scale factor B, or the gain of the system, with units of ohms. The new equation for this particular system can be as follows:

$$u_2 = B \cdot k \cdot i_1 \tag{2.26}$$

where we require u_2 within a certain range and it depends on the distance between the two coils. Therefore, by changing the current through the primary coil of the system, the voltage on the implanted coil could be adjusted for a fixed coupling factor. This gives a required coil (rms) current on the order of 100 mA for the system with weak coupling.

Inductive coupling has many different applications as implanted electronic circuits, RFID, or even as a low-cost, low-power, high-bandwidth interface for interconnection of the modules on the chip [25].

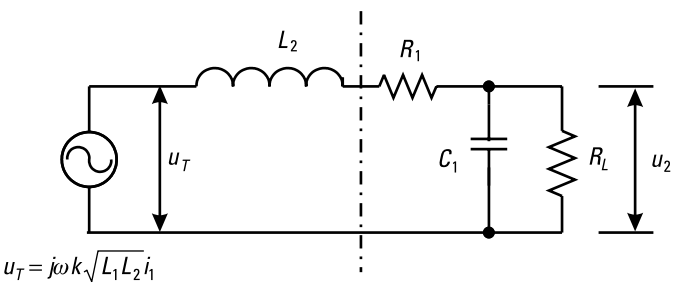


Figure 2.12 Equivalent model of the inductively coupled system.

2.5.4 Medical Implant Communication Service and Wireless Medical Telemetry Service Bands

2.5.4.1 MICS Development and Relevant Standards

As shown earlier, inductive links are short range and often require the external programmer to have contact with the skin of the patient directly over the implant. To overcome these operating-range and low-data-rate limitations, new ultralow-power RF technologies are being developed that operate at much higher frequencies, such as in the 433- and 915-MHz ISM bands and the more recently allocated 402–405-MHz Medical Implant Communication Service (MICS) band. RF integrated circuit technology can now offer low-power, reduced external component count and higher levels of integration, which will open new markets for medical device manufacturers.

From a regulatory viewpoint, the establishment of the MICS band began in the mid-1990s when Medtronic petitioned the FCC to allocate of spectrum dedicated to medical implant use. After gaining wider industry support, the 402–405-MHz MICS band was recommended for allocation by ITU-R Recommendation SA1346 in 1998. The FCC established the band in 1999, and similar standards followed in Europe.

The MICS band, located in the frequency range of 402–405 MHz, is reserved specifically for wireless data communications between implanted medical devices and external equipment. The FCC set aside this band because the signal-propagation characteristics in the band are particularly well suited for implantable applications, due to signal propagation characteristics in the human body, the relative dearth of other users in the band, and the ability to use the band internationally. The MICS use of this band is secondary to the primary users of this spectrum (i.e., Meteorological Aids Service, the Meteorological Satellite Services, and the Earth Satellite Service).

MICS band has general worldwide acceptance and has been approved in the United States, Europe, Canada, Australia, and Japan [26]. In addition, since the device's primary purpose is therapeutic, the communication link is used about 0.005% of the time, further limiting its interference potential.

The common standards should allow patients with implantable devices to obtain care in the United States and Europe. Relevant MICS standards could be found in FCC and ETSI documents [27–29].

The MICS standard is set by ETSI EN 301 839-1 but Specific Absorption Rate (SAR)[9] levels are not specified there; instead, safety limits can be taken from the International Commission on Non-Ionizing Radiation Protection (ICNIRP) [30].

MICS in Canada is regulated through RSS-243, "Active Medical Implant Communications System Devices in the 402–405 MHz Band," and these devices are defined as Category I equipment as per RSS-Gen. Devices certified under this standard are classified as Category I equipment and a Technical Acceptance Certificate (TAC), issued by the Certification and Engineering Bureau of Industry Canada, or a certificate issued by a recognized Certification Body (CB) is required.

9. Specific absorption rate (SAR) is a measure of the rate at which energy is absorbed by the body when exposed to an RF electromagnetic field (i.e., power absorbed per mass of tissue), expressed in watts per kilogram (W/kg).

Insuring the flexibility and scope of potential uses under this new service, the FCC proposed to revise its nomenclature and designate the entire 401–406-MHz band as MedRadio service. To accommodate a wider variety of devices than the current MICS service, which is limited to use of implant devices, the FCC proposed allowing the use of body-worn transmitting devices in the MedRadio service.

There are two principal fields of application for the MICS standard. The first one is for communication between a base station and an implanted device and the second one is for communication between medical implants within the same body. There is also a third possible use, rarely discussed, and that is the communication between medical implants in different bodies; this application is today fairly far-fetched but there are possible applications, such as mesh networking in order to increase the effective communication range.

2.5.4.2 Technical Characteristics of MICS

Technical rules were established to minimize interference and ensure safe coexistence of multiple MICS devices. The maximum transmitting power is very low, EIRP = 25 μW (–16 dBm), in order to reduce the risk of interfering with other users of the same band. The MICS band is broken into 300-kHz-wide channels. The rules specify that devices must listen for other devices before transmitting, called Listen-before-talk (LBT). If interference is encountered, the radio switches channels and listens again, a process known as *frequency agility.*

The rules also allow MICS devices to transmit without prior frequency monitoring in response to a non-RF actuation signal generated by a device external to the body (i.e., manual activation) or in response to a medical implant event (i.e., alert or alarm condition).

The MICS regulations require the system to perform a Clear-channel Assessment (CCA) in which the user scans all 10 of the 300-kHz channels and is allowed to transmit on the channel with the lowest ambient signal level (the least noisy channel). The user can also choose to transmit on the first available channel with an ambient power below a certain threshold (as defined in the standard). The MICS standard requires that the external programmer carry out the scanning process. For this reason, the IMD transceiver should support a low-power method of sniffing for the presence of an external programmer signal.

MICS regulations provide an exception to the CCA procedure in the event of an emergency medical event. For clinically significant medical emergencies, the IMD may transmit immediately on any channel. For example, if an implanted ECG monitor or pacemaker detects a cardiac arrest, the device could transmit immediately to a monitoring base station that, in turn, calls an emergency response service.

There is a growing need for implants, particularly heart implants, to communicate over greater distances than the current rules allow. The power permitted under MICS accommodates, at most, 6 to 8 feet of separation. For instance, it is increasingly difficult, if not impractical, to position implant monitoring equipment near patients in operating environments when physicians and nurses require unrestricted access to patients at all times. In addition, in operating theaters, implant monitoring equipment must be located outside the sterile field, which often means an estimated 30 feet or more between implant and reader. Furthermore, where multiple patients reside in a common areas (e.g., nursing homes, hospital wards, and so

forth), independent sessions with individual patients become increasingly economical and convenient as the distance between implants and programmers increases.

The main problem with MICS is that the spectrum currently allocated to MICS may not be insufficient to support the variety of implants and data rates being demanded by doctors (and patients) in the years ahead. Because of the reliance on spectrum for functionality of medical devices, and in anticipation of even greater usage, the FCC initiated a proceeding to make more spectrum available for medical devices.

2.5.4.3 Wireless Medical Telemetry Service

Wireless medical telemetry service (WMTS) enables monitoring equipment to remotely and unobtrusively observe several patients at one time. Such telemetry systems transmit real-time physiologic data, so it is critical to ensure that data are not lost or delayed.

Increasingly more radios for nonmedical applications are operating in the ISM bands, increasing the likelihood of signal loss and interference. In response to growing concerns about interference resulting from new digital television transmitters, low power television transmitters, and greater use of private land mobile radio equipment, the FCC established the WMTS, dedicating bands of frequencies for interference-free operation of medical telemetry systems.

This is the only frequency spectrum designated strictly for medical telemetry systems and neither video nor voice transmission is permitted; all transmitters operating in the WMTS bands must be registered in the database to ensure interference-free operation. Authorized health care providers, who desire to use wireless medical telemetry devices, before starting the operation must register all devices with a designated frequency coordinator [31].

The WMTS bands are 608–614 MHz, 1,395–1,400 MHz, and 1,427–1,432 MHz. All transmitters operating in the WMTS bands must be registered in the database to ensure interference-free operation. Prior to operation, authorized health care providers who desire to use wireless medical telemetry devices must register all devices with a designated frequency coordinator.

The 608–614-MHz band is shared with the radio astronomy service. There are 13 radio astronomy sites located throughout the United States. These sites have a protected radius of up to 50 miles. If the proposed WMTS deployment should fall within the protected radius of any of these sites, it is necessary to coordinate with the National Science Foundation (NSF).

The 1,395–1,400 MHz is shared with military radar systems. There are 17 radar sites located throughout the United States; these systems can have a protected radius of up to 55 miles. If the proposed WMTS deployment should fall within the protected radius of any of these sites, it is necessary to coordinate with the NTIA.

FCC excluded WMTS equipment from home use as well as the use of WMTS equipment in vehicles, including ambulances, due to potential interference issues. It would be difficult to ensure that WMTS equipment operated in ambulances or other vehicles would not interfere with other WMTS equipment operating on the same or adjacent frequencies at fixed sites in hospitals and health care facilities within the area passed by the ambulance.

Additional requirement for all WMTS transmitters that operate within 8 inches of a person's body (portable WMTS transmitters, for example) is that they must be routinely evaluated to demonstrate compliance with the FCC's RF radiation exposure guidelines. However, WMTS transmitters that operate at 8 inches or more from a patient's body are not required to undergo routine evaluation to demonstrate RF exposure compliance. Mobile WMTS transmitters are designed to normally operate with a separation distance of at least 8 inches from the radiating structures of a device and a patient's body.

Typically, medical telemetry devices are worn on the body of the patient, so it is expected that most WMTS transmitters will be classified as portable transmitters and will therefore be required to demonstrate RF exposure compliance with respect to the SAR limit as specified in 47 CFR 2.1093.

In some situations, the WMTS transmitter is mounted on a patient bed or incorporated within a separate device that is more than 8 inches away from the patient and nearby persons and in those cases, the WMTS transmitter will not be required to undergo routine evaluation to demonstrate RF exposure compliance. Due to relatively low transmitter power (less than 1.5-W EIRP) and the separation distance between the transmitter and the body of persons, the potential for mobile WMTS devices to exceed RF exposure limits is remote.

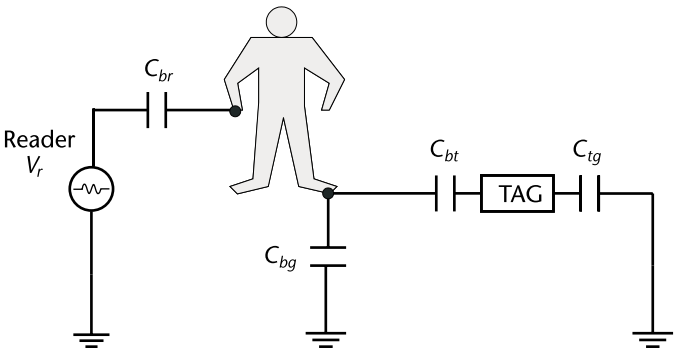
2.5.5 Passive Wearable Electrostatic Tags

Passive wearable electrostatic tags, also called body tags, have also been under development for some time. The wearable tag exploits the human body's natural ability to conduct electric fields and allows the wearer to present tags to tag readers through natural motions such as the grasp of a doorknob or the push of a button. The tag and reader imbue the user's physical gesture with digital meaning [32]. The body tag is also less expensive than other conventional inductive tags because it contains no magnetic flux coupling coil. The body may be modeled as a conductor surrounded by an insulator. Power and data signals may be coupled electrostatically to the body's interior and sent through it.

The human body acts as a poor conductor connecting the tag and the reader (Figure 2.13). However, displacement current, not dc current, passes through the user's body allowing the tag and reader to exchange data and power through the body. We call this type of communication *intrabody signaling.* At low frequencies, the human body appears to be a capacitive load; at higher frequencies the body radiates RF energy. It is possible to send power and data through the body by capacitively coupling displacement current into the body and using the ambient ground reference provided by our environment as the current return path.

The human body is modeled as a solid ideal conductor (the briny interior) surrounded by an ideal insulator (the skin). It is not a good idea to send dc current through the body because it is surrounded by an insulator, not to mention that it could be hazardous to present a constant voltage drop across the interior of the body. However, ac current can be sent through the body by capacitively coupling to its interior and using it as a single low-impedance node in a network of capacitors.

Each of these capacitances is on the order of 10 to 100 pF. Note that body couples to one electrode on the tag, while the tag's other electrode couples to the ambient ground. If we put the body tag into the shoe, these electrodes could be the



C_{rb} = Capacity between the tag reader and the body
C_{bg} = Capacity between the body and ambient ground
C_{bt} = Capacity between the body and the tag
C_{tg} = Capacity between the tag and ambient ground

Figure 2.13 Body tag circuit model.

top and the bottom of an inserted pad. Generally speaking, *wireless body-centric networks* consist of a number of nodes and units placed on the human body or in close proximity, such as on everyday clothing. Currently, it is used to receive or transmit simple information which requires very low processing capabilities.

However, some high-performance and complex units are needed in the future to provide the facilities for powerful computational processing with high data rates, for applications such as video streaming and heavy data communications. These have led to increasing research and development activities in the field of body area network applications for many purposes, with the main interest being health care and patient monitoring and task-specific/fully compatible wearable body networks (e.g., wearable computers) that have been applied in fields such as construction and medicine.

There are three primary criteria for wireless modules for wireless body-centric networks. First, they must support high data rates. Second, they must be small, both of which suggest the use of high frequencies. Third, they must consume a minimum of power, which implies highly efficient links. In terms of antennas and propagation, efficient design requires good understanding of the properties of the propagation channel involved and the development of optimized antennas.

2.6 Ultrawideband Technology

2.6.1 Ultrawideband Description

Ultrawideband (UWB) is a recently approved technology which relies on extremely short pulses that generate signals with very wide bandwidths, sometimes up to several gigahertz. UWB signals go undetected by most conventional receivers, minimizing their threat as harmful interferers. UWB technologies are currently being used in a variety of applications, such as ground penetrating radar, and are likely to be used in a variety of emerging applications, such as through-wall imaging and high-speed data transmission.

Gerald F. Ross first demonstrated the feasibility of UWB waveforms for radar and communications applications in the late 1960s and early 1970s [33]. Originally developed by the Defense Advanced Research Projects Agency (DARPA), the technology was called baseband, carrier-free, impulse communications or time-domain signaling, until the U.S. Department of Defense named it ultrawideband in 1989.

UWB radios are extremely wideband radios with very high potential data rates (Figure 2.14). The concept of ultrawideband communications actually originated with Marconi's spark gap transmitter, which occupied a very wide bandwidth. However, because only a single low-rate user could occupy the spectrum, wideband communications was abandoned in favor of more efficient communication techniques.

The renewed interest in wideband communications was spurred by the FCC's decision in 2002 to allow operation of UWB devices, as system underplayed beneath existing users, using over 7 GHz of bandwidth. These systems can operate in the 3.1–10.6-GHz range.

FCC defines UWB as any signal that occupies more than 500 MHz in the 3.1–10.6-GHz band and that meets the spectrum mask. Given the recent spectral allocation and the new definition of UWB adopted by the FCC, UWB is not considered a technology anymore, but available spectrum for license-exempt use. This means that any transmission signal that meets the FCC requirements for UWB spectrum can be considered UWB technology. This, of course, is not just restricted to impulse radios or high-speed spread-spectrum radios pioneered by companies so far, but to any technology that utilizes more than 500-MHz spectrum in the allowed spectral mask and with the current emission limit's restrictions.

In theory, the system could interfere with all the systems in that frequency range, including critical safety and military systems, license-exempt systems such as 802.11 wireless and Bluetooth, and cellular systems where operators paid billions of dollars for dedicated spectrum use. The FCC's ruling was quite controversial given the vested interest in interference-free spectrum of these users. To minimize the impact of UWB on primary band users, the FCC put in place severe transmit power restrictions. This requires UWB devices to be within close proximity of their intended receiver [34].

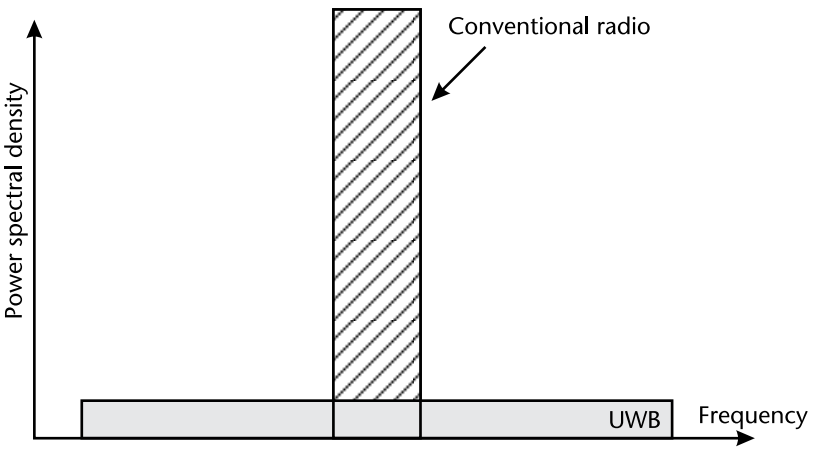


Figure 2.14 Conventional and UWB radio transmission.

There are two common forms of UWB: one is based on sending very short duration pulses to convey information, and the other uses multiple simultaneous carriers. Each approach has its relative technical merits and demerits. The most common form of multicarrier modulation, OFDM, has become the leading modulation for high-data-rate systems, and much information on this modulation type is available in recent technical literature.

Pure impulse radio, unlike classic communications, does not use a modulated sinusoidal carrier to convey information. Instead, the transmit signal is a series of baseband pulses. Because the pulses are extremely short (commonly in the nanosecond range or shorter), the transmitted signal bandwidth is in the order of gigahertz.

2.6.2 UWB Technical Specifications

UWB radios come with unique advantages that have long been appreciated by the radar and communications communities. Their wideband nature allows UWB signals to easily penetrate through obstacles and provides very precise ranging capabilities. Moreover, the available UWB bandwidth has the potential for very high data rates. Finally, the power restrictions dictate that the devices can be small with low-power consumption.

Initial UWB systems used ultrashort pulses with simple amplitude or position modulation. Multipath can significantly degrade performance of such systems, and proposals have been suggested to mitigate the effects of multipath equalization and multicarrier modulation. Precise and rapid synchronization is also a big challenge for these systems. Although many technical challenges remain, the appeal of UWB technology has sparked great interest both commercially and in the research community to address these issues.

UWB has several features that differentiate it from conventional narrowband systems:

- Large instantaneous bandwidth enables fine time resolution for network time distribution, precision location capability, or use as radar.
- Short duration pulses are able to provide robust performance in dense multipath environments by exploiting more resolvable paths.
- Low power spectral density allows coexistence with existing users and has a Low Probability of Intercept (LPI).
- Data rate may be traded for power spectral density and multipath performance.

On February 14, 2002, the FCC issued a First Report and Order, which classified UWB operation into three separate categories, and each category was allocated a specific spectral mask:

1. Communication and measurement systems;
2. Vehicular radar systems;
3. Imaging systems, including ground penetrating radar, through-wall imaging and surveillance systems, and medical imaging.

The FCC ruling, however, did not specifically address precision location for asset tracking or inventory control. These applications, known as *location-aware communication systems,* are a hybrid of radar and data communications that use UWB pulses to track the 2-D and 3-D position of an item to accuracies within a few tens of centimeters, as well as transmitting information about the item, such as its contents, to a centralized database system. The FCC has only specified a spectral mask and has not restricted users to any particular modulation scheme.

As discussed previously, a number of organizations are promoting multicarrier techniques, such as Orthogonal Frequency-division Multiplexing (OFDM), as a potential alternative for high-data-rate communications. Beyond the United States, other countries have been using a similar approach toward licensing UWB technology; in both Europe and Japan, initial studies have been completed, and regulations are expected to be issued in the near future that are expected to harmonize with the FCC mask.

UWB systems are approved for license-exempt use within the United States under FCC Part 15, specifically Subpart F and Part 15.250, permitting both indoor and outdoor use. The FCC Part 15.250 band spans from 5,925–7,250 GHz.

In March 2007, the European Commission (EC) formally adopted a UWB frequency range from 3.4 to 4.8 GHz and 6 to 8.5 GHz, for use in EC member countries, which will establish several frequency limitations requiring UWB vendors to alter their technology to meet those limits. The EC decision designates the frequency bands of 3.4 to 4.8 GHz and 6 to 8.5 GHz for use by UWB RFID tags and interrogators, as well as for other applications, such as data networking. UWB devices utilizing frequencies between 4.2 GHz and 4.8 GHz were permitted only until December 31, 2010, after which time they had to be converted to the 6- to 8.5-GHz band.

2.6.3 UWB Medical Applications

In recent years, many proposals have been made to address the privacy and security issues of the WLAN and RFID systems. One of the proposals is to implement the link using UWB communications, since the use of an advanced modulation scheme offers a new approach to the security [35]. By using the modulation spreading code as a secret parameter of the communications link, it is possible to make eavesdropping extremely difficult and therefore increase the communication reliability.

Recently, a UWB-based data link has been reported in biomedical applications to transmit data since it can offer low-power consumption, low signal power spectrum density, high data rate, and immunity against power interference and noise. To optimize receiver structures and antennas for UWB WBANs with respect to energy efficiency and complexity, the distinct features of the body area network channel have to be considered. Thus, it is necessary to know the propagation mechanisms in the proximity of the human body. Research in [36] is limited to transmission at the head, especially focusing on the link between both ears, and considers direct transmission, surface waves, reflections, and diffraction as possible propagation mechanisms.

The use of UWB radio in RFID/BAN systems could bring significant benefits; for example, reduced risk of interfering with sensitive medical equipment in hos-

pital/health care applications, and very precise positioning of the person. These proposals and ideas are still under development.

2.6.4 Orthogonal Frequency-Division Multiplexing

The basic idea behind OFDM is the use of a large number of parallel narrowband subcarriers instead of a single wideband carrier to transport information. OFDM refers to the use, by a single transmitter, of a set of frequency multiplexed signals with the exact minimum frequency spacing needed to make them orthogonal so that they do not interfere with each other.

The advantages are that such a system is very efficient in dealing with multipath and robust again narrowband interference. Disadvantages include sensitivity to frequency offset and phase noise and peak-to-average problem reduces the power efficiency of RF amplifier at the transmitter.

Multicarrier communications were first used in the late 1950s and early 1960s for higher data-rate high-frequency military communications. Since that time, OFDM has emerged as a special case of multicarrier modulation using densely spaced subcarriers and overlapping spectra, and was patented in the United States in 1970. However, the technique did not become practical until several innovations occurred. Fortunately, the apparently very complex processes of modulating (and demodulating) thousands of carriers simultaneously are equivalent to *discrete Fourier transform* operations, for which efficient fast Fourier transform (FFT) algorithms exist. Thus, integrated circuit implementations of OFDM demodulators are feasible for affordable mass-produced receivers.

Throughout the 1980s and 1990s, other practical issues in OFDM implementation were addressed, such as oscillator stability in the transmitter and receiver, linearity of the power amplifiers, and compensation of channel effects. Doppler spreading caused by rapid time variations of the channel can cause interference between the carriers and held back the development of OFDM until the development of a *coded multicarrier modulation.*

The concept of using parallel data transmission by means of frequency division multiplexing (FDM) was published in mid-1960s, while some early development can be traced back to the 1950s. A U.S. patent was filled and issued in January 1970. The idea was to use parallel data streams and FDM with overlapping subchannels to avoid the use of high-speed equalization, to combat impulsive noise and multipath distortion, and to fully use the available bandwidth. The initial applications were in the military communications.

In the telecommunications field, the terms of discrete multitone (DMT), multichannel modulation, and multicarrier modulation (MCM) are widely used, and sometimes they are interchangeable with OFDM. In OFDM, each carrier is orthogonal to all other carriers (see Figure 2.15); however, this condition is not always maintained in MCM. We could say that the OFDM is an optimal version of multicarrier transmission schemes.

OFDM is also used in asymmetric digital subscriber line (ADSL) Services, digital audio broadcast (DAB), digital terrestrial television broadcast (DVB) in Europe, Integrated Services Digital Broadcasting (ISDB) in Japan, IEEE 802.11a/g, 802.16a, and power line networking (HomePlug). Because OFDM is suitable for high-data-

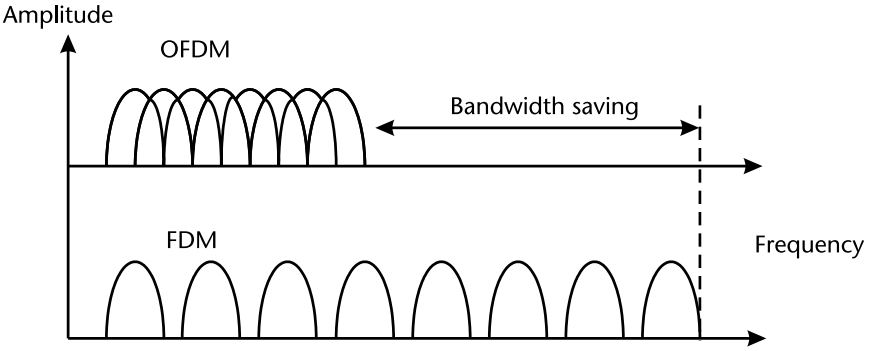


Figure 2.15 Difference between FDM and OFDM.

rate systems, it is also being utilized in the fourth generation (4G) wireless services, IEEE 802.11n (high-speed WLAN) and IEEE 802.20 (MAN).

2.7 Review Questions and Problems

1. List some of the most unusual applications of the wireless technology you have encountered. Describe the principle of operation, frequencies, and potential issues that users might face right now and in the future.
2. A 75Ω load is connected to a transmission line with the characteristic impedance of 50Ω. What is the VSWR at the load terminals? *(Answer: 1.5.)*
3. Calculate the wavelength of the signals that have a frequency of 125 kHz, 13.56 MHz, 915 MHz, and 2.4 GHz. *(Answer: 2.4 km, 22.2 m, 0.328 m, and 0.125 m.)*
4. Although widely accepted as an inventor of radio, Guglielmo Marconi, an Italian engineer working in England, based his work on patents and inventions of Nikola Tesla. Tesla was an inventor of (among other things) alternate current, used today in every household (Figure 2.16).

   In 1911, Tesla refused to share a Nobel Prize in Physics for the invention of radio transmission with Marconi. Tesla's patents from 1900 were reversed in favor of Marconi in 1904 after large private investments (including Edison) were made in Marconi's company. In 1943, the U.S. Su-

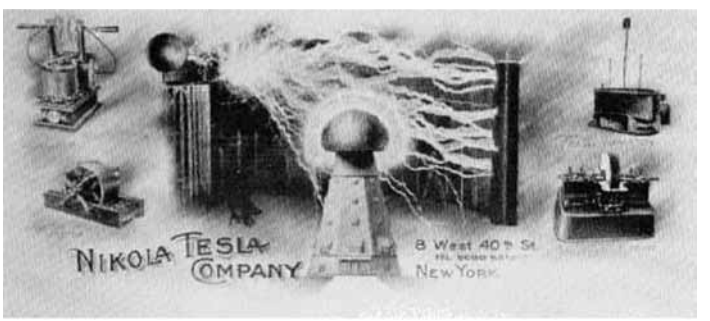


Figure 2.16 Nikola Tesla Company.

preme Court upheld Tesla's radio patent from 1900, and, as it stands today, Nikola Tesla is an official inventor of radio.

Although he is often nearly forgotten, Tesla held over 700 patents, and those patent are still valid. Write an essay describing the productive and very dynamic life of a genius and eccentric, Nikola Tesla.

5. Although it is possible to continually increase the transmitting power of a system in order to achieve a path of any length, it is not necessarily desirable to do so. As the transmit power of a system increases, the potential of the system to cause interference to other services also increases, which limits the use of spectrum in geographically adjacent areas. It is necessary to find a balance between the need of one user for increased power for his or her system and the need of another user for access to a channel to establish a service.

   During the 1920s, radio communication was a veritable free-for-all; anyone possessing radio equipment was allowed to broadcast signals over the air, resulting in chaos. Because interference resulted any time several transmitters operated in near proximity, no one could be assured of reliable communications. By the early 1930s, radio sales and usage plummeted, and the market failure created by this chaos predestined today's regulatory environment. Accordingly, with the passage of the Communications Act of 1934, Congress created the Federal Communications Commission to regulate radio communications in the United States, the District of Columbia, and all U.S. possessions [37]. The FCC has historically controlled access to radio spectrum by allocating specific frequency bands for use by licensed service providers.

   Today, many applications take advantage of the so-called *license-exempt frequency bands*. Research different applications and describe advantages and disadvantages of the wide use of these (license-exempt) frequency bands. Show your results in a form of a table and briefly summarize your findings.

6. Verify whether the following two statements are correct and justify the answer:
   - A horizontally positioned, linearly polarized receiver antenna is unlikely to capture any of the energy emitted by a vertically positioned, linearly polarized transmitter antenna.
   - A circularly polarized transmitter antenna will be able to communicate with any receiver antenna, regardless of its orientation.

7. List some of the ethical issues in the modern wireless (E911, for example) communications systems. Discuss the topic looking from different points of view (for example, an engineer, philosopher, doctor, clergyperson, politician, stay-at-home parent, or businessperson).

8. UWB, a technology that was recently approved by the FCC for a number of communications and sensing applications, is a signaling method that relies on short pulses that create extremely wide bandwidths. In addition to their potential for communications systems, UWB technology can also support the operation of new low-power radar products that can provide precise

measurement of distances or detection of objects underground or behind walls or other structures.

Find at least three suppliers of commercial systems based on the UWB technology. Describe their products and systems.

9. If you were a person in charge of the decision whether wireless human body implants should continue to be developed and used on people, what would you decide? Can you justify your decision in a way to be acceptable to all of the people all of the times?
10. What factors determine the range of a wireless link? Discuss the influence of each individual factor.
11. How would you shape a wire of fixed length to obtain the greatest and the smallest inductance?
12. There are many types of antenna radiation patterns, but the most common are omnidirectional, pencil beam, fan beam, and shaped beam. Discuss them in more detail and find examples of applications. More useful information can be found in [38].

References

[1] Saunders, R. S., and A. A. Zavala, *Antennas and Propagation for Wireless Communication Systems,* 2nd ed., New York: John Wiley & Sons, 2007.

[2] Best, S., *Antenna Polarization Considerations in Wireless Communications Systems*, Cushcraft Corporation, Manchester, NH, 2003.

[3] Huang, Y., and K. Boyle, *Antennas: From Theory to Practice*, New York: John Wiley & Sons, 2008.

[4] ITU-R Recommendation SM.1538-2, "Technical and Operating Parameters and Spectrum Requirements for Short-Range Radiocommunication Devices," 2006.

[5] http://www.bluetooth.com/English/Products/Pages/Low_Energy.aspx (accessed July 28, 2010).

[6] Ergen, S. C., *ZigBee/IEEE 802.15.4 Summary*, Internal Report to Advanced Technology Lab of National Semiconductor, Berkley, CA, 2004.

[7] Cao, H., et al., "Enabling Technologies for Wireless Body Area Networks: A Survey and Outlook," *IEEE Communications Magazine,* December 2009, pp. 84–93.

[8] Valdastri, P., et al., "Transmission Power Requirements for Novel ZigBee Implants in the Gastrointestinal Tract," *IEEE Transactions on Biomedical Engineering,* Vol. 55, No. 6, June 2008.

[9] Johansson, A. J., "Simulation and Verification of Pacemaker Antennas," *25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Cancun, Mexico, September 17–21, 2003.

[10] www.imec.be/ScientificReport/SR2007/html/1384142.html (accessed July 17, 2010).

[11] http://medicalconnectivity.com- (accessed July 19, 2010).

[12] Stankovic, J. A., "Wireless Sensor Networks," Charlottesville, VA: Department of Computer Science, University of Virginia, June 19, 2006.

[13] http://www.engadget.com/2008/10/02/biotronik-gets-go-ahead-for-gsm-based-implant-monitoring-system/ (accessed August 10, 2010).

[14] http://www.medgadget.com/archives/2010/11/portable_wireless_fetal_heartbeat_and_uterine_contractions_monitor.html, (accessed December 2011).

[15] http://www.accessdata.fda.gov/cdrh_docs/pdf5/H050003a.pdf (accessed July 17, 2010).

[16] Yang, G. -Z., (ed.), *Body Sensor Networks*, New York: Springer-Verlag, 2006.

[17] Zhen, B., et al., "Networking Issues in Medical Implant Communications," *International Journal of Multimedia and Ubiquitous Engineering,* Vol. 4, No. 1, January 2009.

[18] Chirwa, L. C., et al., "Electromagnetic Radiation from Ingested Sources in the Human Intestine Between 150 MHz and 1.2 GHz," *IEEE Transactions on Biomedical Engineering,* Vol. 50, No. 4, 2003, pp. 484–492.

[19] Madou, M. J., and R. Cubicciotti, "Scaling Issues in Chemical and Biological Sensors," *Proceedings of the IEEE,* Vol. 91, No. 6, 2003, pp. 830–838.

[20] Ahmadian, M., et al., "Data Transmission for Implantable Microsystem Using Magnetic Coupling," *IEEE Proceedings on Communications,* Vol. 152, No. 2, 2005, pp. 247–250.

[21] Hmida, G. B., et al., "Design of Wireless Power and Data Transmission Circuits for Implantable Biomicrosystem," *Biotechnology*, Vol. 6, No. 2, 2007, pp. 153–164.

[22] Wang, L., et al., "A Programmable Microsystem Using System-On-Chip for Real-Time Biotelemetry," *IEEE Transactions on Biomedical Engineering,* Vol. 52, No. 7, 2005, pp. 1251–1260.

[23] United Kingdom Regulations for Adults Exposed to Radiation in the Band from 10 MHz to 60 MHz, http://www.who.int/docstore/peh-emf/EMFStandards/who-0102/Worldmap5.htm (accessed December 2011).

[24] Sauer, C., et al., "Power Harvesting and Telemetry in CMOS for Implanted Devices," *IEEE Transactions on Circuits and Systems*, Vol. 52, No. 12, December 2005.

[25] Miura, N., et al., "Analysis and Design of Inductive Coupling and Transceiver Circuit for Inductive Inter-Chip Wireless Superconnect," *IEEE Journal of Solid States Circuits*, Vol. 40, No. 4, April 2005.

[26] Recommendation ITU-R SA.1346, "Sharing Between the Meteorological Aids Service and Medical Implant Communications Systems (MICS) Operating in the Mobile Service in the Frequency Band 401-406 MHz," 1998.

[27] FCC Rules and Regulations, "MICS Band Plan," Part 95, January 2003.

[28] FCC Rules and Regulations, 47 CFR 95.601-95.673 Subpart E, Federal Communications Commission, 1999.

[29] ETSI EN 301 839-1, "Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Radio Equipment in the Frequency Range 402 MHz to 405 MHz for Ultra Low Power Active Medical Implants and Accessories; Part 1: Technical Characteristics, Including Electromagnetic Compatibility Requirements, and Test Methods," European Telecommunications Standards Institute, 2002.

[30] ICNIRP, "Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic and Electromagnetic Fields (Up to 300GHz)," *Health Physics*, Vol. 74, No. 4, 1998, pp. 494–522.

[31] Yazdandoost, K. Y., and R. Kohno, "Frequency Band Considerations for the Use of Body Area Network," Doc: IEEE 802.15-07-0728-00-0ban, 2007.

[32] Babak, N., et al., "Passive Wearable Electrostatic Tags: The Bodytag," *Physics and Media*, MIT Media Lab, September 12, 1997.

[33] Bennett, C. L., and G. F. Ross, "Time-Domain Electromagnetics and Its Applications," *Proceedings of the IEEE*, Vol. 66, No. 3, March 1978.

[34] Reed, J., *An Introduction to Ultra Wideband Communication Systems*, Upper Saddle River, NJ: Prentice-Hall, 2005.

[35] Yu, P., et al., "Securing RFID with Ultra-Wideband Modulation," Virginia Tech, Electrical and Computer Engineering Department, Blacksburg, 2006.

[36] Zasowski, T. et al., "UWB Signal Propagation at the Human Head," *IEEE Transactions on Microwave Theory and Techniques,* Vol. 54, No. 4, April 2006.

[37] Carter, K. R., et al., "Unlicensed and Unshackled: A Joint OSP-OET White Paper on License-exempt Devices and Their Regulatory Issues," FCC, Office of Strategic Planning and Policy Analysis, OSP Working Paper Series, May 2003.

[38] Volakis, J. L., *Antenna Engineering Handbook*, 4th ed., New York: McGraw-Hill, 2007.

CHAPTER 3

Automatic Identification Systems

The technologies used in the world of automatic identification and data capture (AIDC) are varied and often used in combinations to provide a broader base of information flow. This chapter attempts to summarize the technologies in common use today and give the reader a basic understanding of the technology and its uses and limitations.

3.1 Bar Codes

Perhaps the oldest of the AIDC technologies, bar code technology, can be looked upon as the best-known and probably the most successful to date. We are all familiar with the basic bar code on a box of cereal or the jar of honey that we buy in the supermarket. This bar code is called UPC/EAN and is but one variation of over 250 bar codes that have been designed over time. Bar codes like this are referred to as *linear bar codes,* as they are made up of a collection of bars and spaces side by side (Figure 3.1). Most of these barcodes have never gained broad acceptance and we usually only consider maybe a dozen of linear barcodes.

Typical data content capacity varies from 8 to 30 characters with some bar codes restricted to numerals only and others using full alphanumeric information.

Linear bar codes are used in many applications where the use of a simple numeric or alphanumeric code can provide the key to a database of product. The most obvious limitation is the amount of data that can be stored in a linear bar code, though other problems can exist with the substrate on which the bar code is printed. The substrate might provide insufficient contrast or poor ink receptivity, which can cause the quality of the bar code to be less than ideal.

A new growth area in the world of bar codes is the 2-D version. Several variations of 2-D bar codes are available, but because these do not all comprise bars and spaces, the more accurate name of 2-D symbologies is used. Two-dimensional symbologies provide a means of storing large amounts of data in a very small space. Stacked symbologies (linear bar codes stacked on top of each other), matrix symbologies (comprising a matrix of light and dark elements, circles, squares, or hexagons), and packet symbologies (a collection of linear symbols "randomly" arranged on a page) are commonly used. Examples of the three types include PDF417, Code 49, Code 16K (stacked), Code One, MaxiCode, Data Matrix, Aztec Code, QR Code (matrix), and Super Code (packet).

Figure 3.1 Linear and two-dimensional (2-D) bar codes.

Two-dimensional symbologies have a major advantage over linear bar codes. They can store vast amounts of data. Individual symbols can store as much as 7,000 numeric-only or 4,200 alphanumeric characters. Many of the symbologies also have the ability to use a device called *structured append* that allows messages to be split over multiple symbols, providing almost infinite storage space.

The disadvantage of the 2-D symbologies is that a special scanner is needed. Matrix symbologies need a vision-based scanner to read the data, though some of the stacked symbologies can be read with a special laser scanner.

3.2 Card Technologies

3.2.1 Magnetic Cards

The first *magnetic stripe cards* were used in the early 1960s on transit tickets and in the 1970s for bank cards. Since then, the use of magnetic stripes continued to grow. Credit cards were first issued in 1951, but it was not until the establishment of standards in 1970 that the magnetic stripe became a factor in the use of the cards. Whether the card is a credit card-sized plastic card, a thin paper ticket, or an airline boarding card, the uses for magnetic stripe technology have grown considerably.

Today with an infrastructure that encompasses every store in the high street giving them an ability to read the information on the magnetic stripe, the technology is everywhere. Although some limitations exist on the amount of information that can be stored on the stripe and the security of the data, solutions from various vendors exist to solve these problems.

With the advent of new technologies, many people have predicted the demise of the magnetic stripe. However, with the investment in the current infrastructure, this is not likely to be any time soon. Magnetic stripe technology provides the ideal solution because it is very inexpensive and readily adaptable to many functions, meaning that it will still be used for many years to come.

3.2.2 Smart Cards

Smart cards are not new; they were invented in 1974 by French journalist Roland Moreno [1] and first used as prepaid phone cards in 1984. Smart cards are credit card–sized pieces of plastic containing a data storage system. The technology was rapidly accepted in Europe because the high cost of telecommunications made on-line verification of transactions very expensive. The smart card provided the mechanism to move that verification off-line, reducing the cost without sacrificing any of the security.

The first plastic cards were used in the United States for club membership, as calling cards after that, and finally as credit cards. Several terms are used to identify cards with integrated circuits embedded in them. The terms *chip card, integrated circuit card,* and *smart card* really all refer to the same thing.

There are two types of smart card. The first type only contains memory and is used to store information; examples of this might include stored value cards in which the memory stores a dollar value that the user can spend in a variety of transactions, such as in pay phones, retail bills, or vending machines.

The second type of card is a true smart card in which a microprocessor is embedded in the card along with memory. Now the card actually has the ability to make decisions about the data stored on the card. The card is not dependent on the unit to which it is attached to make the application work. Because the card contains a microprocessor, various methods can be used to prevent access to the information on the card, providing a secure environment. This security has been touted as the main reason that smart cards will replace other card technologies.

The microprocessor type smart card comes in two types: the *contact version* and the *contactless version*. Both types of cards are embedded with microprocessors; however, the contactless version does not have the gold-plated contacts visible on the card. The contactless card uses a technology to pass data between the card and the reader without any physical contact being made. The advantage to this contactless system is that there are no contacts to wear out, no chance of an electric shock coming through the contacts and destroying the integrated circuit, and the knowledge that the components are completely embedded in the plastic with no external connections. The disadvantage to this is that the card and reader are more complex and hence are more expensive.

The biggest disadvantage today with smart cards is the cost to create a smart-card system. Individual card prices have fallen over the past few years, but they are still high when compared with a magnetic stripe card. The biggest advantage is, of course, the amount of data that can be stored and the security that can be built into the card.

3.2.3 Optical Cards

Optical memory cards use a technology similar to the one used for music CDs or CD-ROMs. A panel of the gold-colored laser sensitive material is laminated in the card and is used to store the information. The material is comprised of several layers that react when a laser light is directed at them. The laser burns a tiny hole (2.25 μm in diameter) in the material, which can then be sensed by a low-power laser during the read cycle. The presence or absence of the burn spot indicates a 1 or a 0.

Because the material is actually burned during the write cycle, the media is a write-once, read-many (WORM) media, and the data is nonvolatile (not lost when power is removed). The optical card can currently store between 4 and 6.6 MB of data which gives the ability to store graphical images such as photographs, logos, fingerprints, X-rays, and so forth.

The major disadvantage of the optical card is the fact that it is a write-once technology, so the amount of data storage available is used up with every piece of new data written. In some applications, this can be considered an advantage because it maintains the complete history of changes made to the card.

3.3 Radio Frequency Identification

3.3.1 RFID Historic Background

The beginning of the RFID was in October 1948 after the paper by Harry Stockman, "Communications by Means of Reflected Power." The popular system Identification of Friend of Foe (IFF) for aircraft was one of the first applications of the RFID [2]. In early 1940s, the British Royal Air Force outfitted airplanes with radio transponders that would respond when interrogated. This allowed pilots and ground crews to distinguish the RAF airplanes from the Luftwaffe's airplanes, which proved to be a decisive advantage in the Battle of Britain.

In 1960s, the electromagnetic theory related to the RFID application was developed, and this was the prelude to the RFID explosion. Commercial activities exploiting the RFID began also during the 1960s and the Electronic Article Surveillance (EAS) application is one example. The EAS is a simple 1-bit tag, because only the presence or the absence of a tag can be detected.

During the rapid development of microelectronic technology during the 1970s, companies, universities, and government laboratories were actively engaged in the development of practical applications of RFID, such as animal tracking, vehicle tracking, and factory automation. The 1980s was the decade for mass deployment of RFID technology. The interest in the United States was mainly for transportation and access control whereas in Europe the greatest interests were for animal tagging, industrial applications, and toll roads.

Since the 1990s many technological developments have been dramatically expanding the functionality of the RFID. Advances in microelectronics, embedded software, and microwave circuit integration are opening the door to new wireless system and expanding the application field of RFID[1].

Benefits of RFID technology are that it allows manufacturers, retailers, and suppliers to efficiently collect, manage, distribute, and store information on inventory, business processes, and security controls. RFID will allow retailers to identify potential delays and shortages, grocery stores to eliminate or reduce item spoilage, toll systems to identify and collect auto tolls on roadways, and suppliers to track shipments and in the case of critical materials, RFID will allow receiving authorities to verify the security and authentication of shipped items.

These uses are seen as only the beginning, and as RFID is deployed across different sectors and services, increasing efficiency and visibility, we can already see other applications and benefits, for example, in the medical and health care field.

3.3.2 RFID System Overview

In general terms, RFID represents a way of identifying objects or people using radio waves. Identification is possible by means of unique numbers that identify objects, people, and information, stored on microchips, which can be read automatically, unlike bar codes, which need to be scanned manually. With recent advancements in

1. A search of the U.S. Patent Office alone will reveal over 350 patents related to RFID and its use and that number is still growing.

the technology, the automatic identification data capture industry is accelerating its efforts to identify new applications to take advantage of RFID.

RFID is fundamentally based on wireless communication, making use of radio waves, which form part of the electromagnetic spectrum; it is not unlike two other wireless technologies, Wi-Fi and Bluetooth. The three technologies are all designed for very different uses and therefore have different functionalities, but there is shared ground among them, with some hybrids starting to appear. RFID systems can utilize both Wi-Fi and Bluetooth (and potentially other technologies, like UWB, for example) and need not see them as competitors.

All RFID systems are composed of three main components:

- *RFID tag*, or transponder, which is located on the object to be identified and is the data carrier in the RFID system;
- *RFID reader*, or transceiver, which may be able to both read data from and write data to a transponder;
- *Data processing subsystem,* which utilizes, in some useful manner, the data obtained from the transceiver.

The essential requirement in an RFID system is to transfer data stored in a tag, to a reader, across a wireless air interface (the region between the tag and the reader). A two-way communication process is required to do this and requires a radio carrier signal suitably modified (modulated) to carry the data. The concept is depicted in Figure 3.2.

The RFID concept works as follows: a reader transmits a signal that is received by an antenna integrated with a small RF chip. In general, the chip is activated only when an RFID reader scans it. When the chip wakes up, it sends the unique identifier number, which the reader passes along to applications such as inventory control and shipping. It is this level of application logic that provides the selection of particular tag and data manipulation criteria for those identified tag(s).

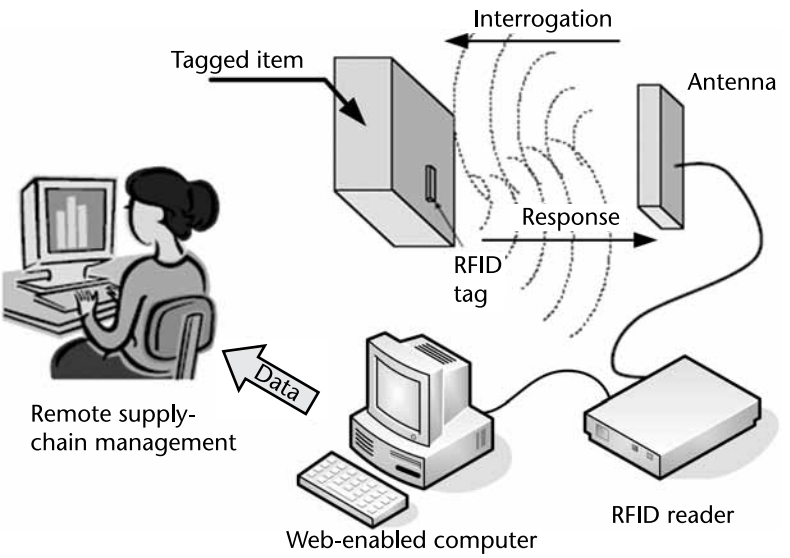


Figure 3.2 Basic components of RFID systems.

Using a host computer application program, specific tags can be selected to be identified, and the data contained within those tags to be acted upon while those tags are active within the *RF portal* (interrogation area). The RF portal is defined as the area where RFID tags can be read or written to; it can be stationary or mobile.

Stationary portals are used mainly in applications where the item containing the RFID tag has to follow some prescribed physical path or flow. An example of this would be a location in which warehouse goods flow through a dock door or with items that travel down a conveyor or assembly line.

Portable or *mobile interrogators* are used in applications when the tagged items do not follow a predefined path. This type of RFID interrogator can be used in conjunction with a portable computing device, such as a portable data terminal or mobile pen computer. Typical applications include asset tracking, picking or moving inventory, inspection, and quality control. In portable or mobile applications, the interrogator is aimed into a certain physical area in which RFID tags need to be scanned. It is more flexible in application scope since the user will define the RF portal by orienting the longitudinal and latitudinal axis of the interrogator. However, because this device operates on batteries, it is limited in its effective range of scanning or RF portal depth of field. The energized period of operation of this type of interrogator must also be controlled so that battery life can be maximized.

There are many potential applications for RFID; the most obvious one is as a more robust replacement to bar codes. However, innovative companies are regularly finding new application for the enhanced range, capacity, and read/write capability. The RFID readers can be big enough for forklifts to pass through or small enough to fit on retail shelves.

RFID operates in license-exempt spectrum space, but the exact frequencies that constitute license-exempt bands may vary depending on the regulations in different countries. Typical carrier frequencies (reader's transmitting frequency) in today's applications range from 125 kHz to 2.45 GHz (with 5.8 GHz also being considered). The frequency bands must be selected carefully for applications because each one has its own advantages and disadvantages.

The RFID systems themselves can achieve high levels of complexity having incorporated memory, data processing capabilities that include communication encryption, and protocols (Figure 3.3). As we will see later, the air interface (communication between the reader and the tag), physical interrogator, data protocol processor, and application commands and interfaces are all covered by the different, but related, standards.

A high-level description of the sequence of communication is as follows:

1. The host manages reader(s) and issues commands, after which the reader and the tag communicate via RF signal.
2. The reader continuously generates an RF carrier sine wave, watching always for modulation to occur. Detected modulation of the field would indicate the presence of a tag.
3. The carrier signal is sent out through the antennas.
4. Carrier signal hits tag(s) (Figure 3.4).
5. Once the tag has received sufficient energy to operate correctly, it divides down the carrier and begins clocking its data to an output transistor, which is normally connected across the coil inputs.

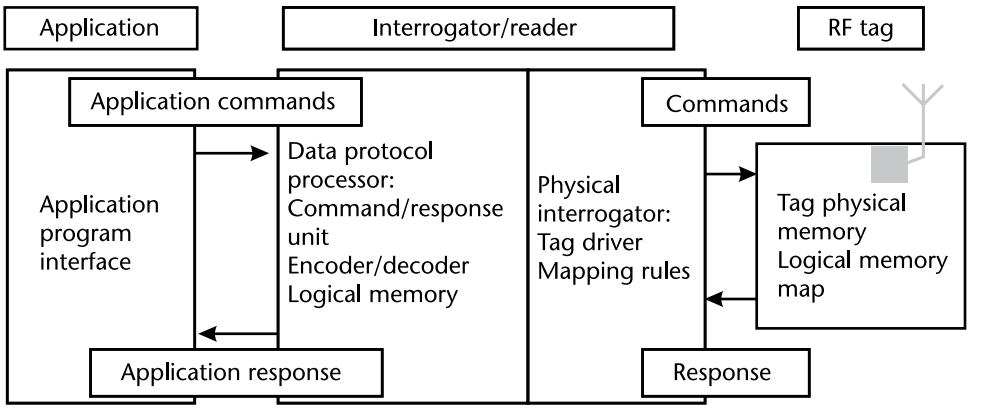


Figure 3.3 RFID system overview.

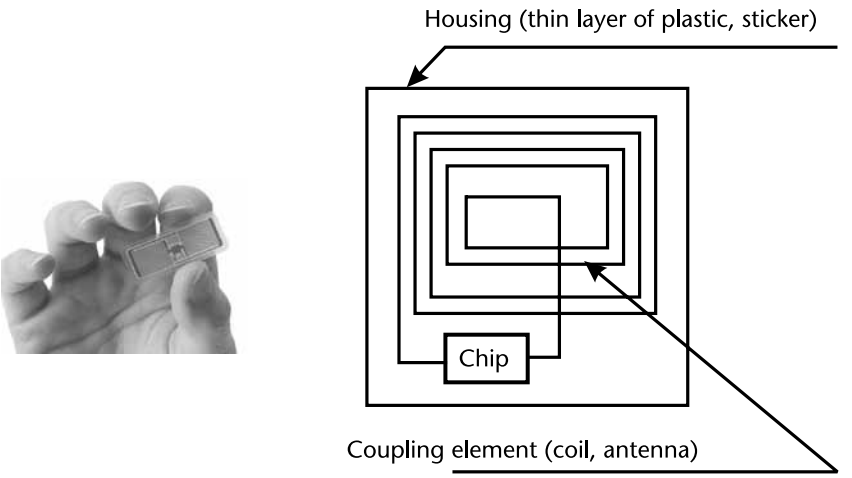


Figure 3.4 RFID tag.

6. Tag receives and modifies carrier signal and sends back modulated signal (*passive backscatter*, which FCC and ITU refer to as a field disturbance device). The tag's output transistor shunts the coil, sequentially corresponding to the data that are being clocked out of the memory array. Inductive coupling and load modulation are used at lower frequencies, as opposed to systems operating at 2.45 GHz and higher bands, where true RF communication links and backscatter principles are used.
7. Shunting the coil causes a momentary fluctuation (dampening) of the carrier wave, which is seen as a slight change in amplitude of the carrier.
8. Antennas receive the modulated signal and send them to the reader.
9. The reader decodes the data. The reader peak-detects the amplitude-modulated data and processes the resulting bit stream according to the encoding and data modulation methods used.
10. The results are returned to the host application.

The general requirements for the tag antenna are small size, high efficiency, and sufficient beamwidth for the reduction of orientation sensitivity. Tags also require simplicity and low manufacturing costs.

3.3.3 Principles of RFID Operation

3.3.3.1 Near-Field and Far-Field Operations

The coupling between tag and reader is achieved in one of two ways, depending upon the carrier frequency used and the system and antenna design (Figure 3.5).

Low- (<135 kHz) and high-frequency (typically 13.56 MHz) systems invariably use reactive (typically inductive) coupling, wherein the predominantly magnetic field component carries the data in the communication between tag and reader, in much the same way as coupling between primary and secondary coils in an air-cored transformer. In these systems the field is effectively tied to its source, and the field that couples with the tag is modulated by means of the tag circuitry, such that the data-related changes can be sensed by the reader.

The second form of coupling is by propagation of the electromagnetic field used to read or interrogate the tag. In these systems field components dissociate from their source in the reader and propagate into free space. The components of an RFID system that largely determine whether an RFID system couples by inductive or propagation means are the antenna and the manner in which it is driven in electrical terms. The term *antenna* we can only use when describing the propagation coupling. Inductive coupling systems do not contain real antenna, just inductively coupled coils.

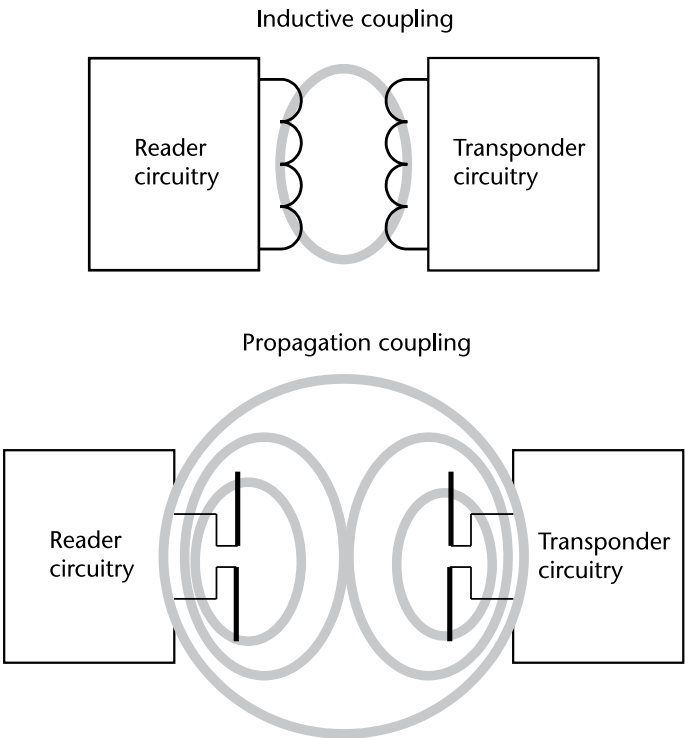


Figure 3.5 Inductive and propagation coupling RFID systems.

For low- and high-frequency RFID systems, the coils are structured and driven in such a way that the propagation component is small or even nonexistent, while the reactive component is predominant. At ultrahigh frequencies and above, the systems are essentially structured and driven to operate in the propagation mode, the antenna dimensions being matched or appropriately related to the carrier wavelength to achieve the desired result. In these systems the reactive component is designed to be suitably small.

The far field begins where the near field ends, although not abruptly, at the certain distance from the transmitting antenna. In the near field, the tag-to-reader communication is achieved via load modulation. Load modulation is achieved by modulating the impedance of the tag as seen by the reader. In the far field, the tag-to-reader communication is achieved via backscatter. *Backscatter* is achieved by modulating the radar cross section of the tag antenna.

3.3.3.2 Inductive Coupling and the Load Modulation

When a tag is placed within the alternating magnetic field created by the reader, it draws energy from the magnetic field. This additional power consumption can be measured remotely as a voltage perturbation at the internal impedance of the reader coil. The periodic switching on/off of a load resistance at the tag therefore effects voltage changes at the reader's coil and thus has the effect of an amplitude modulation of the coil voltage by the remote tag. If the switching on/off of the load resistance is controlled by the tag's stored data stream, then this data is transferred from the tag to the reader. This type of data transfer is called *load modulation*.

In load modulation the carrier signal is modulated by switching impedance from a matched condition to an unmatched condition to alter the reflection coefficient. The data transfer from the transponder to the reader can be achieved in three different ways: load modulation, load modulation by using a subcarrier (frequencies below 30 MHz), or load modulation using subharmonic procedure (above 100 MHz).

The incident radio electromagnetic wave is received by the tag through the coil. The radiated energy is then converted to electrical current and travels down a transmission line configuration with *intrinsic impedance* Z_0 (determined by the material properties of the transmission line) to *load impedance,* Z_L. In a transmission line theory, as already mentioned in more detail in Chapter 2, three special cases are possible:

- Matched load, $Z_L = Z_0$, no reflection;
- Open load, $Z_L = \infty$, full in-phase reflection;
- Shorted load, $Z_L = 0$, full out-of-phase reflection.

At the end of the transmission line, the electric waveform reaches a PIN diode, represented as a switch in Figure 3.6, and is used to toggle between two types of load impedances [3]. When the diode is forward biased, the current is allowed to flow through the matched load making $Z_L = Z_0$, and thus causes the reflection coefficient to equal zero. In the first case, all of the forward-traveling current is absorbed by the load and no power is sent backwards through the transmission line. When no power is transmitted, the tag transmits a logical bit 0.

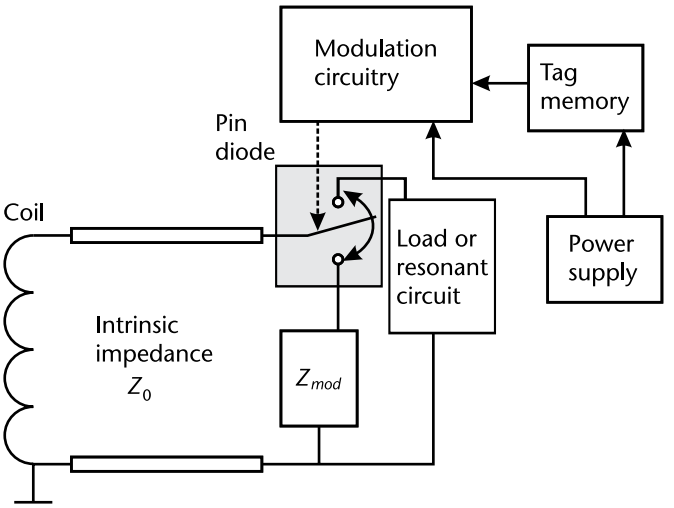


Figure 3.6 Load modulation circuitry.

When the diode is reverse-biased, the load impedance essentially becomes infinite and makes the coefficient of reflection equal to 1. In this case, no power is absorbed, and all of it is reflected back down the line. When all the incident power is reflected, the tag transmits a logical bit 1.

These bits are propagated through the modulated backscatter, and ride on top of the reflected wave to the receiver. There are several methods to encode the data onto the carrier wave. The process of load modulation creates amplitude-modulated sidebands symmetrically placed around the 13.56-MHz interrogation carrier frequency.

Because the coupling between reader antenna and tag is relatively weak and the voltage change created by the tag leads to relatively poor signal-to-noise ratios, reply code modulation with a subcarrier is utilized in most RFID chips. In this improved signaling method, the tag's data reply information is contained in a pair of backscattered sidebands, which are subsequently demodulated in the RF and baseband signal processing sections of the reader to recover the tag's data stream.

In the *subharmonic procedure*, a second frequency (which is usually lower by a factor of 2) is derived by a digital division by 2 of the reader's transmission frequency. The output signal of a binary divider can be modulated with the data stream from the transponder. One popular operating frequency for subharmonic systems is 128 kHz. This gives rise to a transponder response frequency of 64 kHz.

3.3.3.3 Propagation Coupling and the Backscatter Modulation

The term *backscatter modulation* refers to the communication method used by a passive RFID tag to send data back to the reader. By repeatedly shunting the tag coil through a transistor, the tag can cause slight fluctuations in the reader's RF carrier amplitude.

The RF link behaves essentially as a transformer; as the secondary winding (tag coil) is momentarily shunted, the primary winding (reader coil) experiences a momentary voltage drop. The reader must peak-detect this data at about 60 dB down

(about 100 mV riding on a 100-V sine wave). This amplitude modulation loading of the reader's transmitted field provides a communication path back to the reader. The data bits can then be encoded or further modulated in a number of ways.

We know from the field of radar technology that electromagnetic waves are reflected by objects with dimensions greater than around half the wavelength of the wave. The efficiency with which an object reflects electromagnetic waves is described by its reflection cross section. Objects that are in resonance with the wavefront that hits them, as is the case for antenna at the appropriate frequency, for example, have a particularly large reflection cross section.

Figure 3.7 shows a block diagram of a typical passive RFID reader/tag configuration. The reader radiates an unmodulated signal (called the incident wave), which impinges on the tag. The tag antenna intercepts this signal, absorbs part of it, and reradiates the rest. A switch [field-effect transistor (FET)] is connected across the antenna, which, when closed, causes a mismatch in the tag antenna. The mismatch causes a small percentage (say, 10%) of the signal to be absorbed, which, in turn, results in the other 90% being reradiated. The switch, also known as a modulator, is controlled by the tag output data stream and causes the tag data to be modulated onto the incident wave and be reradiated (or backscattered) as a modulated signal. This technique is known as *impedance modulated backscatter* or *backscatter modulation* for short.

The amount of energy intercepted and reradiated is determined by the differential radar cross section of the tag, which, in turn, is a function of the tag antenna aperture and modulation depth.

The operating range of the tag is determined by the reader's transmitted power, the reader's receiver sensitivity, the path loss in both directions between the reader and tag, and the differential radar cross section of the tag. The backscattered signal received at the reader receiver decreases as the fourth power of the distance between the reader and tag (the square of the power in each direction). For example, a typical 900-MHz tag will have a radar cross section of at least $0.005m^2$ and typically $0.024m^2$.

Present-day receiver technology allows a reader to reliably decode a signal having a strength of –80 dBm or higher. This means that the signal arriving back at the reader from a tag must be stronger than –80 dBm.

The ratio of power transmitted by the reader (incident wave) and power returning from the transponder (backscatter wave) can be estimated using the radar equation, a detail of which can be found in Chapter 5.

The reader's RF transceiver block diagram shown in Figure 3.7 uses the homodyne topology and represents the classic approach for backscatter radar where received signals are close in frequency to the transmitted carrier. A homodyne receiver performs a direct conversion of the received RF signals to a zero intermediate frequency (IF) baseband. Elimination of the IF reduces the need to perform image rejection filtering as is typical with other receiver approaches, such as the *superheterodyne* approach. This architecture uses two high compression point MMIC mixers in quadrature along with lowpass filtering.

Reducing continuous wave (CW) carrier leakage into the RF port of the mixers is critical to receiver performance since the resulting phase discriminator generates unwanted dc signal offsets. Leakage of the high-power transmitting signal into

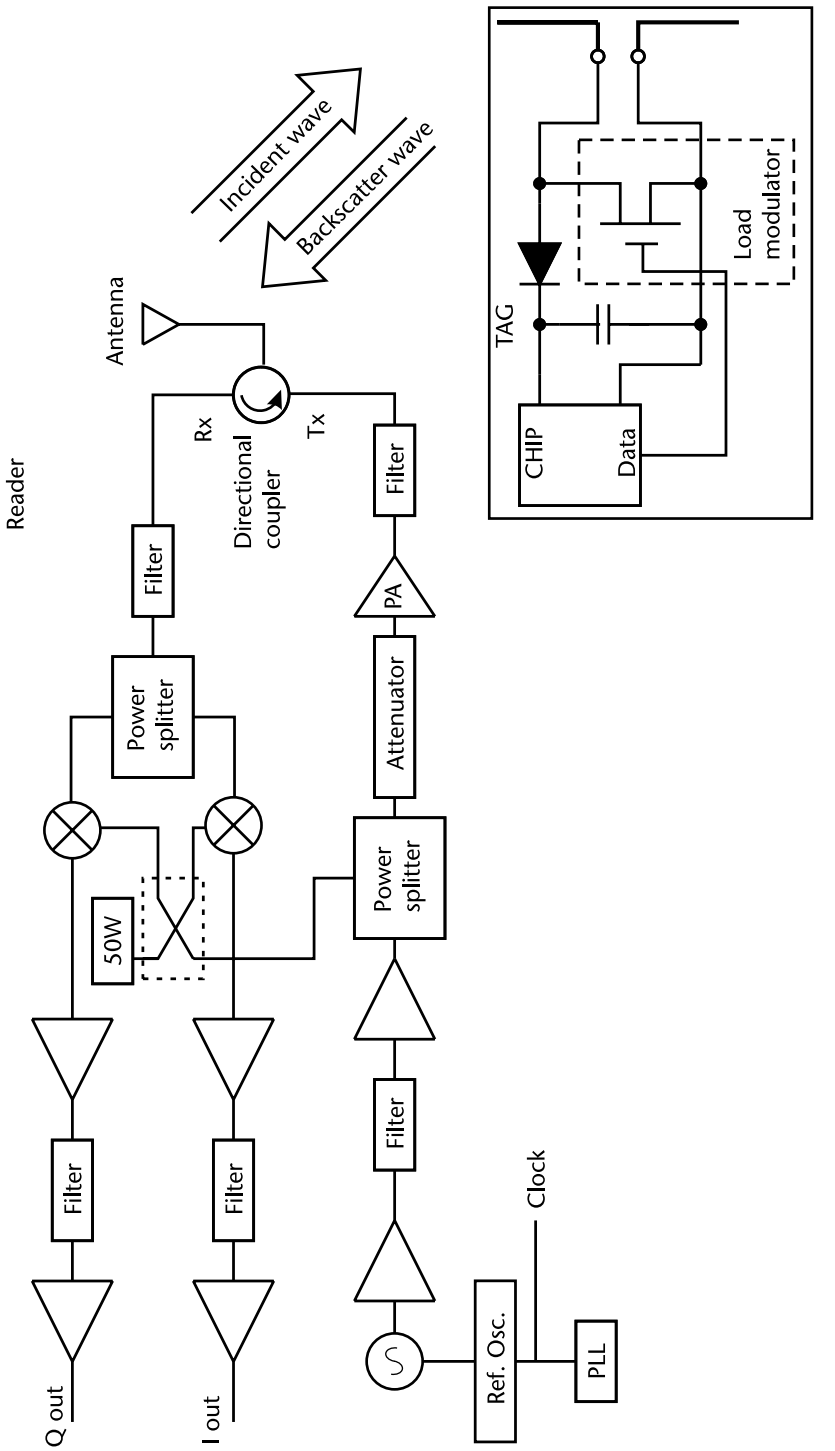


Figure 3.7 Backscatter modulation circuitry.

the receive path can saturate the receiver and significantly degrade receiver signal sensitivity.

The power is supplied to the tag's antenna connections as HF voltage, and after rectification by the diodes, it can be used as turn-on voltage for the deactivation or activation of the power saving power-down mode. The diodes used here are low-barrier Schottky diodes, which have a particularly low threshold voltage. The voltage obtained may also be sufficient to serve as a power supply for short ranges.

Functions performed by the reader may include quite sophisticated signal conditioning and parity error checking and correction. Once the signal from a transponder has been correctly received and decoded, algorithms may be applied to decide whether the signal is a repeat transmission and may then instruct the transponder to cease transmitting. This is known as the *Command Response Protocol* and is used to circumvent the problem of reading multiple tags, in a short period of time. Using interrogators in this way is sometimes referred to as *hands-down polling*.

A more secure but slower tag polling alternative is a technique called *hands-up polling*, which involves the interrogator looking for tags with specific identities and interrogating them in turn. This is contention management, and a variety of techniques have been developed to improve the process of batch (multiple tags) reading.

3.3.4 The Electronic Product Code System

Electronic Product Code (EPC) is an item numbering and networking concept that emerged from research undertaken at the Massachusetts Institute of Technology (MIT) Auto-ID Center, and associated centers, in which RFID data carriers were identified as the method of choice for carrying the EPC numbers [4].

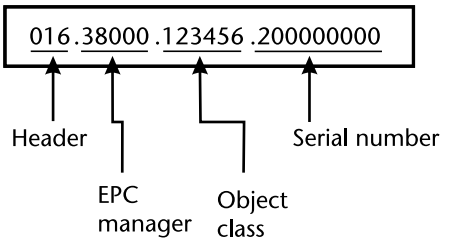
3.3.4.1 Electronic Product Code

The RFID technology itself offers several improvements over its predecessor technologies: the bar code and magnetic stripe cards. The central data feature of RFID technology is the electronic product code, which is viewed by many in the industry as the next generation bar code or the Universal Product Code (UPC). EPC code is a 96-bit code created by the Auto-ID Center that would one day replace bar codes (Figure 3.8).

The EPC has digits to identify the manufacturer, product category, and individual item. It is backed by the United Code Council and EAN International, the two main bodies that oversee bar code standards. The EPC code can carry more data than the UPC code and can be reprogrammed with new information if necessary. Like the UPC, the EPC code consists of a series of numbers that identify the manufacturer and product type. The EPC code also includes an extra set of digits to identify unique items.

The EPC numbering scheme, initially composed of a 96-bit code, is structured as follows:

- An 8-bit header (also known as a EPC version number);



Header: identifies the length, type, structure, version, and generation of EPC
Manager number: identifies the company
Object Class: similar to a stockkeeping unit or SKU
Serial number: specific instance of the Object Class being tagged

Figure 3.8 Electronic Product Code.

- *EPC manager*, 28 bits to facilitate identification of the item manufacturer or source provider, for example;
- *Object class identifier*, 24 bits to facilitate the identification of the type of product, such as a specific stock-keeping-unit;
- *Serial number*, 36 bits to facilitate the unique identification of an individual item.

The header is used to distinguish multiple EPC formats, thus allowing designation of differing bit lengths tags as the technology matures (a 256-bit version has been proposed). The header can also be used to distinguish bit-length field variations to those indicated above with respect to manufacturer, product, and serial number support, thus allowing longer and more manufacturer identifiers for organization with a small number of product types and serial number requirements.

While the initial bit-length designation for EPC was 96 bits, a shorter 64-bit version has been introduced on an interim basis to help facilitate the realization of lower-cost RFID data carrier devices. This seems reasonable on the basis that the full identification capability of the 96-bit version would not be required for some time.

3.3.4.2 Object Naming Service

EPC Object Naming Service (ONS) was introduced to provide a directory service capable of supporting the linkage of EPC numbers with additional data or information concerning the item to which the EPC tag is attached. These additional item-associated data or information may be stored on a server connected to a local network or the Internet. The ONS is analogous to the domain name service used for location of information on the Internet.

3.3.4.3 Physical Markup Language

Physical Markup Language (PML) is structured to allow the information about an item or object to be appropriately specified. The PML is based upon the popular XML metadata language, its syntax, and semantics to be administered

and developed by the governing body (EPCglobal) in conjunction with the user community.

Product definitions within this language markup facility, which began with food items, require the ongoing efforts of the governing body to build a sufficiently inclusive directory. Product descriptions already undertaken by standards bodies, such as the International Bureau of Weights and Measures and the National Institute of Standards and Technology (NIST), are being seen as valuable sources of information in this respect.

In addition to fixed product information, the PML will also accommodate dynamic quantities, such as temperature, humidity or vibration, which may change as a result of some local, environmental, or intrinsic effect, including changes over time (temporal effects). This adds a further dimension to the data gathering and handling processes and, when presented in a PML file, may offer innovative opportunities for process enhancement. For example, condition status information derived dynamically could be used to automatically determine product pricing.

3.3.4.4 Savant Software

Savant is the specification for standard RFID middleware, that is, software that bridges RFID hardware and enterprise applications. It defines an EPC events handling framework and is thus the primary means of data gathering for any RFID deployment. It acts as the central nervous system of the EPCglobal Network [5].

Savant's most basic function is to receive the EPC number and direct a query over the Internet or other established network to the ONS which then returns an address at which the item information is stored. The information is available to, and can be augmented by, Savant systems within the network, ostensibly around the world.

The very high data handling envisaged within the EPC infrastructure indicates the potential need for companies to maintain ONS servers locally to support rapid retrieval of information. The Savant software is being developed to use a distributed architecture with a hierarchical structure to manage data flow. A highly extensive network of Savants is envisaged to support the EPC data management, with Savant platforms running, for example, in factories, stores, distribution centers, and regional support facilities, and even on mobile platforms such as container trucks and cargo planes. Creating such an infrastructure is one of the biggest challenges to realizing the EPC support objectives.

3.3.5 RFID and Biometrics

Biometric technology is the use of human bodily characteristics or "physiological autographs" in an attempt to uniquely and absolutely identify individuals. The earlier forms of unique body characteristics were recognized in the science of fingerprints in the 1970s. In the 1980s, the Automated Fingerprint Identification System (AFIS), developed by NEC Technologies completely changed the role of fingerprints. It combined computer graphics with special software programs and parallel processing to create forensic results.

Today biometric technologies include retina prints, iris prints, signature and handwriting analysis, palm prints and hand geometry, voiceprints, face recognition,

facial thermograms, silhouette identification and gait prints, and even specific task performance and writing styles.

Biometrics is widely used in fields as varied as e-commerce, network access, time and attendance, ATMs, corrections, banking, and medical record access. Due to the apparent ease of use and other factors, biometric technology applications are being used increasingly throughout private businesses and governmental sectors. Of all the mentioned biometric identification systems, iris prints appear to be the most accurate. The iris patterns of each person's eyes are fixed before birth and remain unchanged throughout one's life unless trauma interferes.

Although phenomenal growth in both smart card and biometric technologies has been witnessed, another area of more recent and rapid growth is the merging of these and many other technical elements into the field of RFID. Major initiatives by the United States and other governments aim to combine RFID and biometric technologies in a new generation of identity cards. Together, RFID and biometric technologies promise to reduce fraud, ease identity checks, and enhance security.

As part of its US-VISIT program, the U.S. government has mandated adoption of biometrically-enabled passports by the 27 nations in its Visa-Waiver Program (VWP), among them Japan, most of the nations of Western Europe, and a handful of others [6]. Soon, all passports produced in the United States will carry biometric information. These passports are based on guidelines issued by the International Civil Aviation Organization (ICAO), a body run by the United Nations with a mandate for setting international passport standards.

The ICAO guidelines, detailed in ICAO Document 9303, call for the incorporation of RFID chips into passports. Such chips are present in initial deployments of biometrically enabled U.S. passports and in the biometrically enabled passports of other nations as well. Next generation passports, sometimes called *e-passports*, will be a widespread form of identification within a couple of years. E-passports will contain digitized photographic images of the faces of their bearers.

The standard additionally specifies fingerprints and iris data as optional biometrics, and the goal is the strong authentication through documents that unequivocally identify their bearers. Data integrity and physical integrity are vital to the security of ID cards as authenticators. For authorities to establish someone's identity with certainty, for example, the passport must carry a photograph of irrefutable pedigree, with a guarantee that no substitution or tampering has taken place. Without this guarantee, passports can be forged, enabling unauthorized persons to enter a country.

Strong authentication requires more than resistance to tampering. Data confidentiality, that is, secrecy of data stored on ID cards, is also critical. Protecting biometric and biographical data is essential to the value and integrity of an authentication system. In particular, data secrecy affords an important form of protection against forgery and spoofing attacks. Therefore, protecting e-passport data against unauthorized access is a crucial part of the security of the entire system. For a full review on the work leading to these decisions, see [7].

Confidentiality protection for stored data is important for other reasons as well. Both RFID and biometrics are highly privacy-sensitive technologies. Sensitive data, such as birth date or nationality, are carried on passports. The privacy, physical safety, and psychological comfort of the users of next generation passports and

ID cards will depend on the quality of data-protection mechanisms and supporting architecture.

3.3.6 Challenges of RFID Implementation

It has to be emphasized that the implementation of the technology itself is not a difficult exercise; however, to gain the full benefits of implementing RFID within an enterprise, a more holistic view needs to be taken. A large volume of data is created by implementing RFID, which has to be turned into information and intelligence. At the present time, enterprise resource planning technology suppliers are developing extensions to their products to work with RFID systems [8]. The following list represents the potential challenges to be considered when implementing an RFID solution. Some of those challenges are discussed in more detail in this chapter.

Business Issues

- No proven return on investment;
- Cost of initial implementation;
- Data sharing between supply chain partners;
- Intellectual property issues;
- Environmental (disposal) issues;
- Consumer privacy objections;
- Lack of organizational expertise;
- Lack of historic data.

Technology Issues

- Technology standards and interoperability;
- Reliability and maturity of technology;
- Data integration and evolving middleware;
- Environmental issues (temperature, moisture);
- Spectrum congestion and frequency availability;
- Security of data on tags and readers;
- Accuracy of tag reading;
- Volume of data produced.

Large Volumes of Data

Readers can scan each RFID tag several times per second, which generates a high volume of raw data. Although the data are redundant and discarded at the reader level, processing large volumes of data can be difficult.

Operational Speed

The RFID system must provide accurate reads at all levels, item, case, and pallet, without requiring any reduction in throughput.

Product Information Maintenance
When the reader processes high volumes of RFID tags, the attributes of each tagged product must be continually retrieved from a central product catalog database, a process that results in challenges for large-scale implementations.

Configuration and Management of Readers and Devices
When a large number of readers and related hardware devices are deployed across multiple facilities, configuration and management can be challenging. The implementation of automated devices for these processes is essential.

Data Integration Across Multiple Facilities
In an enterprise with multiple facilities that are geographically distributed, it is increasingly difficult to manage data in real time and instantaneously aggregate it into the central IT facility, a process that can place a significant burden on the network infrastructure.

Data Ownership and Partner Data Integration
When there are different companies involved in business processes, such as the retail supply chain, it can create issues pertaining to the ownership and integration of the data, thereby compromising the integrity of the solution architecture.

Data Security and Privacy
Depending on the nature of the business application and the solution scenario, security and privacy challenges could have a significant impact on the architecture.

Cost
At an average cost of around 20 to 30 cents apiece, RFID tags are still too costly, especially for retail applications and certainly for use on inexpensive and low-margin products, such as a 50-cent candy bar or a $1 bar of soap. This is a key reason why mass-market consumer retail businesses operating on very thin profit margins have been slow to adopt RFID-based smart shelf and smart checkout technology. RFID tag developers are working to lower the cost of tags to 10 cents, or even 5 cents, over the next few years.

Materials
RFID signals are easily blocked. Over short ranges, these signals can be attenuated by certain materials (the most common is packing made from metallic substances). Over longer ranges, the signals, which are much weaker than commercial radio broadcast signals, can be blocked by common objects, including the human body. Researchers are working to solve this problem by using novel designs for tag antennas and more sensitive reader arrays.

Tag Form Factor/Size
Full flexibility in tag sizing is required in order to accommodate the smallest items, as well as cases and pallets, with the same high level of reliability and performance for all tags. In addition, tags must be able to be rigid as well as flexible, for example,

to accommodate the curve of a pill bottle. And regardless of how tiny or flexible the tag must be, read-range requirements must still be met.

Tag Proximity and Orientation

RFID tags and readers are orientation dependent, so tags must be positioned properly relative to readers in order for the antenna coils to exchange signals. The solution to this problem will come with the development of multiple-reader systems that use an array of readers positioned to cover all the possible orientations for tagged items that might, for example, be found in a display bin in a store. Part of this solution will involve protocols to coordinate the operation of these reader arrays.

Environmental Noise

Because RFID equipment may be operating in the environment that generates electromagnetic energy (for example, cordless phones, mobile radios, fluorescent lighting, electrical equipment, and other RFID readers), the RFID system selected must be able to reject the interference from these products in order to ensure predictable and reliable system performance.

Accuracy of Tag Reading

RFID readers often experience *false negatives* and *false positives*. A false negative occurs when a valid tag passes within the prescribed range of an RFID reader, but the reader does not read the tag. This can happen for many reasons, including when a case tag is buried deep inside a pallet, when reader signals are blocked or absorbed by substances such as metal or water, or when a case tag is not oriented properly (tag reads are more successful when tags are perpendicular to reader signals).

A false positive occurs when a tag accidentally passes within range of an RFID reader but was not intended to be read. False negatives and false positives often occur with closely packed items where multiple tags in close proximity shadow each other.

Competing Technical Standards

Competing standards prevent the universal adoption of RFID readers and tags. Different manufacturers are developing tag protocols that operate at different frequencies, with a variety of packet. Ideally, a single standard should be adopted to make all tags compatible with all readers. Both the cost and standardization challenges are being addressed by individual companies and by the Auto-ID Center and the International Organization for Standardization (ISO), industry consortia working to set standards for RFID tags.

3.4 Wireless Sensor Networks

3.4.1 Basics of Wireless Sensor Networks

Wireless sensor network (WSN) is the generic name for the technology consisting of a broad range of devices. Basically, any collection of devices equipped with a processor, having sensing and communication capabilities that are able to organize them into a network created in an ad hoc manner falls into this category.

Research in the field of WSNs has increased tremendously during the last few years, and the initial projects have shown that building cheap smart sensors that can network is possible and the addition of the wireless communication capabilities to sensors increased their functionality dramatically. WSNs bring monitoring capabilities that will forever change the way in which data is collected from the ambient environment [9].

The field of sensor networks is a relatively new one. Scientists from various communities approached this research area with enthusiasm and brought together knowledge from the various domains of computer science, electrical engineering, telecommunications, radiocommunications, and so forth. The initial directions of research were specific to each of these fields, everyone trying to adapt their knowledge to make WSNs a reality.

Let us take, for example, the traditional monitoring approach of a remote location for a given phenomenon, such as recording the geological activity, monitoring the chemical or biological properties of a region, or even monitoring the weather at a certain place. The old approach was to build rather big and robust devices. Besides the sensor pack itself, these devices contained a big power supply and local data-storage capabilities. A team of technicians would have to travel together to the destination being monitored to place these expensive devices at predefined positions and calibrate all the sensors. Then they would come back after a certain amount of time in order to collect the sensed data. If, by misfortune, some hardware failed, then nothing could be done about it and the information about the phenomenon itself would be lost.

The new approach is to construct inexpensive, small sized, energy-efficient sensing devices. Because hundreds or thousands of these devices will be deployed, the reliability constraints for them will diminish. No local data storage is needed anymore because they will process locally and then transmit wirelessly the observed characteristic of the phenomenon to one or more access points connected to a computer network. The individual calibration of each sensor node is no longer needed because it can be performed by localized algorithms. The deployment will also be easier, by randomly placing the nodes (e.g., simply throwing them from a plane) onto the monitored region.

Wireless sensor networks are one of the most important tools of the new era of computing. They are the simplest intelligent devices around, having as their main purpose monitoring the environment surrounding us and alerting us of the main events happening. Based on the observation reported by these instruments, humans and machines can make decisions and act on them.

3.4.2 Applications of Wireless Sensor Networks

At this moment a large variety of sensors exists and sensors have been developed to monitor almost every aspect of the ambient world: lighting conditions, temperature, humidity, pressure, the presence or absence of various chemical or biological products, detection of presence and movement, and so forth.

The sensor networks field is rapidly evolving, and although a large number of sensor network prototypes exist at this moment, the possible new application areas are still being explored. The typical application that one can think of

has as the main goal some sort of monitoring, with the most common one being environmental monitoring. Some of the potential applications are listed as follows.

3.4.2.1 Intelligent Warehouses

Each item contained inside the warehouse will have a sensor tag attached that will be monitored by the sensor nodes embedded into the walls and shelves. Based on the read data, knowledge of the spatial positioning of the sensors, and time information, the sensor network will offer information about the traffic of goods inside the building, create automatic inventories, and even perform long-term correlations between the read data. There will be no need for manual product scanning.

3.4.2.2 Environmental Monitoring

This is the widest area of applications envisioned up to now; a particular application in this category is disaster monitoring. The sensor nodes deployed in the affected areas can help humans estimate the effects of the disaster, build maps of the safe areas, and direct the human actions towards the affected regions.

A large number of applications in this category address monitoring of wildlife. This scenario has an increased complexity. The area of deployment is no longer accessible in an easy manner and is no longer safe for the sensor nodes. There is hardly any infrastructure present, and a large number of nodes have to be scattered around in a random manner; in addition, the network might contain moving nodes.

3.4.2.3 Intelligent Highways

Cars have integrated sensors and these sensor nodes will communicate with each other collecting information about the traffic, routes, and special traffic conditions. On one hand, new information will be available to the driver of each car. On the other hand, a global view of the whole picture will also be available. The two main constraints that characterize this scenario are the large number of nodes and their high mobility. The algorithms employed will have to scale well and deal with a network with a continuously changing topology.

3.4.2.4 Military Applications

Factors such as rapid deployment, self-organization, and increased fault tolerance make wireless sensor networks a very good candidate for usage in the military field. They are suited for deployment in battlefield scenarios due to the large size of the network and the automatic self-reconfiguration at the moment of the destruction/unavailability of some sensor nodes. Typical applications are for monitoring of friendly forces, equipment and ammunition, battlefield surveillance, reconnaissance of opposing forces and terrain, targeting, battle damage assessment, and nuclear, biological, and chemical attack detection and reconnaissance.

3.4.2.5 Health Care Applications

An increasing interest is being shown to the elder population. Sensor networks can help in several areas of the health care field. The monitoring can take place both at home and in hospitals.

At home, patients can be under permanent monitoring, and the sensor networks will trigger alerts whenever there is a change in the state of the patient. Systems that can detect their movement behavior at home, detect any fall, or remind them to take their prescriptions are being studied. Also, in hospitals, sensor networks can be used in order to track the position of doctors and patients (their status or even errors in the medication), expensive hardware, and so forth.

Wireless sensor networks can also use sensors implanted inside of the human body; a combination of sensors, RFID principles of operation, and propagation effects inside of the human body are described in more detail in Chapter 7.

3.4.2.6 Home Applications

The home is the perfect application domain for the pervasive computing field. We can imagine a future with all the electronic appliances forming a network and cooperating together to fulfill the needs of the inhabitants. These electronic appliances will have to identify each user correctly, remember their preferences and habits, and, at the same time, monitor the entire house for unexpected events. The sensor networks have also an important role here, being the eyes and the ears that will trigger the actuator systems.

3.4.3 Concept of Ambient Intelligence

The next step from RFID systems is ambient intelligence (AmI), a new field described as a seamless environment of computing, wireless sensor networks, advanced networking technology, and specific interfaces. AmI will have networking technology embedded in everyday objects such as furniture, clothes, vehicles, appliances, roads, and smart materials. This environment should be aware of the specific characteristics and the needs of users, should be capable of responding intelligently to spoken or gestured indications of desire, and should possibly even result in systems that are capable of engaging in intelligent dialogue.

AmI will be unobtrusive and simple in implementation and usage [10], and based on three key, fairly new, technologies: *ubiquitous computing*, *ubiquitous communication,* and *intelligent user interfaces*. By providing an intelligent environment, innovative intelligent personal health services can be developed while improving the quality and cost control at the same time [11].

The WSN and BAN are the necessary technology for the development of the concept of AmI where users (patients) are provided of services depending on their context [12]. Intelligent interfaces can empower people with severe motion impairments that can result from nonprogressive disorders, such as cerebral palsy, or degenerative neurological diseases, such as amyotrophic lateral sclerosis (ALS), multiple sclerosis (MS), or muscular dystrophy (MD) [13].

Although the application of the AmI vision may lead to a dramatic lowering of costs (reduction of time to diagnosis and time to treatment, outpatient diagnosis

and treatment, and so forth), the risk of dehumanization and depersonalization of the patient should be carefully considered. The problem may exist in a progressive dehumanization and identification of the patient with the collection of his or her vital parameters. In other words, the risk is that the patient will be progressively disembodied, reduced to the sum of his or her biological and physiological functions.

A second risk may arise from the possibility for the patient to monitor directly data detected and stored by wearable biometric devices. This capability may contribute to increase an awareness of the patient's body, but it may also increase the likelihood of self-diagnosis, with potential serious implications for the patient's health [14].

3.4.4 Sensor Networks Design Considerations

3.4.4.1 Sensor Networks Topology

The basic issue in communication networks is the transmission of messages to achieve a prescribed message throughput and quality of service (QoS). QoS can be specified in terms of message delay, message due dates, bit error rates, packet loss, and economic cost of transmission, transmission power, and so forth.

Depending on QoS, the installation environment, the economic considerations, and the application, one of several basic network topologies may be used. A communication network is composed of nodes, each of which has computing power and can transmit and receive messages over communication links, wireless or wireline. The basic network topologies are shown in Figure 3.9 and include fully connected, mesh, star, ring, tree, and bus.

A single network may consist of several interconnected subnets of different topologies. Networks are further classified as local area networks (LANs) (e.g., inside one building) or wide area networks (WANs) (e.g., between buildings).

Mesh networks are regularly distributed networks that generally allow transmission only to a node's nearest neighbors. The nodes in these networks are generally identical, so that mesh nets are also referred to as peer-to-peer networks. Mesh networks can be good models for large-scale networks of wireless sensors that are distributed over a geographic region, for example, personnel or vehicle security surveillance systems.

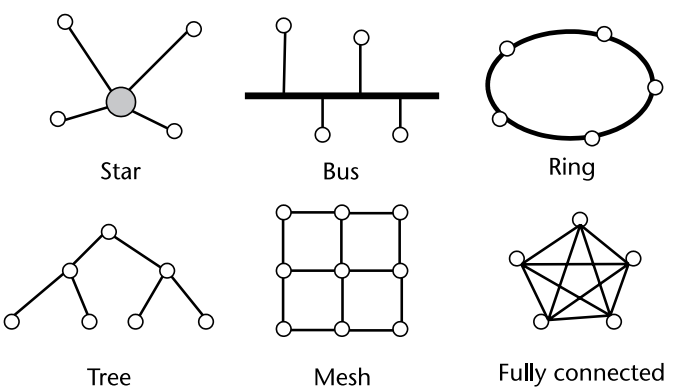


Figure 3.9 Basic network topologies.

Note that the regular structure reflects the communications topology; the actual geographic distribution of the nodes does not need to be a regular mesh [15]. Since there are generally multiple routing paths between nodes, these nets are robust to failure of individual nodes or links. An advantage of mesh nets is that, although all nodes may be identical and have the same computing and transmission capabilities, certain nodes can be designated as *group leaders* that take on additional functions. If a group leader is disabled, another node can then take over these duties.

The required transmission power in a wireless link increases as the square of the distance between source and destination. Therefore, multiple short message transmission hops require less power than one long hop. In fact, if the distance between source and destination is R, the power required for single-hop transmission is proportional to R^2. If nodes between source and destination are taken advantage of to transmit n short hops instead, the power required by each node is proportional to R^2/n^2. This is a strong argument in favor of distributed networks with multiple nodes, that is, mesh networks.

If we take a look at the number of sensors deployed with respect to the area covered, we can give the following categorization:

- *Coarse grained sensor networks:* In this category usually fall the sensor networks made up of devices, each covering a large area. These devices are usually large and expensive, because they are equipped with high quality sensors. The network topology is usually a star topology. The sensor nodes themselves are fixed.
- *Fine grained sensor networks:* This category comprises the networks made up of s large number of cheap devices, equipped with low-quality sensors having small amounts of resources available. The network topology is usually a multihop network. The large number of sensors and the dense deployment compensates the low quality of the sensors, the network as a whole producing high-quality results.

3.4.4.2 Sensor Networks Standardization and Compatibility

Desirable functions for sensor nodes include ease of installation, self-identification, self-diagnosis, reliability, time awareness for coordination with other nodes, some software functions and DSP, and standard control protocols and network interfaces. There are many sensor manufacturers and networks on the market today. It is too costly for manufacturers to make special transducers for every network on the market so different components made by different manufacturers should be compatible. Therefore, in 1993 the IEEE and NIST began work on a standard that resulted in IEEE 1451, an industry-wide open standard for intelligent sensors (smart sensors) [16].

The objective of this standard is to make it easier for different manufacturers to develop smart sensors and to interface those devices to networks. Under the original concept of IEEE 1451, a sensor is divided into two parts. The first, IEEE 1451.1-1999, called a Smart Transducer Interface Module (STIM), contains the sensing element (strain gage, thermocouple, vibration sensor, and so on), the appropriate signal-conditioning circuits and A/D converter, and a Transducer Electronic Data

Sheet (TEDS), a memory chip that identifies the type of sensor, its make and model, its calibration information, its scale factor, and more. IEEE 1451.2-1997 defines the basic STIM, while IEEE 1451.1 defines the Network Capable Application Protocol (NCAP) (Figure 3.10).

Although 1451.1 and 1451.2 worked well, they did not cover enough configurations, and another substandard was started, IEEE 1451.3-2003, "Standard for a Smart Transducer Interface for Sensors and Actuators, Digital Communication and Transducer Electronic Data Sheet (TEDS) Formats for Distributed Multidrop Systems" (often called Dot3), which allows multiple transducer modules (called *transducer bus interface modules*) of varying complexity and data rates to be multidropped to one NCAP via a local transducer bus.

One limitation of IEEE 1451 was the lack of backward compatibility, that is, it did not address the large number of legacy sensors devices with analog outputs and no digital communications already in use. IEEE 1451.4-2004, "Standard for a Smart Transducer Interface for Sensors and Actuators-Mixed-Mode Communication Protocols and Transducer Electronic Data Sheet (TEDS) Formats" (also called Dot4), was proposed as a way to do this. Unlike Dot2 and Dot3 sensors, a Dot4-compliant sensor has an analog output and no A/D converter, but it does contain a TEDS.

Other related standards are:

- IEEE 1451.0-2007, "Standard for a Smart Transducer Interface for Sensors and Actuators—Common Functions, Communication Protocols, and Transducer Electronic Data Sheet (TEDS) Formats";

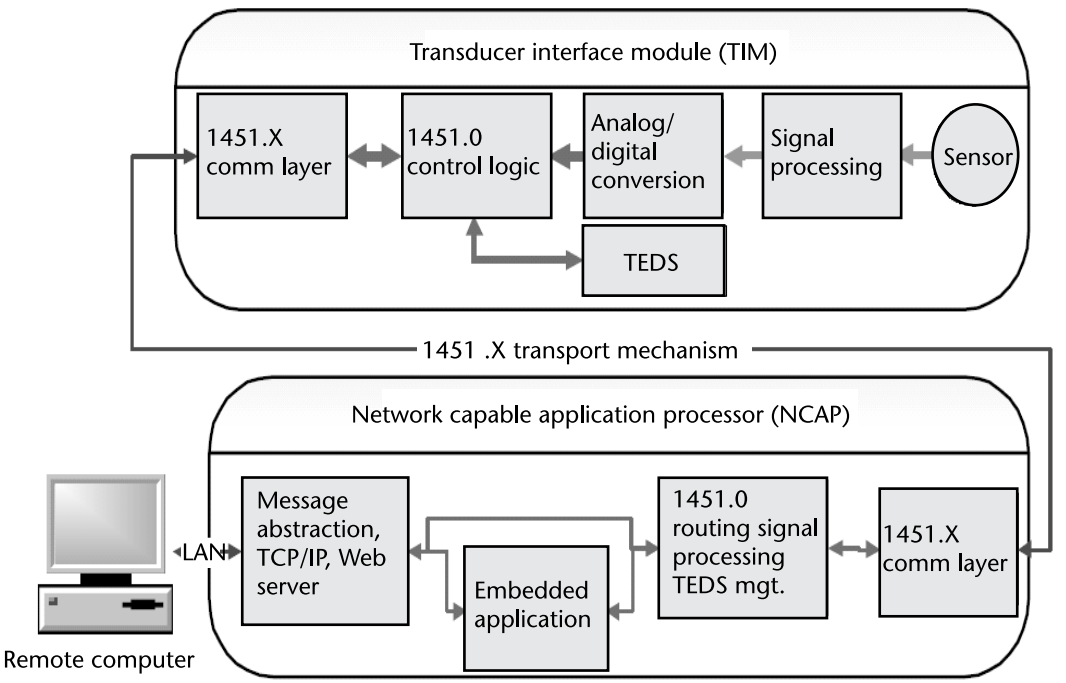


Figure 3.10 IEEE 1451 smart transducer concept.

- IEEE 1451.5-2007, "Standard for a Smart Transducer Interface for Sensors and Actuators—Wireless Communication Protocols & Transducer Electronic Data Sheet (TEDS) Formats";
- IEEE 1451.7-2010, "Standard for a Smart Transducer Interface for Sensors and Actuators—Transducers to Radio Frequency Identification (RFID) Systems Communication Protocols and Transducer Electronic Data Sheet Formats."

The Proposed IEEE 1451.6 Standard is a new and developing standard that combines IEEE 1451 and Intrinsically Safe (IS) technologies[2].

3.4.5 The Future of RFID Sensing

Industry is fast moving toward employing networked, digital, and wireless communications technologies for sensors. Using wireless connectivity for sensor networks increases the flexibility in deployment and reconfiguration and thus reduces the overall infrastructure cost. These advantages will enable sensor networks to monitor complex environments for applications ranging from industrial automation to battlefield surveillance to environment monitoring to the telemetry of a first responder's health condition.

RFID devices are going to play a key role in automated universal identification system for accessing, securing, and tracking assets, personnel, equipment, and products throughout the supply chain. Combining RFID devices and sensors could expand the overall functionality and capability of the above applications.

In today's complex supply chain, even if technology standards are high, visibility still performs at very low efficiency levels. Asset tracking and management, anticounterfeit and in-transit visibility are few of the major contributions of RFID technology to the real world. Several suppliers already have met or are preparing to meet their retailers' mandates. However, what if suppliers could not only find where their products were located, but also receive information about their conditions and status?

Location is just one part of the challenge. According to the U.S. Department of Agriculture, 10% of all perishable products spoil before they reach the consumer. The solution lies in adding several sensors to the RFID solution. So far, RFID has enabled a number of end users to gain a higher visibility of their supply chain, increase inventory efficiency, reduce out-of-stock items, and gain higher anticounterfeit protection. RFID sensors can create and provide a new layer of protection and advance supply chain visibility.

Efficient sensor networks require the sensor nodes to be cheap, to consume little energy, to be multifunctional, to be small, and to have the ability to communicate both among themselves and with other networks. Compared with mobile ad hoc networks, wireless sensor networks differ in various ways, including the larger number of nodes, the dense deployment, and the attribution of fault-proneness, the

2. Intrinsically safe is a condition of safety in a hazardous environment, such as in areas where explosive gasses or other flammable items exist. Any intrinsically safe instrument must not cause any type of ignition in any form under normal operation. In case of failure, it must not produce any form of hot spots.

frequent topology changes, the main use of broadcast communication instead of point-to-point communication, and the limitations in power, storage, and processing units.

3.5 RFID Applications

RFID systems have been deployed in limited numbers for years; two of the most predominant have been in the form of toll road collection transponders and security badges. Toll road authorities around the country have equipped drivers with transponder that are connected to their credit cards, which allows them to pay their tolls at 40 mph rather than stopping to throw quarters into a basket and slowing down the flow of traffic.

Security badges have been equipped with RFID chips to allow centralized control of access to facilities and specific rooms within buildings. These can also be used to track the locations of people in a facility by identifying the door through which they last passed. Today, RFID has the potential for applications in virtually any area of industry, commerce, and services where items are handled and associated data collected and processed. These include:

- Supply chain logistics;
- Product authentication;
- Tracking and traceability;
- Security, ticketing, and access control;
- Lifetime item identification;
- Transient carrier labeling;
- Animal and specimen identification;
- Airline baggage handling.

The following sections will illustrate just a small number of examples; possibilities are truly endless and limited only by our imagination.

3.5.1 Supply Chain Logistics

Global supply chain logistics are expected to be the largest and fastest-growing application for RFID. This will most likely be done through smart labeling of cases, cartons, and pallets. The key benefits are the ability to read the entire contents of mixed pallets all at once during material handling operations such as truck loading or unloading.

Managing pallets, totes, and other returnable transit containers with RFID represents one of the most dramatic cost-saving opportunities that this technology can provide. Many returnable containers are never brought back from customer sites after shipment, forcing companies to carry excess inventory to ensure adequate supplies of shipping materials where they are needed. Identifying returnable containers with smart labels or fixed tags enables companies to augment their legacy

bar code shipping applications by automatically recording materials shipped to customers.

Fast-reading RFID enables instant identification of the shipping container and all of the individual items inside. For shipping, RFID readers can help packers quickly locate and aggregate all the items needed to complete an order.

The same principle is applied to improve warehouse picking. Workers scan shelves and bins with an RFID reader that automatically detects the storage location of the sought items. The system can also detect items that are stored in the wrong location and alert operators to the problem. Using RFID for these applications enables items to self-report their locations, rather than requiring human intervention to find them, thus reducing errors, saving labor, and lowering costs.

Wal-Mart Stores Inc., one of the originators of the RFID movement, continues to expand its RFID capability to additional facilities. Aside from the initial RFID implementation to track inventory, in the near future, customers will be able to enjoy advantages such as automatic warranty activation on electronics, freshness assurance on foods, thanks to cold chain monitoring, and enhanced product safety as a result of faster, more accurate recalls and better freshness monitoring.

3.5.2 Product Authentication

The role of product authentication is to answer whether a given product is genuine or counterfeit (e.g., a product that infringes a trademark). An explicit way to authenticate products is needed in supply-chain applications because counterfeits can be very similar or even identical to authentic products. The starting point of automated nondestructive product authentication is to insert a special label or security feature into products, like a hologram or a water mark, and to authenticate this label. Product authentication can take place in single item level (*smart label*[3]) or in aggregated levels.

Generally, multiple similar units are authenticated simultaneously, for example, when a shipment arrives to a retail store. The desired level of security, which can be defined as the effort that an illicit actor has to undertake to break or bypass the security mechanism, has a major impact on the cost of a product authentication system. While minimizing the cost, the level of security should be high enough to protect the item over its entire lifespan. Because different products have widely varying security requirements, different levels of security and thus different solutions are needed.

The level of security of product authentication system is defined by the level of security of a single security feature and by the granularity of the security features. By *granularity,* we mean how many products use an identical security feature; for example, applying weak but unique security features to all products can be more secure than using strong but identical features on the same products. One conceptual problem of automated product authentication is that it is only the security fea-

3. Label with integrated covert inscriptions helps eliminate the illegal practice of empty vials being refilled with fake drugs and resold. A multipart label has integrated security features; a label is functionally destroyed during its initial use. When the label's tear strip is removed, the inscriptions "opened" and "used" appear, making undetected reuse of a glass container with an original label virtually impossible.

ture that is authenticated and not the product itself; therefore, difference between label and product authentication should be made.

The general requirements of a product authentication system in the supply chain application are listed here:

- The system needs to be used by multiple parties from multiple locations.
- The authentication of products that are unknown to the system should be supported.
- The cost and effort to perform a check need to be low.
- The optimal solution should allow also the customers to authenticate products.
- The product authentication system needs to have an appropriate level of security.

Among these requirements, the level of security demands most attention in the system design. The level of security can be considered as the resistance against attacks that are conducted against the authentication system. In supply chain applications, product authentication is typically performed under the supervision of authorized personnel, thus restricting the possible attacks of counterfeit players.

The general attack scenarios of illicit actors against product authentication system can be divided into following four categories:

- Omission of security features that are applied on the genuine objects refers to the counterfeiters' not taking any explicit actions to fool the authentication. These products form a considerable part of the counterfeit trade due, for example, to consumer demand of counterfeits.
- The use of misleading security features means that the fake products are equipped with security features whose role is to make the products avoid closer inspection. Interviews with brand owners and customs reveal that this scenario, together with the aforementioned one, is dominant, especially for all goods that are mass produced or where the consumers do not regularly check for the object's authenticity.
- The removal and reapplying of authentic security features remain a threat in all automated product authentication systems if not explicitly addressed by binding the product and the label. However, because acquiring and reapplying authentic labels are costly, this attack does not threaten authentication systems in large scales.
- The cloning and imitation of security features are the most obvious attack that a product authentication system has to resist. As the underlying problem of counterfeits is that the products themselves can be cloned, the first line of defense is to integrate such security features into products that are hard to be replicated.

The benefits of RFID compared to old authentication technologies include nonline-of-sight reading, item-level identification, the nonstatic nature of security features, and cryptographic resistance against cloning. RFID systems, in general,

comprise transponders, readers, or interrogators and an online database, sometimes referred to as the back-end server. In many applications, RFID transponders are already being used for authentication, for example, in access control. Although RFID product authentication is very close to RFID access control, when it comes to the used authentication protocols, product authentication needs specific solutions because of the specific application requirements.

Resisting cloning and forgery is one of the most important security properties of authentication tags. The simplest cloning attack against an RFID tag only requires reading the tag serial number and programming the same number into an empty tag. There are two essential obstacles against this kind of replication. First, even the low-cost transponders (e.g., EPC Class-1, Generation-2) have a unique factory programmed chip serial number [or transponder ID (TID)] that is similar to the unique MAC address of PC network cards. Cloning a TID would therefore also require access to hardware manufacturing.

The second obstacle against cloning is to place read-protected secrets on tags and to check if the tag knows these secrets, for example, by cryptographic challenge-response protocols. Even though this can provide a significant improvement to a tag's ability to resist cloning, many ways remain in which to conduct a cloning attack against a single tag. These attacks include side-channel attack, reverse-engineering and cryptanalysis, brute-force attack, physical attacks, and different active attacks against the tag.

In addition, shared secrets–based product authentication approaches are always vulnerable to data theft, where the secret PIN codes or encryption schemes of valid products are stolen or sold out by insiders, which would enable criminals to create phony tags. This scenario is especially interesting for adversaries because it would allow them to clone a large number of tags.

Other RFID security issues that have to be considered in product authentication comprise resistance against denial-of-service (DoS) attacks. In general, DoS causes loss of service to users. Even though it cannot be used to fool the product authentication, it can pose a threat for the overall process. In RFIDs, DoS attack can be conducted, for example, by jamming the readers with hidden blocker tags or by desynchronizing tag and a database entry.

We assume that product authentication is normally performed under the surveillance of authorized personnel or by the customer, which narrows down the possible attack scenarios. Therefore, active attacks, in which the adversary would need to participate in the authentication session and use special devices in the proximity of the reader (e.g., replay, relay, and man-in-the-middle attack), are not considered as realistic threats against RFID product authentication.

3.5.3 Agriculture and Animals

RFID is already used to manage commercial livestock by farmers, improving farming efficiency and allowing them to be traced back to their origins (in case of disease). The technology can also be used to effectively develop a *track-and-trace* of meat, to keep track of pets and their vaccination records, for easy retrieval of pets by owners, to identify animals, and to reduce the potential spread of diseases when crossing borders (see Figure 3.11).

Figure 3.11 Glass transponder for the identification of animals (and humans).

The technology is also being used for fish and wild game to track migrations, breeding patterns, population, and so forth and to prevent poaching and illegal exporting of endangered species and ivory tusks.

Researchers have noticed recent cases of cannibalism in polar bears in the Arctic (something that is previously undocumented) and further investigation will be required. A study on polar bears in Alaska, by members of the U.S. Geological Survey (USGS), shows that some types of RFID tags can be read from as far away as 1,500 feet, while the reader is in motion (in this case, a helicopter).

In terms of savings, the RFID ear tags cost $35 and the battery lasts 5 years. In comparison, the older satellite radio collars cost $4,000, with batteries only lasting 2 years. Obviously, the radio collars can be tracked at a greater distance.

3.5.4 Intelligent Transportation Systems

Traffic congestion in the largest cities of the world is a growing problem that has to be taken into account seriously, not only by governments, but also by the private sector. After an extensive survey and different alternatives being analyzed, the concept of intelligent vehicle/highway system (IVHS) is proposed as the best solution [17].

IVHS has been called by different names depending on the developing area and the application purposes—intelligent transportation system (ITS), intelligent cars and automated highways systems (AHS), and automated vehicle/highway system (AVHS) are just some of the different names basically describing the same concept.

IVHS is an intelligent transportation system in which vehicles and highways will exchange information through a two-way communication system (see Figure 3.12). The automated highways will have a set of lanes on which vehicles with specialized sensors and wireless communications systems could travel under computer control at closely spaced intervals. This type of arrangement is called a *platoon*.

The vehicles could continuously exchange information with other vehicles and traffic-control centers about speed, acceleration, braking, obstacles, road conditions, and other vehicle data. Sensor data can be processed and sent back to each vehicle, guaranteeing a continuous exchange of information.

The highway system will know the destinations and planned routes of individual vehicles. In that way the system can coordinate traffic flow more efficiently, reduce speed fluctuations, monitor unsafe vehicle operation and traffic shock waves, maximize highway capacity, and minimize avoidable traffic congestion. In

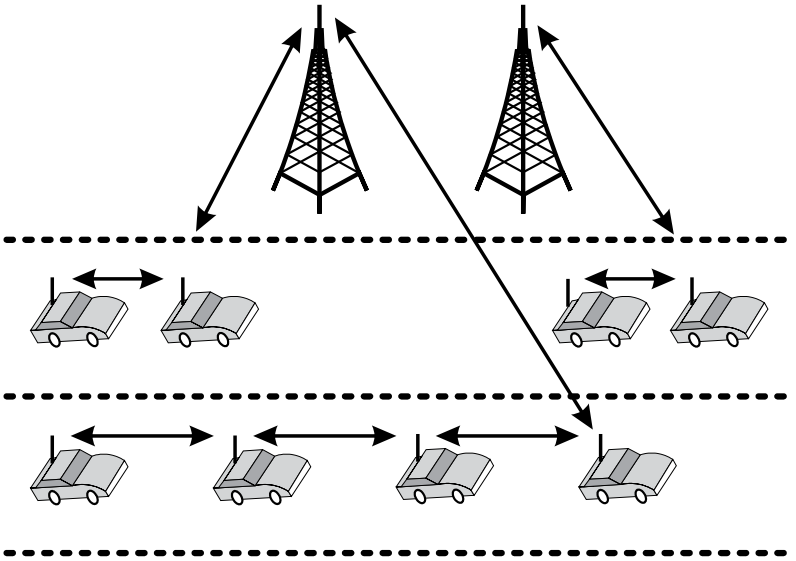


Figure 3.12 Intelligent transportation system.

addition, the system will respond rapidly to changing highway conditions. The vehicles might use several types of devices to sense its environment, such as magnetometers, visual sensors, infrared sensors, laser sensors, and accelerometers. Each vehicle has to have a powerful computer to process sensory data and the information that come from the traffic control centers.

IVHS America has become the coordinating and planning entity in which the individual activities of state and local authorities, companies, and universities have a central orientation for constructing a national IVHS program. VERTIS (Vehicle, Road, and Traffic Intelligence Society) in Japan and ERTICO (European Road Transport Implementation Coordination Organization) perform similar activities as major coordinators of the intelligent vehicle/highway system programs.

Some proposals have suggested using RFID technology for positioning [18]. This technique, however, would not replace GPS; rather, it is a complementary technique. RFID tags need to be installed on a road in a manner that could maximize the coverage and the accuracy of positioning. On installation, necessary information such as coordinates of the location in which the tag is installed needs to be written on each tag. The accuracy of this position information is very critical for this technique to be successful. The position information can be acquired by using differential GPS (DGPS) or some other methods, which would take much longer time to compute the location.

Contrary to GPS in navigation systems in which real-time positioning is necessary, the time for getting the accurate information would be tolerated since this computation would take place once. Vehicles then need to be equipped with an RFID reader that can communicate with the tags on a road.

No matter how accurate the RFID positioning is, it only gives the position where the tags are. Therefore, the vehicles need also to be equipped with a GPS receiver and inertial sensors such as a gyroscope for positioning when there are no tags around. While driving, the vehicles constantly monitor the presence of a tag. On detection, the reader retrieves the information from the tag including a lane

marker. The deployment should be done step by step; places such as tunnels from which getting GPS signal is not an option should be the first, intersections the next, then urban areas, and then nationwide. Due to this nationwide scale, governmental actions would be necessary.

3.5.5 Document Management

Books and other materials are identified with smart labels that carry a unique, tamperproof ID code. Librarians at the circulation desk and patrons read the tags with RFID readers to check items in and out. The process is faster and more accurate than with traditional optical bar code labels. Some economic facts that help justify installing this system are as follows:

- A lost book typically costs a library around $45;
- An average library can have as many as 22 million items circulating each year;
- With RFID smart labels on items, check in and check out saves 1.5 minutes per transaction.

Besides the unique ID number, these labels can be programmed with additional information, such as type of media and storage location. In the retail RFID space, the EPCglobal suite of RFID specifications mandates that tags support an irrevocable kill command. In the library setting, however, tags must be reused to check in loaned items. Irrevocably killing a tag is not an option. The tag has to be rewriteable so libraries do not have to replace a book's digital identification tag when updating a book's status or flagging a book for reservation. Libraries are finding new ways to take advantage of tagged items, such as gathering statistics on what items are most often used.

Aside from the libraries, RFID can also be used to improve the management of important individual document files in industries such as insurance companies and law offices, where the loss of such files can cause severe problems. RFID improves the tracking of documents so that files can be quickly located and document workflow more easily tracked.

Each file is tagged with a smart label that contains a unique ID and human-readable information. The file description is entered into a database along with its tracking number. The file can be assigned certain parameters such as expiration date, permitted movement, and persons authorized to see it. Over time, the database can build up an audit trail of the handling and workflow history of each document file.

3.5.6 Pharmaceutical and Health Care Industry

A number of pilots are already under way in the pharmaceutical supply chain and health care markets for item-level management. Although much remains to be learned about the efficiencies and safeguards that can result from the use of RFID solutions in these markets, companies implementing RFID pilots are experiencing process improvements and safety benefits today. Suppliers to the medical industry,

from garment to surgical instrument providers, as well as health care institutions managing blood and tissue sample processing, are investigating the viability and reliability of HF technology solutions, and they are seeing significant returns in the field [19].

3.5.6.1 Tracking Medications

State and federal laws have been enacted in order to address the increasing threat to public health posed by counterfeit drugs. At the federal level, the U.S. FDA enacted the Prescription Drug Marketing Act (PDMA), which went into effect in December 2006, with a mandated requirement that every drug must have a full pedigree.[4] In addition, many states are enacting their own pedigree laws. Florida's pedigree law, which enables companies to select either paper or electronic record keeping, requires full documentation of the drug from the manufacturer to the store, requiring each shipment to be accompanied by the amount of the drug, dosage form and strength, lot numbers, name and address of each owner with owner signature, and complete shipping information.

California's pedigree law, which specifies the requirement for an electronic pedigree, went into effect on January 1, 2007. Many other states are considering similar laws.

The need and value of RFID in the pharmaceutical industry are well recognized and documented. With the need to meet U.S. federal and state regulatory compliance in the very near future, many companies in the pharmaceutical industry are aggressively seeking to implement RFID now (Figure 3.13). Leading companies have already defined critical strategic and business needs, as well as the system specifications that will meet those needs. These requirements are some of the most demanding and stringent of any RFID applications to date.

For pharmaceutical manufacturers and distributors, RFID will provide a solution for three critical issues: improvement of counterfeit protection; regulatory compliance by providing the ability to create a complete electronic pedigree, automatically, without human intervention; and rapid and cost-effective recalls by

Figure 3.13 RFID tag on the pill bottle.

4. Pedigree is information required to ensure the security and authenticity of a drug as it travels through each step in the pharmaceutical supply chain.

providing the ability to instantly and automatically identify the location of product that must be recalled.

3.5.6.2 Locating Tissue Samples

What was previously a painstaking and time-consuming task of locating and identifying samples can now be completed quickly with a simple pass of an RFID reader over the existing inventory.

Tissue-sample processing labs are using miniature HF tags to create efficiencies in locating single test tube samples among the hundreds in the laboratory at any given time. The RFID tag contains a unique serial number as well as memory that can be read, modified, and protected. The serial number is then linked to a database containing critical information on each tissue sample, including patient data and tissue treatments. Using a fixed desktop or lightweight handheld reader at a distance of a few inches, researchers and laboratory technicians searching for a specific sample on a tray of 100 tubes can quickly and easily read all of the tags in less than a few seconds.

3.5.6.3 Patient Identification and Care

The RFID-based system allows medical professionals to use RFID-enabled wristbands to identify patients and to update their status, location, and medical information in the system's electronic whiteboard automatically. The Navy[5] implemented a new system to replace a labor-intensive, entirely manual system consisting of pen and paper, cardboard tags, and a centrally located whiteboard to show patient movement throughout the hospital. With the new electronic system, each patient receives an HF-enabled wristband, on which basic identification information is stored.

Medical professionals use a handheld RFID device to read the unique ID number, and to add or change data to create a digital treatment record that travels with the patient as he or she is moved throughout the facility. Using a wireless LAN, patient information is transferred to an electronic patient management system, further eliminating manual reentry of data at a central computer terminal.

3.5.6.4 Matching Blood Samples to Patients

In the trial, the doctor or nurse taking the blood sample enters the patient information into a handheld RFID device at the start of the blood sampling process. This data is stored on an RFID label (HF type) on the patient's blood sample tube and can be read by fixed readers and automatically transferred to the facility's database, enabling a fully automated process and replacing an entirely manual one. In addition, because the patient data is entered in electronic format at the beginning of the process, the integration of results into the patients' records is quick and simple.

5. The U.S. Navy is using HF technology to more efficiently track the status and location of hundreds of wounded soldiers and airmen, prisoners of war, refugees, and others arriving for treatment at Fleet Hospital Three, a 9-acre, 116-bed facility in Southern Iraq.

Trials have shown significant reductions in administration time, both during sampling and laboratory processing.

3.5.6.5 Context-Sensitive Medicine

By using active RFID tags, management of resources becomes more intelligent. For example, equipment needing servicing can automatically inform the technician. Patient care can be scheduled, and reminders will be sent out exactly when action is required.

The location of equipment, doctors, and patients inside a hospital can be tracked in real time. This allows the patients to move around freely while it also allows doctors and nurses to tend to more patients.

3.5.7 Indoor Localization for First Responders

Most of the research and development for indoor localization include that of a wireless network that integrates communications, precise tracking, and data telemetry, for use in hospital and manufacturing environments. In contrast, the system used by first responders is intended for an environment that is potentially much less friendly to RF propagation: the in-building environment that may contain smoke, dust, or flames.

The technology is intended to leverage advances in ubiquitous RFID tag technology and in combination with recent advances in miniaturized inertial sensors [20]. This research is a joint effort by components of three NIST laboratories: the Wireless Communication Technologies Group of the Information Technology Laboratory (ITL) and the Fire Fighting Technology Group of the Building and Fire Research Laboratory (BFRL) in Gaithersburg, Maryland, and the Radio-Frequency Fields Group of the Electronics and Electrical Engineering Laboratory in Boulder, Colorado.

The most widely used navigation system today is the Global Positioning System (GPS), which enables position determination through the measurement of time delays of signals from multiple satellites in known (moving) positions; the time-delay measurements are based on cross-correlating received satellite signals with local replicas to identify the signals' digital code position in time relative to the common reference. The difficulty in using GPS indoors and in urban canyons is that the line of sight to the GPS satellites is obscured or severely attenuated. Without four good satellite signals, the GPS position solution is inaccurate. In addition, with weak signals, the GPS receiver continually loses lock and must spend an inordinate amount of time in attempting to acquire the signals.

The research aims to find the way to implement a low-cost, reliable means for tracking firefighters and other first responders inside buildings, where navigation using GPS is not reliable or the GPS signal may have been disabled temporarily to prevent exploitation by terrorists. Even if the GPS signals are not blocked or obscured for tactical advantage, the reception of GPS signals inside most buildings is not reliable.

Prior to the establishment of GPS many techniques and devices for navigation have been used. Today's navigation devices implement some very old navigation techniques, such as dead reckoning and waypoint navigation.

Dead reckoning (DR) is the process of estimating position by advancing a known position using course, speed, time, and distance to be traveled, in other words, figuring out where a person will be at a certain time if he or she holds the planned speed, time, and course. The usefulness of the technique depends upon how accurately speed and course can be maintained in the air and on the sea, and the uncertainty of the DR position grows with time, so that it is necessary to check the position regularly with a fix of some kind, perhaps an RFID tag.

It has been noted that 0.6–2 GHz is the best frequency band for propagation inside the buildings.

3.5.8 Passive Keyless Entry

When in 1993, a major German insurance company put up pressure to either protect vehicles against massively increasing theft or else renounce to full insurance coverage, nobody believed that RFID as a then-emerging technology would see one of its first major successes, still unbeaten in numbers by any other application today. Objection was raised as to the reliability of potential systems and their suitability in an automotive environment; and, of course, the lack of standards in what then seemed to be a collection of proprietary systems was seen to be a major obstacle. An extremely fast but substantial development effort was undertaken by a few powerful automotive industry suppliers, resulting in a miniaturized ignition key system and its car lock transceiver counterpart to be fitted to new cars a year later.

This first generation system, still fitted to certain models of most major car makers, consisted of a 64-bit read-only, rod- or brick-type transponder, embedded in the ignition key of the vehicle and a transceiver antenna with its electronics package on printed circuit board (PCB), integrated around and behind the lock. When the presented key does not match one of the prestored codes, the protection consists of safely cutting power supply to the starter, the fuel pump, the system ignition, and other system elements required for the vehicle's operation during driving for the case.

Specifications for the miniaturized transponders were very stringent, requiring maximum read distance in a metal-loaded environment over industry-practice thermal ranges with only a few parts-per-million failure rate allowed. Three mainstream systems were adopted by the majority of the automotive industry and have since been fitted to millions of new cars with great success, and immobilizers today, for example, in Europe, are part of the normal equipment of cars.

Hands-free passive keyless entry (PKE) applications require bidirectional communication between the base station and transponder units. The base unit inside the vehicle transmits a low-frequency command that searches for a transponder in the field. Once located, the transponder in the vehicle owner's possession then automatically responds to the base unit. The base unit then unlocks the car doors if a valid authentication response is received.

The company's couriers use an automatic keyless entry and ignition system that has RFID transponders embedded within a Velcro wristband. The FedEx system uses RFID readers mounted at each of the four doors to the delivery vehicle and a reader mounted on the right side of the steering column near the ignition switch.

When the courier places his or her transponder wristband within 6 inches of the readers, the transponder's code is compared to ones in the system's memory. If

it is a match, the door unlocks for 5 seconds. The courier simply pulls on the door handle to enter the vehicle, while the three remaining doors stay securely locked to prevent unauthorized entry. To start the vehicle, the courier pushes a button on the right side of the steering column. The courier pushes another button near the start button to turn off the vehicle.

3.5.9 Military Applications

The U.S. Department of Defense (DoD) became involved in RFID during the 1990s due to the identification of supply chain challenges. During Operation Desert Storm in 1991, logistics and materiel distribution were major problems.

The Defense Logistics Agency (DLA) became known for mountains of unopened shipping containers in the middle of the Saudi Arabian desert. The lack of supply chain visibility required 25,000 of the 40,000 containers to be opened in order to identify their contents. A Defense Research Projects Agency (DARPA) grant was awarded to identify whether RFID could help prevent similar supply chain problems in the future, resulting in several initiatives over the next few years.

Evaluations of the DLS's effectiveness during Operation Desert Storm focused on the high cost of the DLA's supply chain. In 1995, the Joint Total Asset Visibility office was formed with a charter to provide asset visibility in-storage, in-process, and in-transit to optimize the DoD's operational capability. This had several results. First, it organized all RFID supply chain initiatives under one office, instead of being managed by individual armed forces or distribution depots. Second, it provided a source for funding future RFID initiatives. By 2004, the DLA had spent over $100 million on RFID initiatives; this level of funding would not have been available under the previous organizational structure. Finally, the implementation plan tied RFID usage to the overall strategic goals mandated by the department's charter. By 2004, the DoD had joined EPCglobal.

When it comes to RFID, the DoD and the armed services have learned a thing or two from private industry. The DoD, for instance, is adding passive RFID tags to cartons and pallet, and rather than reinvent the wheel, the DoD is working with established industry technologies. Similarly, the armed services are implementing active RFID solutions to track mobile assets, like containers in the field and work in process. The Defense Appropriations Act for fiscal year 2007 has had a total of $17 million added for projects either directly or indirectly related to RFID [21].

According to the DoD's policy, which was finalized and released in July 2004, by January 1, 2005, suppliers had to put passive RFID tags on all individual cases, all cases packaged within a pallet, and all pallets of packaged troop rations, clothing, individual equipment and tools, personal items, and weapons systems repair parts and components shipped to the two DLA distribution centers.

The Tobyhanna Army Depot, for instance, is using a Real-time Locating System (RTLS) to streamline the repair and overhaul of defense electronic systems. The Army maintains, repairs, and overhauls command, control, communications, intelligence, and reconnaissance systems at the 1.9-million-square-foot refurbishment center in Pennsylvania, including the Army's radar system, which detects and tracks enemy mortar and artillery shells in Afghanistan and Iraq. These systems are shipped from the field to Tobyhanna. There, each system is disassembled, overhauled, and tested before being shipped out into the field again.

The RTLS allows Tobyhanna to track hundreds of components during the process that might otherwise get misplaced in the massive facility. The system will expedite the refurbishment process, enabling the Army to return systems to the field up to 35 days sooner than in the past [22].

The U.S. Army is developing an RFID system to track weapons usage. In an effort to make sure that ground vehicles' weapons are properly maintained, Benét Laboratories is working on a system using sensors and passive RFID tags to record the number of rounds fired. The first vehicle to demonstrate the sensor system will be the M1 Abrams Main Battle Tank. The sensor systems will include microelectromechanical systems (MEMS) sensors integrated with RFID tags to help the vehicle's operator track how many times a weapon has been fired. This allows the gunner to determine whether it can be depended on to function properly. To date, vehicle and weapon operators have been required to log manually, on paper, how many times a weapon has been fired and the types of munitions used and then bring that information to the vehicle maintenance depot.

The first phase of the deployment, which will measure applied force, will focus only on using the MEMS sensors to count the rounds fired. However, the MEMS sensors will eventually be used to measure a fired round's physical effects on the gun barrel as well, including the intensity of the applied force, heat and vibration. This will assist not only in counting the number of rounds fired, but also in recording the physical effects of the type of munitions fired in each instance, resulting in an accurate indication of the health and maintenance requirements of the barrel.

An operator can access the weapon-firing count on a tablet PC inside the vehicle. The tablet includes an RFID reader, and its data can be downloaded by Army personnel to keep a record of the weapon's firing history. When a vehicle operator prepares the weapon for firing, the reader automatically sends an RF signal energizing the passive RFID tag, which transmits its unique RFID number and a count of each time the weapon has been fired.

The tablet PC will include integrated data storage, communications, and display technologies. The MEMS will incorporate an RFID tag and a low-power microprocessor with limited memory, which would continue to record rounds fired even if the rest of the system were to fail [23].

3.5.10 Other RFID Applications

Gillette (which announced in early 2003 that it would purchase 500 million RFID tags) has worked with retailers to test smart shelves, as an adjunct to item-level tagging, for inventory control. With a reader on each shelf and a tag on each package of razor blades, the data proprietor would always know how many packages are on the shelves, without having to count them.

Michelin has begun fleet-testing for RFID for passenger and light truck tires (Figure 3.14). Each tire's unique ID number will be associated in an external database with the vehicle identification number (VIN) of the car on which it is mounted, and with information describing when and where the tire was made, its maximum inflation pressure, its size, and so on. The transponder consists of a UHF/SHF RFID chip, circuit board, and two antennas. The antenna leads must be tuned to resonate at 868 and 960 MHz when cured into the tire. The current design has

Figure 3.14 RFID tire transponder.

a nominal recommended tuned length of around 61 mm end to end. Exact tuned length will always depend on tire type.

Casinos are putting RFID tags in chips to block counterfeiting, identify stolen chips, and track gamblers' play. An Italian manufacturer has introduced a washing machine equipped to read RFID washing instruction tags in clothing.

A German supermarket, for a brief time, inserted RFID tags in supermarket loyalty cards, which gave the store the capability, while someone carrying the loyalty card was in the store, to pull up his or her entire buying history without his or her being aware that the query was taking place and without any other basis for the store's knowing who he or she was.

It is possible to imagine a whole lot of uses, indeed, for a technology in which objects can be uniquely identified without direct contact. This is the ideal technology if you want the milk in your refrigerator to notify you (or your supermarket) if you have failed to drink it by its pull date. Indeed, you could tie a slightly more elaborate tag to a nanosensor that checks for spoilage directly.

Students in an Osaka, Japan, elementary school will be getting RFID chips in their schoolbags, name tags, or clothing, to be read by readers installed in school entrances and exits.

The University of Washington's Brockman Memorial Tree Tour uses RFID and PDA technologies [24]. The combined use of these technologies increases the ease of locating and positively identifying each tree along the lengthy, spacious tour around the campus. Each tree is individually profiled including details such as origin, data planted, types of flowers, and other pertinent information. A map, located at the rear, visually locates each tree by its unique number. Originally, a nameplate was attached to each tree along the tree tour for quick identification from a distance. Over time, however, many nameplates have been removed, damaged, or destroyed due to weather, growth, or vandalism. The optimum solution proposed was to embed RFID tags within the trees and retrieve the ID numbers via a tree-penetrating RF signal. The updated tour information is professionally read and recorded into PC-formatted sound files. The files are then embedded within the electronic tree tour version.

At the click of a button, the text for each tree is played back through the onboard speaker on the PDA device. The advantage is that the user can simultaneously observe the tree and receive audio information. Also, the high-volume capability of the chosen portable information device enables groups to share a single system with one person designated as a chief navigator and tree scanner. In the electronic version of the tree tour, a map with significantly higher detail is displayed on a PDA device with a crisp color screen at a magnification roughly four times that of the paper version.

The industry's biggest score to date, however, could be in the works in Europe, where rumor has it that the European Central Bank is working with vendors on weaving RFID into the fabric of its bank notes. The technology, most probably for incorporation into larger bills, would enable money to carry its own history. Hence, it would become more difficult for kidnappers to ask for unmarked bills. It would also enable law-enforcement agencies to follow the money in illegal transactions. The U.S. government is said to have expressed interest as well.

A 2.45-GHz, high-frequency, ultrasmall antenna, embedded in an ultrasmall IC chip (a μ-chip or mu-chip), for RF identification is shown in Figure 3.15. An ultrasmall ($0.3 \times 0.3 \times 0.06$ mm) RFID chip, called the μ-chip, has been developed for use in a wide range of individual recognition applications. The chip is designed to be thin enough to be applied to the paper and paper-like media widely used in retailing to create certificates that have monetary value, as well as to tokens. It was designed and fabricated using 0.18-μm standard CMOS technology.

A conventional general-purpose type μ-chip needs an external antenna to transmit the 128-bit ID number inside the chip. In this new type of μ-chip, only a small IC chip is required for batteryless data transmission operations, because the embedded antenna on the chip is able to receive electromagnetic power from the reader [25].

This miniaturization enables RFID devices to be easily inserted inside a bank note or gift card. Furthermore, this feature also enables the identification devices to be attached to a narrow surface, or inside thin materials. The embedded antenna is formed by using gold-plating technology, widely used in the fabrication process of packaging connection terminals on semiconductor wafers. Therefore, the antenna is not only ultrasmall, but also very easily made on the wafer by the fine and batch antenna-forming process.

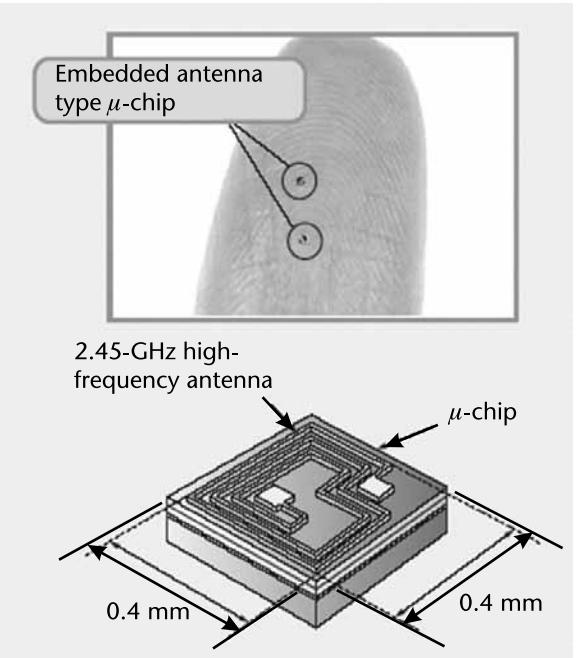


Figure 3.15 A 2.45-GHz antenna imbedded in a chip.

Chinese jails have used RFID to upgrade security by fitting its 6,000 prisoners with RFID wristbands. The new prisoner identification program at Jiangsu Longton Jail in the Nanjing Jiangsu province uses 13.56-MHz ISO 15693 tags and readers. Each wristband contains a smart label with encrypted data including the inmate's name, ID number, and security level. As inmates move around the prison, they present their wristband to a reader that records their identity and the time they entered and left a particular area. Guards are equipped with handheld readers for real-time spot checks and roll calls. Prisoners are required to check into the system for roll call at various times during the day, and if an inmate is not recorded by the system at the appropriate time an alarm indicates which prisoner has not been identified. Guards can view that prisoner's last check-in point and locate him or her quickly.

To deal with security threats posed by the volume of containers shipping in the United States, the U.S. Customs Service is proposing the Container Security Initiative (CSI) to identify high-risk containers and secure them with tamper-detection systems. The initiative aims to expedite processing of containers prescreened at points of embarkation in overseas megaports participating in the initiative.

The CSI's basic goal is to first engage the ports that send the highest volumes of container traffic into the United States, as well as the governments in these locations, in a way that will facilitate detection of potential problems at the earliest possible time. To meet this requirement, high-end RFID tags could periodically monitor electronic seals on the containers during transit. This class of application requires tags that can integrate sensor management electronics, such as analog-to-digital converters, and digital data interfaces. Tampering can also be detected in real time, and the tags, as the lowest level of a multitier architecture, can relay data to alert the shippers or customs authorities of tampering as it occurs.

Similarly, tags are used extensively to monitor transport of the high-value goods in the United States as well as worldwide. For example, Norway is a major exporter of salmon, so the RFID tags record the temperature in the containers so that the buyer can verify product freshness. This can be especially important when the shipments are bound for southern locations such as Italy, Spain, or North Africa. The biosensor will have sensitive biological film coatings that will undergo changes in material properties on contact with target pathogens like *Salmonella* and *Escherichia coli*[6].

Ski resorts are using smart label technology in lift tickets and passes. They not only authenticate the ticket as genuine and valid for that date and time, but also increase the traffic-flow rate on the lifts since passes can be read on the fly. The reader signals if there is a problem with a particular ticket. Skiers pass quickly through lift gates without having to remove their gloves to swipe a card, making their experience more pleasurable.

Many nightclubs use a similar system for tagging their patrons and making it easy to exit and reenter the club at any time during the night.

6. *Escherichia coli* (commonly abbreviated *E. coli* and named after Theodor Escherich) is a Gram-negative rod-shaped bacterium that is commonly found in the lower intestine of warm-blooded organisms. Most *E. coli* strains are harmless, but some can cause serious food poisoning in humans and are occasionally responsible for product recalls.

3.6 Other Developments in AutoID Systems

3.6.1 RuBee

RuBee is the commercial name for what is officially known as long-wavelength ID (LWID) as defined by the IEEE. The moniker was given to the technology by engineers at Miami-based Visible Assets, who coined the name for LWID technology after the hit 1967 Rolling Stones' song, "Ruby Tuesday."

RuBee networks operate at long wavelengths and accommodate low-cost radio tags at ranges to 100 feet. RuBee networks and tags are distinguished from most RFID tags in that they are unaffected by liquids and can be used underwater and underground. RuBee devices will be able to be used as implantable medical sensors having a 10- to 15-year battery life, depending on the number of reads and writes. The ability of RuBee tags to maintain performance around steel, so they work well when steel shelves are present, removes a key obstacle for low-cost deployment of RFID in retail, item-level tracking environments.

Tags based on the RuBee technology can be either active or passive, and all operate at a low frequency of 132 kHz, rather than the HF (13.56 MHz) or UHF (916 MHz) ranges used in the most widely deployed RFID systems today. Because low frequencies have significantly smaller bandwidth for data transfer, relative to higher frequencies, only about 6 to 10 RuBee tags (active or passive) can be read per second, while several dozen passive HF tags and several hundred passive UHF tags can be interrogated in that same span of time. This is not foreseen as a problem because RuBee users are interested in tracking the locations of assets rather than the passage of fast-moving tagged goods through portals, for which passive UHF tags are optimized.

The new IEEE RuBee standard, P1902.1 "RuBee Standard for Long Wavelength Network Protocol," will allow for networks encompassing thousands of radio tags operating below 450 kHz and represents candidate specification, which the standards group is using as a starting point for the future standardization of the technology.

IEEE P1902.1 will offer a real-time, tag-searchable protocol, using IPv4 addresses and subnet addresses linked to asset taxonomies that run at speeds of 300 to 9,600 baud. RuBee networks are managed by a low-cost Ethernet enabled router. Individual tags and tag data may be viewed as a stand-alone Web server from anywhere in the world. Each RuBee tag, if properly enabled, can be discovered and monitored over the World Wide Web using popular search engines (e.g., Google).

RuBee is the only wireless technology presently approved by the U.S. Department of Energy (DoE) for use in high security, top-secret areas in which RFID, Wi-Fi, Bluetooth, and Zigbee have all been banned [26].

In July 2006, the FDA classified 1902.1 as a Non-Significant Risk (NSR) Class 1 device in medical visibility applications. In May 2007, a peer-reviewed study was published by the Mayo Clinic showing that RuBee has no effect on pacemakers or implantable cardioverter defibrillators. A second Mayo study has shown RuBee has no electromagnetic interference or electromagnetic compatibility in the operating room. RuBee tags are also expected to meet the Intrinsically Safe ANSI 913-88 standard in a Zone 0 or Zone 1 explosive atmosphere as well as the DoD HERO (Hazards of Electro Magnetic Radiation to Ordinance) standards.

3.6.2 Visible Light Tags

Visible light tags use visible light instead of radio waves for the communication between tags and readers. Visible light tags can, for example, be used in a hospital where the use of radio waves is restricted or used underwater. The read range and scope can be modified by using optical devices, and people can visually see the communication range/scope since it uses visible light.

They are upper compatible with ISO15693, can replace part of existing RFID infrastructure, and could be used in hospitals, underwater, and other places where radio communication does not work. Obviously, visible light tags are not restricted by the radio interference, so there is no need to obtain a license, and actually they can be installed in an environment that is already crowded with RFID readers.

3.6.3 RFID and Printable Electronics

The term *printable electronics* refers to circuitry created out of conductive inks using a wide variety of printing technologies, old and new [27]. Much of the buzz in printable electronics is about ink-jet printing offering cost-effective device creation in very small volumes. But there are many other ways of creating printable electronics. These include: nanoimprint lithography (NIL), offset lithography, gravure, and flexographic printing.

Different applications may be suited to different printing technologies, since each application has its own requirements and each printing technology has its own advantages and disadvantages. For example, ink-jet printing holds out the prospect of the economic creation of customized/small production run circuits, while conventional printing processes using masks are well suited to something closer to mass production. Nanoimprint lithography, a technique that may or may not be considered part of the printed electronics sector, is probably the only production approach considered today that might genuinely be considered to be capable of creating features at the nanoscale.

The ability to cost-effectively create circuitry as part of other printed products is finding early use in greeting cards, but could become a big revenue generator through the success of printed RFIDs and smart packaging. It is possible to imagine some future smart packaging product, entirely created by multiple in-line printing processes, that encompasses high quality graphics, RFIDs, sensors, and even a small display to show pricing as it varies.

3.6.4 RFID and Mobile Phone Integration

RFID can be combined with mobile phone technologies by inserting either a transponder or a reader into a smart phone. A smart phone featuring a transponder connects to the wireless network and communicates with RFID readers, while a smart phone featuring a reader also connects to a wireless network but to retrieve RFID tag information. The smart phone transponders transfer their data to readers over the phone network.

Communication between transponder and reader must be secured by cryptographic means because RFID too poses the risks of virus and hacker attacks. As with RFID in general, communication between transponder and reader depends

on the range. All information is stored in the wireless network's central database. Smart phones with RFID readers can be used for detecting RFID-tagged products' Web sites via the wireless network for additional information on materials, origin, and so forth, locating the coordinates of anything or anyone that has been RFID tagged; for retrieving information from tagged items (computers, cars, furniture, ads); and for occasionally updating this information via the RFID reader.

RFID tags can also be used to call people. Combining RFID with cellular technology enables instant access to information concerning items, which leads to improved services, work efficiency, and performance. RFID provides direct, up-to-date data and diminishes manual typing of passwords, keys, and codes. Overall, RFID combined with mobile technology facilitates processes the same way RFID does in general.

3.7 Review Questions and Problems

1. If RFID has been around so long and is so great, why are only a few companies using it? Discuss at least two main reasons for the slow deployment rate of RFID systems.
2. Is RFID better than using bar codes? Will RFID ever completely replace bar codes?
3. Figure 3.16 represents flexible substrate. What is an importance of materials like this one in the development of the RFID technology?
4. Over the last 10 years, a concept of smart skin has been under development. This skin consists of an array of force sensing cells (i.e., sensels) that can measure the spatial distribution and magnitude of forces perpendicular to the sensing area. Tactile sensors have been the subject of extensive research for use in robotic applications.

   Recently, researchers have also developed a tattoo-like film that would allow doctors to monitor a patient's vital signs without the wiring and electrodes, just by using embedded electronic sensors in a film thinner than the

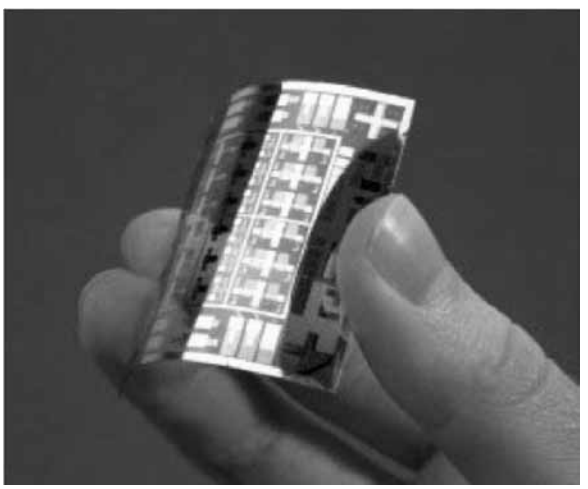


Figure 3.16 Flexible substrate.

diameter of a human hair and then placing it on a polyester backing. In addition to monitoring heart rate and temperature, the device could monitor brain waves, aid muscle movement, sense the larynx for speech, emit heat to help heal wounds, and perhaps even be made touch-sensitive and placed on artificial limbs.

Discuss different applications of this concept. What problems do you foresee in long term in medical applications? What about the use in AmI and other smart systems? How about sports (for example, intensity and duration of body stress)? How about the idea of the smart skin embedded in the floor and used to identify people by analyzing their footstep force profiles?

5. Using RFID in the pharmaceutical industry could prevent most of the 1.25 million adverse reactions and 7,000 patient deaths that occur annually in the United States as a result of drug errors, according to the *Meta Group Consultancy*. Discuss the issue.
6. Transponder systems equipped with one or more sensors are able to measure continuously or time-discreet physical values such as temperature, pressure, or acceleration. Measurements and data storage outside the electromagnetic field of the reader are only possible if an energy source such as a battery or solar cell is available.

   A hard hat, as shown in Figure 3.17, worn by building workers or motorcyclists, has been supplied with an RF tag and a sensor network with three acceleration sensors [28]. The maximum acceleration values and their distribution can be measured in case of an accident. With a reader, it is possible to read out these measurements quickly. Try to envision other applications of wireless sensors where human lives could be protected and/or saved.

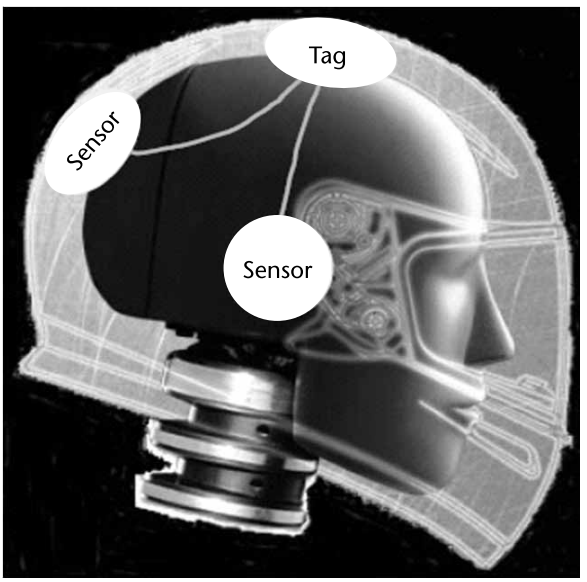


Figure 3.17 Motorbike helmet with sensor network.

7. There are four major frequency ranges at which RFID systems operate. As a rule of thumb, low-frequency systems are distinguished by short reading ranges, slow read speeds, and lower cost. Higher-frequency RFID systems are used where longer read ranges and fast reading speeds are required, such as for vehicle tracking and automated toll collection. Microwave requires the use of active RFID tags.

   Analyze Table 3.1 and answer the following questions:

   a. You are planning an RFID system for tracking the location of people within the building. What frequency band(s) would you consider for this application? Elaborate your answer.

   b. You are planning an RFID system for counting a large number of fast-moving and cheap products on the conveyer belt. What frequency band(s) would you consider for this application? Elaborate your answer.

   c. You are planning to deploy RFID system, but the environment already contains a large WLAN network with many users. What RFID bands may not be suitable for your system? Elaborate your answer.

8. Generally speaking, people are concerned about the social consequences of a world full of embedded wireless implants, tags and readers. WBANs are a part of AmI; monitoring and tracking of implanted devices outside of the designed purpose are a critical issue, as they provide details about the actual person.

   The WSN and BAN are the necessary technology for the development of the concept of AmI where users (patients) are provided services depending on their context. Society may need laws to specify who can access personal data logs and for what purpose. In Europe, the Data Protection Act already limits access to computer records of this kind, and the United States should probably enact similar legislation.

   What are your thoughts on the application of AmI for medical purposes? Should people be concerned with unexpected and undesired side effects in the attempt to improve human lives using the latest technology? Should we even continue with the development of the AmI? Write a short essay and provide arguments that support your answer.

Table 3.1 RFID Summary Table

Frequency	*Range*	*Tag Cost*	*Applications*
Low frequency (125–148 kHz)	3 feet	$1+	Pet and ranch animal ID; car key locks
High frequency (13.56 MHz)	3 feet	$0.50	Library book ID; clothing identification; smart cards
Ultrahigh frequency (915 MHz)	25 feet	$0.50	Supply chain tracking: box, pallet, container, trailer tracking
Microwave (2.45 GHz)	100 feet	$25+	Highway toll collection; vehicle fleet ID

References

[1] www.rolandmoreno.com (accessed May 7, 2011).

[2] Tedjini, S., et al., "Antennas for RFID Tags," *Joint sOc-EUSAI Conference*, Grenoble, France, 2005.

[3] Griffin, J., "A Radio Assay for the Study of Radio Frequency Tag Antenna Performance," Georgia Institute of Technology, Fall 2005.

[4] *RFID Compendium & Buyer's Guide 2004-2005,* AUTO ID Service Providers Ltd., on behalf of AIM U.K.

[5] Fuhrer, P., et al., "RFID: From Concepts to Concrete Implementation," University of Fribourg, Department of Informatics, 2005.

[6] Molnar, D. A., "Security and Privacy in Two RFID Deployments, with New Methods for Private Authentication and RFID Pseudonyms," Research Project, Department of Electrical Engineering and Computer Sciences, University of California Berkeley, 2006.

[7] www.icao.int/Pages/default.aspx, last accessed December 2011.

[8] Griffin, S., and C. Williams, "RFID Futures in Western Europe," White Paper, Juniper Research, 2005.

[9] Dulman, S. O., "Data-Centric Architecture for Wireless Sensor Networks," Ph.D. Dissertation, University of Twente, 2005.

[10] Rouvroy, A., "Privacy, Data Protection, and the Unprecedented Challenges of Ambient Intelligence," *Studies in Ethics, Law, and Technology*, Vol. 2, Issue 1, Article 3, The Berkeley Electronic Press, 2008, http://www.bepress.com/selt/vol2/iss1/art3 (accessed August 24, 2010).

[11] Chiarugi, F., et al., "Ambient Intelligence Support for Tomorrow's Health Care: Scenario Based Requirements and Architectural Specifications of the eu-DOMAIN Platform," 2006.

[12] Fernández, L., et al., "Wireless Sensor Networks in Ambient Intelligence," *Technologies for Health and Well-Being*, Instituto ITACA, Universidad Politécnica de Valencia, 2007.

[13] Nakashima, H., et al., *Handbook of Ambient Intelligence and Smart Environments*, New York: Springer-Verlag, 2009.

[14] Gaggioli, A., et al., "From Cyborgs to Cyberbodies: The Evolution of the Concept of Techno-Body in Modern Medicine," *Psychology Journal,* Vol. 1, No. 2, 2003, pp. 75–86.

[15] Lewis, F. L., "Wireless Sensor Networks," Research Head, Advanced Controls, Sensors, and MEMS Group, Automation and Robotics Research Institute, The University of Texas at Arlington, 2004.

[16] http://www.nist.gov/el/isd/ieee/ieeedocuments.cfm (accessed May 7, 2011).

[17] Martin, A., et al., "Intelligent Vehicle/Highway System: A Survey, Part 1," Florida International University, Department of Mechanical Engineering, 2000.

[18] Chon, H. D., et al., "Using RFID for Accurate Positioning," *Journal of Global Positioning Systems*, Vol. 3, No. 1-2, 2004, pp. 32–39, White Paper.

[19] White Paper, "Item-Level Visibility in the Pharmaceutical Supply Chain: A Comparison of HF and UHF RFID Technologies," Philips Semiconductors, TAGSYS, Texas Instruments Inc., July 2004.

[20] Miller, L. E., et al., "RFID-Assisted Indoor Localization and Communication for First Responders," National Institute of Standards and Technology (NIST), 2005.

[21] H.R. 5631, "Defense Appropriation Act for Fiscal Year 2007," June 16, 2006, http://www.congress.gov.

[22] http://www.rfida.com/apps/dodrfid.htm, last accessed December 2011.

[23] Swedberg, C., "US Army Developing RFID System to Track Weapons Usage," *RFID Journal*, November 9, 2006.

[24] Hoyt, S., et al., "A Tree Tour with Radio Frequency Identification (RFID) and a Personal Digital Assistant (PDA)," *IECON'03*, 2003.

[25] Usami, M., "An Ultra-Small RFID Chip: μ-Chip," Central Research Laboratory, Kokubunji City, Tokyo, 185-8601, Japan Hitachi Technology, 2004-2005.

[26] An Oracle White Paper, "An Introduction to RuBee Technology," An Oracle & Visible Assets Inc., January 2010.

[27] NanoMarkets LC, "Printable Electronics; Roadmaps, Markets and Opportunities," Executive Summary, September 2005.

[28] Fischer, W. J., et al., "Smart RF-Transponder Chips with On-Chip or External Sensors for Usage in Portable Systems," Department of Electrical Engineering, Dresden University of Technology, Dresden, Germany, 2003.

CHAPTER 4

RFID Standards Development Challenges

4.1 Regional Regulations and Spectrum Allocations

The growing need for interoperable products demands both the harmonization of regulatory controls on spectrum usage and international standards to support compatibility or interoperability of RFID systems. These are important factors when considering identification and data capture solutions that are required to operate in different countries and/or require compatibility for the purposes of data capture and transfer.

RFID data carriers (tags) and associated systems are generally considered to be part of a general category of radio-based short-range devices (SRDs) designed to operate in regions of the electromagnetic spectrum that do not require operating licenses and do not incur operating fees. The use of the electromagnetic spectrum, particularly for radio usage, is carefully controlled and spectrum allocations are specified to avoid systems interfering with one another. The license-exempt regions of the spectrum are generally allocated for industrial, science, and medical (ISM) applications. Unfortunately, the regulations for spectrum usage vary from country to country.

For the user of RFID systems, the expectation is that the necessary regulatory requirements will be satisfied. Although transparent to the user, the manufacturers of RFID systems will need to have a detailed understanding of the requisite regulations and associated standards and be able to demonstrate to users that their products are compliant.

Generally speaking, the manufacturers are required to identify and comply with the essential regulations and standards in respect of:

- Spectrum allocations and associated operating constraints;
- Health and safety;
- Electromagnetic compatibility;
- Avoidance of interference with other spectrum users;
- Compliance with national interface regulations;
- Other regulations and directives that may arise from time to time concerning system usage.

Manufacturers may also be required to carry out essential testing for compliance purposes. The necessary requirements to be met will be available from the spectrum management authorities in respective countries.

For low frequency (<135 kHz) and high frequency (13.56 MHz), there is a fair degree of allowed usage worldwide for RFID purposes. However, at the UHF carrier frequency, the situation is more complicated. Differences between the United Kingdom and the United States in the allocation of mobile phone usage, for example, prevent the use of common carrier frequencies and associated bands.

Country-specific regulatory controls specify the field strengths and power levels allowable for devices and systems operating at the different carrier frequencies. These levels naturally have a determining influence upon the ranges that are achievable for the reactively coupled and propagation coupled systems. It is therefore important to establish and confirm what is allowable within the country in which the technology is to be used.

The standardization process and efforts to agree regionalization of frequency usage are likely, in the fullness of time, to provide a foundation for wider open systems exploitation. It must be stressed that considerations for usage of RFID in different countries should include up-to-date information on spectrum allocation and constraints on usage obtained from the appropriate regulatory authority within the country concerned, as well as any encompassing regulations operating at union and international levels.

It is understood that the use of any modern technology requires some form of harmonization at the national and international levels. Standardization ensures compatibility and interoperability between different manufacturers and technical applications. This chapter will deal with the standards landscape and present areas for future standardization initiatives. In addition to the spectrum-related initiatives, RFID standardization is needed in the following areas:

- *Air interface and protocols*: Communication between tags and readers, readers and readers, RFID systems and other wireless communication systems;
- *Data structures*: Organization of the data (e.g., on a tag);
- *Conformance*: Tests ensuring that products meet the standards;
- *Applications*: Use of the technology for a particular purpose.

The importance of standards cannot be overemphasized for technologies, like RFID, that have universal relevance and significant potential for open systems usage. In addition, RFID systems generate radio signals that can interfere with other radio applications. Therefore, they must comply with regulations that state which frequencies may be used and the maximum power of transmission.

In Europe, the European Radiocommunications Committee (ERC) regulates the use of frequencies. In the United States, radio devices have to comply with licensing regulations by the FCC's licensing regulations. RFID systems are not separately regulated, but they need to comply with regulations based on the frequency used. The regulation of other countries often corresponds to either the U.S. or the European regulations. One exception is Japan, which has its own regulation system. However, it is often enough just to consider the FCC rules of the United States and the European ETSI standards.

4.2 Key Players in RFID Standardization

Various players are involved in working with and standardizing RFID technologies. This section aims to present some major players in RFID standardization activities and spectrum allocation.

The Automotive Industry Action Group (AIAG), an association of 1,600 members involved in the automotive and truck manufacturing supply chain, has been developing RFID specifications for the automotive industry. A general standard exists (ARF 156: Application Standard for RFID Devices in the Automotive Industry) and is accompanied by several standards dealing with special sectors of the automobile supply chain, such as AIAG B-1157, a standard to identify tires and wheels with RFID. AIAG B-11 has been developed together with EPCglobal. AIAG also holds regular conferences on the use of RFID in the automotive industry.

The European Article Numbering (EAN) and the Uniform Code Council (UCC), which administers and manages the EAN·UCC standards system in the United States and in Canada, launched the Global Tag Initiative (GTAG) in March 2000. It is a standard that covers UHF RFID technology and data formats. The air interface aspects of GTAG have now been merged with ISO 18000 Part 6. EAN International and UCC started EPCglobal as a joint venture.

At a European level, the European Radiocommunication Office (ERO) has been working on radiocommunications policies and frequency allocation that are important to RFID technologies. The ERO is the permanent office that supports the Electronic Communications Committee (ECC). This committee, in turn, is the telecommunications regulation committee for the European Conference of Postal and Telecommunications Administrations (CEPT). The ERC Decision ERC/DEC (01)04 of 2001 addresses the use of nonspecific SRDs, such as RFID, in certain UHF frequency bands.

Other important CEPT regulations related to RFID technologies include CEPT/ERC 70-03 (relating to the use of SRDs), CEPT T/R 60-01 (low-power radiolocation equipment for detecting movement and for alert, EAS), and CEPT T/R 22-04 ("Harmonisation of Frequency Bands for Road Transport Information Systems, RTI").

The European Telecommunications Standards Institute (ETSI) has also been very active in the field of RFID standardization. In 2004, ETSI TG34, the technical group for electromagnetic compatibility and radio spectrum matters, completed, in cooperation with EPCglobal, the standard ETSI EN 302 208-xx[1], for example, ETSI EN 302 208-1 V1.3.1 (2010-02): "Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Radio Frequency Identification Equipment operating in the Band 865 MHz to 868 MHz with power levels up to 2 W; Part 1: Technical requirements and methods of measurement." It allows readers to use more power and operate in a wider UHF band. Prior to that, ETSI had already developed the ETSI EN 300-220 standard.

1. ETSI EN 302 208-2 V1.3.1 (2010-02) is a document that includes improvements to the previous version of the standard that take advantage of technical developments within the RFID industry. In particular, this includes the ability for multiple interrogators to transmit simultaneously on the same channel. This provides significant improvements in spectrum efficiency and system performance. As a consequence, listen-before-talk is no longer a requirement.

The International Air Transport Association (IATA) is studying RFID technologies for airline baggage management; a subgroup of the Baggage Working Group (BWG) is responsible for this. IATA has already adopted 13.56 MHz and ISO/IEC 15693 in its "Recommended Practice for Airline Baggage RF Identification" (RP1745).

Within the framework of the International Civil Aviation Organization (ICAO), the Technical Advisory Group on Machine Readable Travel Documents (TAG/MRTD) has been working on international travel documents (e.g., a passport or visa) containing eye- and machine-readable data. Specifications for the design of these travel documents are contained in ICAO Doc 930376. An annex to that document contains provisions on contactless integrated circuits, referring in particular to the ISO/IEC 14443 standard.

Within the International Committee for Information Technology Standards (INCITS), the Technical Committee T6 deals with RFID standardization. T6 has developed NCITS 256, an RFID standard for use in item management. INCITS is accredited by the American National Standards Institute (ANSI).

The ISO and the International Electrotechnical Commission (IEC) have undertaken major activities to standardize RFID technologies in various areas. The work on RFID standardization in ISO and IEC has been carried out mainly under ISO/IEC Joint Technical Committee 1 (JTC 1) subcommittees 17 (Identification Cards and Personal Identification) and 31 (Automatic Identification and Data Capture Techniques). Inside of SC31, WG481 is responsible for RFID for Item Management. Other groups within the ISO framework that work on RFID topics include ISO TC 23/SC 19 (Agricultural Electronics), ISO TC 104/SC 4 (Identification and Communication), and ISO/TC204 (Transport Information and Control Systems).

The International Telecommunication Union (ITU) is the United Nations' (UN) specialized agency for telecommunications. Two of its three sectors are dealing with RFID-related issues: the Radiocommunication Sector (ITU-R) and the Telecommunication Standardization Sector (ITU-T). As global spectrum coordinator, ITU-R plays an essential role in the management of the radio-frequency spectrum. It is governing the use of the radio spectrum by some 40 different services around the world. ITU-R Study Group 185 is responsible for spectrum management. The Telecommunication Standardization Sector (ITU-T) creates globally agreed and globally accepted ICT standards.

The Universal Postal Union (UPU) is the UN specialized agency for cooperation between postal services. The UPU Technical Standards Board has been developing standards to use RFID technologies in postal applications for several years. Among these standards are: S25-1G ("Data Constructs for the Communication of Information on Postal Items, Batches and Receptacles"), S23-1 Part A,B,C,G (RFID) and Radio Data Capture (RDC) Systems), UPU RF 0001.2 ("Identification and Marking Using RFID Technology: Data Schemes"), UPU Snn-1 ("Identification and Marking Using RFID Technology: Reference Architecture and Terminology"), and UPU Snn-3 ("Identification and Marking Using RFID: System Requirements and Test Procedures").

4.3 ISO and EPC Approach

Standards and specifications may be set at the international, national, industry, or trade association level, and individual organizations may call their own specifications standards. Many industry standards and specifications set by individual organizations are based on international standards to make implementation and support easier and to provide a wider choice of available products. Standards can be applied to include the format and content of the codes placed on the tags, the protocols and frequencies that will be used by the tags and readers to transmit the data, the security and tamper-resistance of tags on packaging and freight containers, and applications use.

The ISO and the EPCglobal have both been leading figures in this debate. The ISO has their 18000 standard and the EPCglobal Center has introduced the EPC standard. Wal-Mart has decided to use the EPC standard, where the DoD wants to use the EPC for general purposes, but use the ISO standard for air interface. This means a lot of pressure on the ISO and EPCglobal to come to some kind of an agreement.

What is EPC? The Auto-ID Center has proposed a new electronic product code (EPC) as the next standard for identifying products. The goal is not to replace existing bar code standards, but rather to create a migration path for companies to move from established standards for bar codes to the new EPC. To encourage this evolution, we have adopted the basic structures of the Global Trade Item Number (GTIN), an umbrella group under which all existing bar codes fall. There is no guarantee that the world will adopt the EPC, but the proposal already has the support of the Uniform Code Council and EAN International, which are the two main bodies that oversee international bar code standards. Other national and international trade groups and standards bodies are working together.

The ISO is based in Geneva, and its standards carry the weight of law in some countries. All ISO standards are required to be universal around the world, so users of ISO RFID standards will not have to worry if their systems comply with the different regulations on frequencies and power output for each country where they do business. The ISO is very active in developing RFID standards for supply chain operations and is nearing completion on multiple standards to identify items and different types of logistics containers. As appropriate, EPCglobal submits its standards to ISO for review and ratification as ISO considers EPC as being a subset of its standards.

Some in the industry anticipate a future where everyday objects and appliances are connected to the Internet, and each other, via RFID tags. This so-called Internet of things will enable people to interact more with their environment, and their belongings with each other. EPCglobal, the not-for-profit standards organization, is driving the development of a universal electronic product code system and a global information network to enable automatic identification of items in the supply chain.

EPC standard is mainly backed by the U.S. vendors while ISO standard more widely supported and accepted in Europe. Japan has UID (ubiquitous ID), while China has announced its intent to create its own RFID standards. Even though EPC has the strongest commercial support, the true global standards are still under development and will be subjected to geopolitical forces.

Appropriate standards allowing numerous companies to create interoperable products are a key prerequisite to widespread use of RFID tags. ISO/IEC 15693 forms part of a series of international standards that specify a contactless smart card. It is titled "Identification Cards-Contactless Integrated Circuit(s) Cards—Vicinity Cards"[2] and has three parts: Part 1 (physical characteristics), Part 2 (air interface and initialization), and Part 3 (anticollision and transmission protocol). It specifies a 13.56-MHz RFID protocol, originally proposed by Texas Instruments and Philips Semiconductors in 1998, defining data exchange between RF tags and readers, and collision mediation when multiple tags are in a reader's RF field. Compliance guarantees that RF tags and readers using the ISO/IEC 15693 protocol will be compatible across companies and geographies.

4.4 RFID Systems and Frequencies

4.4.1 Power Emissions Conversion

Before discussing the details of the regulations, note that there are several accepted means of describing one of the most important regulated parameters of an RF device: its radiated emissions. All license-exempt devices have some limitation on the amount of output power, or radiated energy, that they can produce. What differs among the various regulatory agencies is the means of describing this limit. Radiated energy can be described in terms of:

- Electrical field strength (E) measured at some distance from the radiator;
- Effective isotropic radiated power (EIRP);
- Effective radiated power (ERP).

The electrical field strength (E) is perhaps the most precise way of describing the actual RF energy present at a point in space that a receiving antenna could use. Since the RF energy decreases with increasing distance from the transmitting antenna, the regulatory limits based on electrical field strength are specified at a specific distance from the transmitting antenna. While the electrical field strength may be precise, it is often not as useful from a design perspective as the effective isotropic radiated power (EIRP).

The EIRP is the power that would have to be supplied to an ideal antenna that radiates uniformly in all directions in order to get the same electrical field strength that the device under test produces at the same distance; such an antenna is called an *isotropic radiator*. Given a distance r from the transmitting antenna, the $EIRP$ can be calculated from E using the following formula:

$$EIRP = 10\log\frac{4\pi E^2 r^2}{0.377} = 10\log\frac{E^2 r^2}{0.03}\ [\text{dBm}] \tag{4.1}$$

2. Vicinity cards are cards that can be read from a greater distance as compared to proximity cards. ISO/IEC 15693 systems operate at the 13.56-MHz frequency and offer a maximum read distance of 1–1.5m. As the vicinity cards have to operate at a greater distance, the necessary magnetic field is less (0.15 to 5 A/m) than that for a proximity card (1.5 to 7.5 A/m).

where the *EIRP* is expressed in dBm, *E* is the electrical field strength in V/m, *r* is distance in meters, and 0.377 is expressed in V^2 (unit of measurement.) Contrary to the European regulations, the U.S. regulations are in most cases specified in terms of field strength, not power. The field strength shall be measured at 3m from the device under test.

The effective radiated power (ERP) is similar to the EIRP. It is the power that would have to be supplied to a half-wave dipole to get the same electrical field strength that the device under test produces at the same distance. A half-wave dipole represents more realistic antenna than an isotropic radiator. Since a half-wave dipole radiates more energy in some directions and less in others, it is said to have antenna gain in the direction of the most energy.

The amount of antenna gain is usually expressed in decibels relative to an isotropic radiator, or dBi. The maximum gain of a half-wave dipole is 2.15 dBi. Thus, in the direction of maximum gain, the amount of power required to produce the same electric field is less with a half-wave dipole than it is with an isotropic radiator (Table 4.1). If both the ERP and the EIRP are expressed using a logarithmic scale, such as dBm, it holds that:

$$ERP = EIRP - 2.15 \text{ dB} \tag{4.2}$$

or

$$EIRP = 1.64\ ERP \tag{4.3}$$

If one of the three parameters (EIRP, ERP, or E) at a distance *r* is given, the other two can be calculated. These reference parameters are used to define the permitted radiated power in RF systems, and they are often quoted in RFID reader and tag specifications.

In the United States, engineers tend to use EIRP while in Europe engineers more commonly use ERP.

4.4.2 North American and International Frequency Bands

The RF spectrum is a scarce and shared resource, used nationally and internationally, and subject to a wide range of regulatory oversight. In the United States, the FCC is a key regulatory body that allocates spectrum use and resolves spectrum conflicts. The ITU is a specialized agency of the United Nations, which plays the same role internationally.

Table 4.1 ERP/EIRP Conversion Example

	EIRP	*Gain*	*ERP Power Fed to the Antenna*
Isotropic antenna	4 W	1	4 W
Dipole antenna	4 W	1.64	2.44 W
Antenna	4 W	3	1.33 W

Development of RFID systems suitable for international deployment faces some considerable problems as a result of disparate frequency regulations and the transmitting power of the reader. It is not only that the frequency bands have been allocated differently in different countries, but the permitted transmitting power for the reader also varies, which means that identical models can differ considerably in their range.

The frequency ranges of 125 kHz and 13.56 MHz are generally regarded as having been largely standardized. The standardization process for the remaining frequency ranges is being carried out on the basis of the ISO/IEC 18000 standards.

Both the U.S. and EU regulatory agencies place limitations on the operating frequencies, output power, spurious emissions, modulation methods, and transmit duty cycles, among other things. RFID tags and readers fall under the category of SRDs, in which, although they do not normally require a license, the products themselves are governed by the laws and regulations which vary from country to country. Today, the only globally accepted frequency band is the HF 13.56 MHz.

For passive UHF RFID the problem is much more complicated, as frequencies allocated in some countries are not allowed in others, due to their proximity to already allocated bands for devices such as mobile phones and alarms. This discontinuity has resulted in the ITU dividing the world into three regulatory regions (Figure 4.1):

- *Region 1:* Europe, Middle East, Africa, and the former Soviet Union, including Siberia;
- *Region 2:* North and South America and Pacific Rim east of the International Date Line;
- *Region 3:* Asia, Australia, and the Pacific Rim west of the International Date Line.

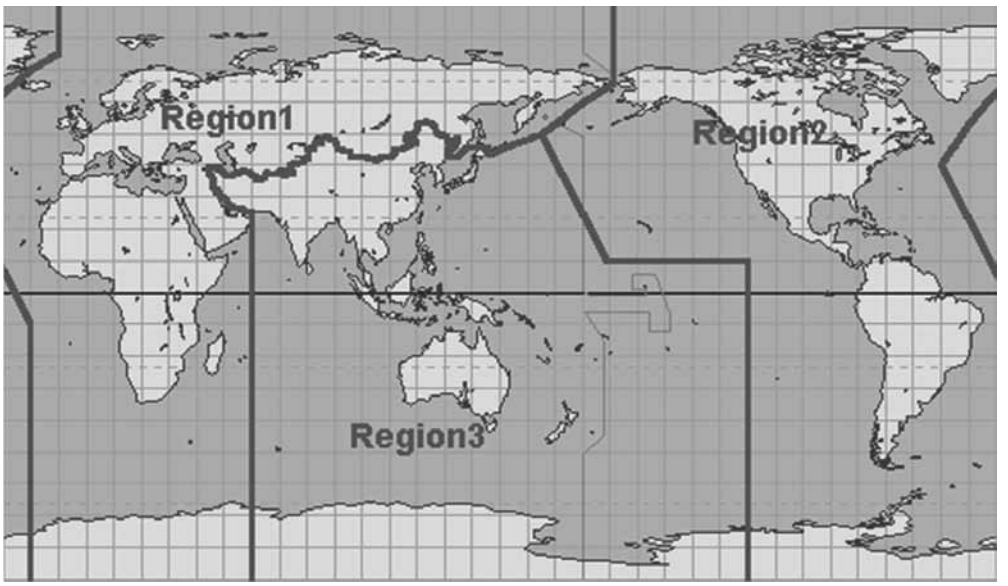


Figure 4.1 ITU regions.

The main regulatory bodies in different regions are:

- In the United States, the FCC;
- In Europe, the CEPT;
- In Japan, the Ministry of Public Management, Home Affairs, Posts and Telecommunication (MPHPT).

4.4.3 RFID Interoperability and Harmonization

When considering interoperability for global use of RFID devices, it is first necessary to consider spectrum allocation. Perhaps the most essential aspect in worldwide acceptance of RFID is to have spectrum available in all relevant markets so that RFID can be used where needed [1]. RFID systems in the United States, as with other ubiquitous RF devices used by the general public, are license-exempt.

In the United States, RF radiation from intentional and unintentional license-exempt radiators is regulated by the FCC under Part 15 of Title 47 of the *Code of Federal Regulations*. These regulations delineate the technical specifications, such as allowable frequency, power limits, and other operational constraints, under which an intentional radiator may be operated without an individual license. Part 15 also allows operation of RFID systems over a broad range of frequencies, but places limits on the allowable output power of the system.

Every available operating frequency has a specific power limit associated with it. The combination of frequency and allowable power level are the factors that dictate the functional range of the particular RFID application, whether over a range of centimeters or hundreds of meters. These allowable power levels are particularly relevant for the power output of the readers, which have a far higher power output than the tags.

Country-specific regulatory controls specify the field strengths and power levels allowable for devices and systems operating at the different carrier frequencies. These levels naturally have a determining influence upon the ranges that are achievable for the reactively coupled and propagation coupled systems. It is therefore important to establish and confirm what is allowable within the country in which the technology is to be used.

In addition to power limits, there are restrictions on the use of certain bands by license-exempt devices to prevent potentially harmful interference to systems used for services such as safety, search and rescue, aeronautical communications, and scientific research. It should also be noted that license-exempt systems operating in United States under Part 15 must accept any interference from other systems in these bands, including interference from other license-exempt devices.

Many countries do not have rules for license-exempt systems like the United States. Instead, they allocate spectrum on a primary or secondary basis and require that all radio transmitters be licensed by the government. In addition, some countries that allow RFID currently do not recognize active tags within their regulations. Similar to the United States, several other countries do not allocate spectrum specifically for RFID but to categories of service such as SRDs.

In 2006, the ITU-R released a recommendation SM.1538-2[3] outlining the spectrum requirements and regulatory approaches applicable to SRD in Europe, the United States, China, Japan, and South Korea. Table 4.2 contains a partial list of countries currently using RFID and the frequency bands allowed for RFID operations.

Each frequency has advantages and disadvantages relative to its capabilities. Generally, a lower frequency means a lower read range and a slower data read rate but increased capabilities for reading near or on metal or liquid surfaces.

There are exceptions in terms of allowable frequency and functional use of the bandwidth for critical operations both in the United States and in other countries. For example, the allowable bandwidth and power for RFID devices are not generally the same from ITU region to region. In ITU Region 1, which covers most of Europe, the industrial, scientific, and medical (ISM) radio band (13.553–13.567 MHz) is much narrower than the bandwidth that is allowed in the United States for RFID.

Also, use of the band at 433 MHz for active RFID in the United States is intended for container tracking and is allowed at higher power levels than generally permissible for license-exempt devices in this band. However, operations of such systems are limited primarily to industrial locations, such as railheads and shipyards, and the systems must be registered with the FCC.

Table 4.2 RFID Operational Frequencies

Band	*Frequency*	*System*	*Regions/Countries*
Low frequency (LF)	125–134 kHz	Inductive	United States, Canada, Japan, and Europe
High frequency (HF)	13.56 MHz	Inductive	United States, Canada, Japan, and Europe
Very high frequency (VHF)	433.05–434.79 MHz	Propagation	In most of Europe, United States (active tags at certain locations must be registered with the FCC), and under consideration in Japan
Ultrahigh frequency (UHF)	865–868 MHz	Propagation	Europe, Middle East, Singapore, Northern Africa
Ultrahigh frequency (UHF)	866–869 and 923–925 MHz	Propagation	South Korea, Japan, New Zealand
Ultrahigh frequency (UHF)	902–928 MHz	Propagation	United States, Canada, South America, Mexico, Taiwan, China, Australia, Southern Africa
Ultrahigh frequency (UHF)	952–954 MHz	Propagation	Japan (for passive tags)
Microwave	2.4–2.5 and 5.725–5.875 GHz	Propagation	United States, Canada, Europe, Japan

3. ITU-R is presently exploring whether the traditional approach to spectrum management, allocating different bands to different services, needs reconsideration, in light of the fact that hybrid services are becoming common and one channel can now carry a many different services (thanks to TCP/IP and packetization), while frequency-agile equipment allows for dynamic/flexible band sharing. This work may also include revising Recommendation ITU-R SM.1538-2: "Technical and operating parameters and spectrum requirements for short range radiocommunication devices," which is now withdrawn.

In Europe, however, the 433.05–434.79-MHz band is an ISM band and is allocated on a primary basis to SRDs such as RFID. In addition, many nations in East Asia (e.g., China, Japan, and South Korea) are currently developing their own regulations for RFID; for example, Japan is in the process of revising its regulations to allow the 950–956-MHz band to be used for license-exempt, low-power, passive tag RFID systems. They are also establishing a license structure for high-power (power levels up to any level not hazardous to humans at specified distances) passive RFID systems that will be used in industrial areas.

In general, RFID chips can be used to track products grouped in various hierarchies:

- Individual items or single packages containing multiple items for consumer purchase;
- Cartons or cases of multiple items;
- Pallets of multiple cartons or cases;
- Loads (e.g., truckloads, shiploads, or railcar loads) of multiple pallets.

The products at each of these levels may be assigned an RFID label that is associated with information pertaining to at least one adjacent hierarchical level. For example, an RFID label on a pallet may be associated in a database with the RFID labels for each carton on the pallet, or may be associated with data pertaining to the RFID label from the truckload.

Figure 4.2 illustrates needs for different requirements for different types (layers) of RFID tagging. The RF power, the frequency used, and therefore the reader-tag distance will be different for different applications.

Although some of the ISM frequency bands are internationally recognized (e.g., 13.56 MHz and 2.45 GHz), others are not. In the United States, 915 MHz is recognized as an ISM band, while 433 MHz is not; however, in Europe the 915 MHz is not recognized, but 433 MHz is. Differences like this add to the challenges of harmonization of RFID bands.

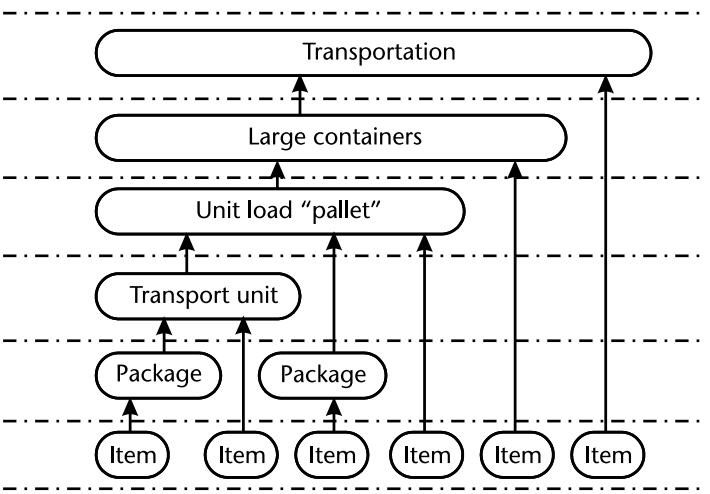


Figure 4.2 Different layers of RFID tagging.

In recognition of the complications associated with international harmonization of frequency bands, the DoD requires that passive RFID systems be both multimode and multiband to meet global requirements. Accordingly, the DoD has issued a set of policies mandating system performance and functionality for implementing RFID systems within the DoD supply chain. Additionally, it specifies operation in the 860–960-MHz frequency range and requires that passive RFID systems be capable of operating in this frequency range.

4.4.4 Advantages and Disadvantages of Using 125-kHz Frequency

The 125/134.2-kHz low-frequency RFID tags are known as being robust, durable replacements for bar code or other fragile identification technologies. The 125 kHz is used in many industrial applications since it is the most "metal friendly" of all RF frequencies.

Some examples of industrial applications are process controls (manufacturing floor), automated guided vehicles (AGVs), authentication/verification (printers, UV bulbs, fluids, paints), vehicle identification, personnel ID, and specialized markets (oil and gas pipe, pharmaceutical), and so on. Advantages and disadvantages of this band are summarized next.

The advantages of the 125-kHz RFID are:

- Works well in metal surroundings;
- Works well around the objects with the high water content;
- Anticollision (ability to read more than on tag at a time) rarely necessary;
- Low sensitivity to interference;
- Simple and inexpensive but sufficient for such applications as animal ID and car immobilizers.

The disadvantages of the 125-kHz RFID are:

- Short read range: up to 5 feet, although in most cases only few inches;
- Low read rate;
- Limited EEPROM memory size;
- Tags are usually read-only tags (simple tags with hardwired code), although read/write tag also exist;
- Passive (inductive) tags are most common in this frequency band.

4.4.5 Advantages and Disadvantages of Using the 13.56-MHz Frequency

Much of the current activities today revolve around an RFID frequency of 13.56 MHz. The reason for this attention is that there are several RFID technology providers who have designed their product offerings based on this particular frequency; 13.56 MHz is, after all, an ISO standard for smart card applications.

Smart cards differ from RFID tags in that they are used for more than data storage. In some smart card applications, the card itself contains a microprocessor that can define parameters for data storage and byte allocation.

Smart card applications require that certain RFID attributes be present for implementations of that technology to be successful. Since 13.56 MHz provides for RFID near-field read/write capability, it is ideally suited to smart card applications in which secure financial transactions are being transmitted. Understandably, a smart card user would not want his or her bank account number to be able to be read at a distance of 3–5m. Why are these attributes important? If a technology, or specific frequency in this case, was targeted for a particular application (smart cards), then it may not have the necessary attributes that would be needed for another RFID application, such as item management.

The 13.56-MHz inductive passive RFID systems are one of the mainstreams RFID products, and some of the main advantages (and disadvantages) of this band are listed as follows.

The advantages of the 13.56-MHz RFID are:

- Frequency band available worldwide as an ISM frequency;
- Well suited for applications requiring reading small amounts of data and minimal distances;
- Popular smart card frequency;
- ISO 15693, ISO 14443, and HF EPC standardization for the air interface;
- Robust reader-to-tag communication;
- Excellent immunity to environmental noise and electrical interference;
- Well-defined and localized label interrogation zones;
- Minimal shielding effects from adjacent objects and the human body;
- Penetrates water/tissue well;
- Freedom from environmental reflections that can affect UHF and microwave systems;
- Good data transfer rate;
- High clock frequency and synchronous subcarrier;
- On-chip capacitors for tuning transponder coil can be easily realized;
- Uses normal CMOS processing, cheap ICs, and disposable tags;
- Cost-effective antenna coil manufacturing;
- Low RF power transmission so EM regulation compliance does not cause problems;
- No user licenses for reader systems required (license-exempt ISM band);
- Possible to use the systems in industrial and in hazardous environments with potential for explosive substances.

The disadvantages of the 13.56-MHz RFID are:

- Government-regulated frequency (United States and Europe recently harmonized);
- Does not penetrate or transmit around metals;
- Large antennas (compared to higher frequencies);

- Larger tag size than those on higher frequencies;
- Tag construction requires more than one surface to complete a circuit;
- Reading range of less than 1.0m (3 feet).

4.4.6 Operation in the 433-MHz Band

The 433-MHz band has rarely been used for RFID applications in the past, but it seems that is starting to draw more attention lately. Major applications using this frequency are cargo handling, container locations, Real-Time Location Systems (RTLS), and asset tracking.

This band is only allocated as an ISM band in ITU Region 1 (including Europe). The 433-MHz RFID systems are subject to certification under ETSI EN 300 220. ISO 18000-7 is the standard for the air interface for the 433-MHz RFID and ISO 24730-3 is the related 433-MHz RTLS standard.

The 433-MHz systems use active tags because of the low power allowance of 10 mW. Selecting an optimal RF for operation of an active RFID system requires consideration of several factors, including technical performance, regulatory issues, and coexistence with other technologies. The 433 MHz has been selected as the optimal frequency for global use of active RFID from a broad range of radio frequencies against these parameters and also a frequency used for ISM appliances [2].

Two key technical performance parameters of an active RFID system are directly related to the frequency of operation: maximum communication range and propagation within crowded environments. Frequencies between 100 MHz and 1 GHz offer the best technical performance in terms of range for active RFID. The use of active tags provides ranges of hundreds of meters outdoors, making this technology useful in the very large outdoor facilities used for storage and trans-shipment.

Implementation of active RFID in this band for some years has also shown that 433-MHz active RFID can be used without interfering with other systems in the same band.

The advantages of the 433-MHz RFID are:

- Works well in RF-impaired environments;
- Affordable components plus deployment;
- No special environmental configuration;
- Multiple tag designs to match asset groups and application;
- Supports a high density of tags;
- 10–15-foot accuracy suitable for most applications;
- Instant detection provides perimeter control capability.

The disadvantages of the 433-MHz RFID are:

- High price if used as active tags;
- Increased size and weight;
- A need for maintenance (changing batteries, for example).

4.4.7 Operation in the 900-MHz Band

Traditionally, passive transponders operate at 125 kHz or 13.56 MHz using coils as antennas. These transponders operate in the magnetic near field of the base station's coil antenna, and their reading distance is typically limited to less than 1.0m (3 feet). A problem of these systems is the low efficiency of reasonably large antennas at such low frequencies.

Due to great demand for higher data rates, longer reading distances, and small antenna sizes, there is a strong interest in UHF frequency band RFID transponders, especially for the 868–915-MHz bands. The frequency band of 902–928 MHz is one of the ISM bands in the United States (FCC, Part 15.247 regulations) and Canada, commonly abbreviated as the 915-MHz ISM band. In this band, there are no restrictions to the application or the duty cycle as there were intended for the control and periodic applications only. Furthermore, the allowed power output is considerably higher.

Because of the lack of restrictions and higher allowed power, the 915-MHz ISM band is very popular for license-exempt short range applications including audio and video transmission. FCC section 15.249 allows 50 mV/m of electrical field strength at 3-m distance in the frequency band of 902–928 MHz. This corresponds to an EIRP of –1.23 dBm. The harmonics limit is one hundredth of the fundamental level, 500 μV/m, corresponding to an EIRP of –41.23 dBm [3].

The frequency of operation for the reader-to-tag communication in passive UHF RFID is not fixed. The reader does a frequency hopping in the ISM band in UHF for communicating with the tags. The frequency hopping avoids interference that might occur due to other devices using some part of the ISM band's spectrum. Also, the modulation schemes used in the reader-to-tag communication depend on the type of the protocol being read.

The U.S. readers will have a maximum output power of +30 dBm (1W), the European readers will have a maximum output power of +27 dBm, and the Japanese readers will have a maximum output power of +30 dBm.

A new requirement for a dense reader mode is being introduced in the United States that will be used when there are a high number of readers in close proximity. Most 1-W readers to date in the United States are designed for low-density usage, and their emissions and out-of-band requirements are defined by the FCC. For the U.S. and Japanese readers, a saturated power amplifier can be used, which allows for higher power amplifier efficiency (PAE) on the order of 50%. Regulations require the European and dense reader mode readers to operate the power amplifier in the linear range, which reduces PAE to approximately 30%.

Even higher output power can be used if the system employs some form of spread spectrum such as frequency hopping or direct-sequence spread spectrum. The reason such allowances are made is that spread-spectrum systems are less likely to interfere with other systems than are single-frequency transmitter, and they are often more immune to interference from other systems.

For more up-to-date details on UHF regulations in Europe, see publication ERC REC 70-73 [4].

The advantages of the 915-MHz RFID are:

- License-exempt ISM band in United States;
- Read range of up to 20 feet (6m);
- Significantly higher data rates than LF RFID;
- Widespread use;
- Can cover dock door portals up to 9 feet wide; therefore, suitable for inventory tracking applications including pallets and cases.

The disadvantages of the 915-MHz RFID are:

- Absorbed by liquids, although Gen2 tags and antenna designs are less susceptible;
- Unpredictable performance near metal, although Gen2 tags and antenna designs less susceptible;
- Interferes with existing bands in some countries, but receiving certification;
- Common sources of interference include 900-MHz cordless phones, older 900-MHz wireless LANs, metal supports, equipment, and cabinets;
- As a consequence of sometimes unpredictable radio propagation due to reflections, some tags may be successfully read from large distance away from the reader, while neighboring tags may receive little power from the reader.

Efficient antennas are most readily constructed with dimensions on the order of a half of the wavelength, in this case around 6 inches (150 mm), an inconveniently large size for many desirable applications.

It is possible to compress the linear dimension by bending the conductive regions, or using large-area structures, though inevitably with some compromise in radiation resistance and thus performance. At these frequencies tags are also sensitive to environmental effects, and optimal designs may change depending on the material to which the tag is to be attached.

4.4.8 Operation in the 2.45- and 5.8-GHz Bands

The basic operating principle of microwave 2.45-GHz RFID systems is energy and data transmission using propagating radio signals (E-field transmission). This is exactly the same principle as used in long-range radio communication systems.

An antenna of the reader generates a propagating radio wave, which is received by the antenna in the tag. A passive power tag converts the signal to DC voltage to supply the tag with energy. Data transmission from the reader to the tag is done by changing one parameter of the transmitting field (amplitude, frequency, or phase). The return transmission from the tag is accomplished by changing the load of the tag's antenna (amplitude and/or phase).

In this context, the microwave, 13.56-MHz and 125-kHz systems use the same principle. For microwave RFID systems, this method is called *modulated backscatter*. Alternatively, a signal of different frequency can be generated, modulated, and transmitted to the reader. The latter types of systems are referred to as *active RF*

transmitter tags. The 2.4–2.4835-GHz band is another ISM band covered by FCC sections 15.247 and 15.249.

The advantages of the 2.4-GHz RFID are:

- The 2.4-GHz band is a worldwide license-exempt band, and this is an important advantage compared to, for example, the 902–928-MHz band;
- The 2.4-GHz band also has a wider bandwidth than the 902–928-MHz band, which means more available channels;
- Good reflections off metal surfaces allow better propagation in cluttered environments;
- It has reasonable propagation through nonconductive materials, such as wood and wood-based products, natural and synthetic garments, and plastics;
- Because tags operating in the E-field do not require antennas with extremely low impedances, inexpensive flexible antennas able to withstand considerable bending are achievable.

The disadvantages of the 2.4-GHz RFID are:

- The active components are more expensive and have higher current power consumption;
- It has reduced propagation distance for the same power;
- Moisture and moisture-containing substances can exhibit energy absorbing mechanisms at microwave frequencies;
- For ranges in excess of 1.0m (3 feet), multipath effects and fading need to be considered;
- It shares spectrum allocation with spread-spectrum radios, microwave ovens, Bluetooth, WLANS, TV devices, and so forth;
- It is more susceptible to electronic noise than lower UHF bands, for example, 433-MHz and 860–930-MHz bands;
- Regulatory approvals are still in process.

For single-frequency or other systems that do not qualify as spread-spectrum, the same transmit power limits as in the 902–928-MHz band apply. The only difference is that an averaging detector can be used in the 2.4-GHz band, allowing a higher peak output power, with the limitation that the peak electrical field strength must not be more than 20 dB above the average value. Thus, similar to the control and periodic applications described earlier, the transmitting strength can be up to 20 dB larger than the limits for continuous signals if the duty cycle is reduced accordingly.

As with the 902–928-MHz band, larger transmitting power levels are allowed if the spread spectrum is used. The criteria for a frequency-hopping system in the 2.4-GHz ISM band are:

- The transmitter hops pseudo-randomly between at least 15 nonoverlapping frequency channels.
- The average time of occupancy at any frequency must not be larger than 0.4 second within a time period of 0.4 second multiplied by the number of channels.

The permitted peak transmit power measured at the antenna input of a frequency-hopping system with at least 75 hopping frequencies is +30 dBm. For systems with less than 75 but at least 15 hopping frequencies, a peak transmitting power of +21 dBm is allowed. Similar to the 902–928-MHz band, the power has to be reduced if the isotropic antenna gain is larger than 6 dBi. This allows a maximum EIRP of +36 dBm in a system with at least 75 channels or +27 dBm in a system with less than 75 but at least 15 channels.

Systems operating at microwave frequencies may be considered to exhibit quasi-optical features with the facility to form well-defined beams. By exploiting these spatial directivity features, systems can be configured for interrogating defined areas. However, costs for transponders in the category are generally higher, by a factor of 2 or more, than lower-frequency devices.

The antenna structures often used for microwave tags yield a directional beam, in contrast with the near-spherical field patterns associated with lower-frequency antenna structures. Each antenna structure will exhibit a primary forward lobe and, depending upon antenna dimensions in relation to the wavelength, a number of side lobes. Because of these features, the direction in which the antennas are deployed (directivity) in relation to a readable tag will influence the range and ability to detect the tag.

Today 2.45-GHz tags are available in many different shapes and with different functionality, influenced by applications and its requirements. Unlike inductive RFID tags, which require substantial surface area, many turns of wire, or magnetic core material to collect the magnetic field, UHF and microwave tags can be very small, requiring length in only one dimension. Thus, in addition to a longer range over the inductive systems, the UHF and microwave tags are easier to package and come in a wider variety of configurations. Tag lengths of 20 to 100 mm (1–4 inches) are typical. The tag's thickness is limited only by the thickness of the chip as the antenna can be fabricated on thin, flexible materials.

The 5.8-GHz band offers the advantages of the less congested band, resulting in less interference (the longest range is up to 1,000m), but it has a number of disadvantages; for example, it is not available in the United States or many other countries, orientation of the antennas is very important, and chips are difficult and expensive to build. Microwave RFID has been effectively used in special applications where the directionality, range (particularly with active tags), and very fast data transfer features are required. These areas include asset tracking, factory automation (particularly in automotive manufacturing), toll systems, and barrier-based access control.

In the United States, the FCC has been requested to provide a spectrum allocation of 75 MHz in the 5.85–5.925-GHz band for intelligent transportation services (ITS).

4.5 ISO/IEC 18000: RFID Air Interface Standards

4.5.1 About the 18000 Standards

ISO/IEC 18000 is a series of standards being created by ISO/IEC/JTC creating RFID air interface standards for the item management. ISO/IEC 18000 has been developed in order to:

- Provide a framework to define common communications protocols for internationally usable frequencies for RFID and, where possible, to determine the use of the same protocols for all frequencies such that the problems of migrating from one to another are diminished;
- Minimize software and implementation costs;
- Enable system management and control and information exchange to be common as far as is possible.

The standard contains the following parts:

- 18000-1 Part 1; Reference architecture and definition of parameters to be standardized;
- 18000-2 Part 2; Parameters for air interface communications below 135 kHz;
- 18000-3 Part 3; Parameters for air interface communications at 13.56 MHz;
- 18000-4 Part 4; Parameters for air interface communications at 2.45 GHz;
- 18000-5 Part 5; Parameters for air interface communications at 5.8 GHz (withdrawn due to the lack of global interest);
- 18000-6 Part 6; Parameters for air interface communications at 860 to 960 MHz;
- 18000-7 Part 7; Parameters for air interface communications at 433 MHz.

As can be seen, each of these parts deals with a different aspect of RFID. The first part is the defining document that explains how the standard works, and the rest are divided by frequency. All of the parts of ISO/IEC 18000 have been published, but work items are currently open to revise parts 6 and 7.

4.5.2 ISO/IEC 18000-1:2008

ISO/IEC 18000-1:2008 defines the generic architecture concepts in which item identification may commonly be required within the logistics and supply chain and defines the parameters that need to be determined in any standardized air interface definition in the subsequent parts of ISO/IEC 18000.

The subsequent parts of ISO/IEC 18000 provide the specific values for definition of the air interface parameters for a particular frequency or type of air interface, from which compliance (or noncompliance) with ISO/IEC 18000-1:2008 can be established. In addition, it provides description of example conceptual architectures in which these air interfaces are often to be utilized. ISO/IEC 18000-1:2008 is an enabling standard that supports and promotes several RFID implementations

without making conclusions about the relative technical merits of any available option for any possible application.

4.5.3 ISO/IEC 18000-2:2009

ISO/IEC 18000-2:2009 defines the air interface for RFID devices operating below 135 kHz. The purpose of ISO/IEC 18000-2:2009 is to provide a common technical specification for RFID devices that can be used by ISO committees developing RFID application standards.

ISO/IEC 18000-2:2009 is intended to allow for compatibility and to encourage interoperability of products in the international marketplace. ISO/IEC 18000-2:2009 defines the physical layer used for communication between the interrogator and the tag and further defines the communications protocol used in the air interface.

Two types of tag are defined by ISO/IEC 18000-2:2009: Type A and Type B, which differ only by their physical layer. Both support the same inventory (anticollision) and protocol. Type A tags are permanently powered by the interrogator, including during the tag-to-interrogator transmission, and operate at 125 kHz. Type B tags are powered by the interrogator, except during the tag-to-interrogator transmission, and operate at 125 kHz or 134.2 kHz.

4.5.4 ISO/IEC 18000-3:2010

It is an international standard for passive RFID item-level identification and provides physical layer, collision management system, and protocol values for RFID systems for item identification operating at 13.56 MHz in accordance with the requirements of ISO/IEC 18000-1.

ISO/IEC 18000-3:2010 has three modes of operation, intended to address different applications. The modes, although not interoperable, are noninterfering.

4.5.5 ISO/IEC 18000-4:2008

It defines the 2.45-GHz protocols that support ISO/IEC 18000-1. Each of the specific physical/data link configurations is defined in a separate subclause. The configuration descriptions include a physical layer and a data link layer.

ISO/IEC 18000-4:2008 defines the air interface for RFID devices operating in the 2.45-GHz ISM band used in item management applications. ISO/IEC 18000-4:2008 provides a common technical specification for RFID devices that can be used by ISO committees developing RFID application standards. ISO/IEC 18000-4:2008 is intended to allow for compatibility and to encourage interoperability of products for the growing RFID market in the international marketplace.

ISO/IEC 18000-4:2008 defines the forward and return link parameters for technical attributes including, but not limited to, operating frequency, operating channel accuracy, occupied channel bandwidth, maximum EIRP, spurious emissions, modulation, duty cycle, data coding, bit rate, bit rate accuracy, bit transmission order, and, where appropriate, operating channels, frequency hop rate, hop sequence, spreading sequence, and chip rate. ISO/IEC 18000-4:2008 further defines the communications protocol used in the air interface.

ISO/IEC 18000-4:2008 contains two modes. The first is a passive tag operating as an interrogator talks first, while the second is a battery-assisted tag operating as a tag talks first.

4.5.6 ISO/IEC 18000-6:2010

It defines the air interface for RFID devices operating in the 860–960-MHz ISM band used in item management applications. It provides a common technical specification for RFID devices that can be used by ISO committees developing RFID application standards. ISO/IEC 18000-6:2010 is intended to allow for compatibility and to encourage interoperability of products for the growing RFID market in the international marketplace.

The standard defines the forward and return link parameters for technical attributes including, but not limited to, operating frequency, operating channel accuracy, occupied channel bandwidth, maximum EIRP, spurious emissions, modulation, duty cycle, data coding, bit rate, bit rate accuracy, bit transmission order, and, where appropriate, operating channels, frequency hop rate, hop sequence, spreading sequence, and chip rate. It further defines the communications protocol used in the air interface.

ISO/IEC 18000-6:2010 specifies the physical and logical requirements for a passive-backscatter, interrogator-talks-first (ITF) or tag-only-talks-after-listening (TOTAL) RFID system. The system comprises interrogators and tags, also known as labels. An interrogator receives information from a tag by transmitting a continuous-wave (CW) RF signal to the tag; the tag responds by modulating the reflection coefficient of its antenna, thereby backscattering an information signal to the interrogator. The system is ITF, meaning that a tag modulates its antenna reflection coefficient with an information signal only after being directed to do so by an interrogator, or TOTAL, meaning that a tag modulates its antenna reflection coefficient with an information signal upon entering an interrogator's field after first listening for interrogator modulation in order to determine if the system is ITF or not.

ISO/IEC 18000-6:2010 contains one mode with four types. Types A, B, and C are ITF. *Type A* uses pulse-interval encoding (PIE) in the forward link and an adaptive ALOHA collision-arbitration algorithm. *Type B* uses Manchester coding in the forward link and an adaptive binary-tree collision-arbitration algorithm. *Type C* uses PIE in the forward link and a random slotted collision-arbitration algorithm. Type D is TOTAL based on Pulse Position Encoding or Miller M=2 encoded subcarrier.

ISO/IEC 18000-6:2010 specifies:

- Physical interactions (the signaling layer of the communication link) between interrogators and tags;
- Interrogator and tag operating procedures and commands;
- The collision arbitration scheme used to identify a specific tag in a multiple-tag environment.

This standard was revised and published in 2010 with the addition of a Type D and support for sensors and battery assist. The addition of these options to the

standard meant that the document had grown to almost 500 pages. A new work item to break this standard into smaller parts was approved at the end of 2010, and the committee met, for the first time since that approval, to review the work that has been done. The standard has been broken into five parts:

- Part 6: Parameters for air interface communications at 860 MHz to 960 MHz General;
- Part 61: Parameters for air interface communications at 860 MHz to 960 MHz Type A;
- Part 62: Parameters for air interface communications at 860 MHz to 960 MHz Type B;
- Part 63: Parameters for air interface communications at 860 MHz to 960 MHz Type C;
- Part 64: Parameters for air interface communications at 860 MHz to 960 MHz Type D.

4.5.7 ISO/IEC 18000-7:2009

It defines the air interface for RFID devices operating as an active RF tag in the 433-MHz band used in item management applications. It provides a common technical specification for RFID devices that can be used by ISO technical committees developing RFID application standards.

ISO/IEC 18000-7:2009 is intended to allow for compatibility and to encourage interoperability of products for the growing RFID market in the international marketplace. ISO/IEC 18000-7:2009 defines the forward and return link parameters for technical attributes, including, but not limited to, operating frequency, operating channel accuracy, occupied channel bandwidth, maximum power, spurious emissions, modulation, duty cycle, data coding, bit rate, bit rate accuracy, bit transmission order, and, where appropriate, operating channels, frequency hop rate, hop sequence, spreading sequence, and chip rate. ISO/IEC 18000-7:2009 further defines the communications protocol used in the air interface.

ISO/IEC 18000-7, Parameters for active air interface communications at 433 MHz, was last published in 2010. This standard has been opened to revise the technology to include multichannel utilization and more efficient communication techniques needed to address increased market and application needs for higher and more secure data transmissions. This revision will also specify the sensor interface and a universal mechanism to allow for other services to be enabled on the tag.

4.6 UHF and EPCglobal Gen 2

The EPC is a universal identifier for any physical object. It is used in information systems that need to track or otherwise refer to physical objects. Large subsets of applications that use the EPC also rely upon RFID tags as a data carrier.

Nevertheless, it is important to note that the EPC and RFID are not synonymous: EPC is an identifier, and RFID is a data carrier. RFID tags contain other data

besides EPC identifiers (and in some applications may not carry an EPC identifier at all), and the EPC identifier exists also in non-RFID contexts.

4.6.1 The EPC Class Structure

The RFID Class Structure, depicted in Table 4.3, provides a framework to classify tags according to their primary functional characteristics. The RFID Class Structure classifies tags as belonging to one of five classes: Class 1 (identity tags), Class 2 (higher functionality tags), Class 3 (semipassive tags), Class 4 (active ad hoc tags), or Class 5 (reader tags).

Each successive class within this framework builds upon, that is, is a superset of, the functionality contained within, the previous class, resulting in a layered functional classification structure. Class 1 forms the foundation of this framework [5]:

- *Class 1 identity tags* are designed to be the lowest-cost, minimum-usable-functionality tag classification. Identity tags are pure passive RFID tags that are expected to implement a resource discovery mechanism and store a unique object identifier only. The signaling and modulation defined for Class 1 tags are the foundation for all passive communication within this hierarchy.
- *Class 2 higher functionality tags* build upon the identity tag by providing more functionality, such as a tag identifier and read/write memory, while still maintaining a pure passive power and communication scheme.
- *Class 3 semipassive tags* add an on-tag power source, such as a battery, to their higher functionality foundation. Semipassive tags combine passive communication with an on-tag power source that enables a tag to operate without the presence of a passive tag reader (i.e., a Class 5 reader tag).
- *Class 4 active ad hoc tags* encompass the Class 3 semipassive tags and, in addition, are ad hoc networking devices that are capable of communicating with other Class 4 tags using active communication and with Class 5 reader tags using both passive and active communication. Because they may initiate communication, Class 4 tags are necessarily active. Functionally, these tags lie in the realm of ubiquitous computers or *smart dust*.
- *Class 5 reader tags* encompass the functionality of a Class 4 active ad hoc tag and are able to power and communicate with pure passive Class 1 and Class 2 tags and communicate with Class 3 tags via passive communication.

Table 4.3 The EPC Class Structure

Class 1	Read-only passive identity tags, no battery
Class 2	Passive tags with additional functionality, such as memory or encryption
Class 3	Semipassive tags (battery-assisted); may support broadband communication
Class 4	Active tags that may be capable of broadband, peer-to-peer communication with other active tags in the same frequency band and with readers
Class 5	Essentially reader tags; they can power other Class 1, 2, and 3 tags and also communicate with other Class 4 tags and with each other wirelessly

The main technological difference between passive and semipassive (battery-assisted) labels has to do with the source from which the tags receives the appropriate amount of power in order for the IC to reach the excitation level (or to *wake up*) and send signals back to the reader. Passive tags gather energy from the reader's signal, whereas semipassive tags contain an integrated power source and have no need to receive energy from the reader. This enables semipassive labels to outperform passive labels in reliability (up to 100% read and write rates), even for liquids, metals, and foils, and in increased ranges of up to 60 feet.

The EPC Tag Data Standard[4] defines the EPC and also specifies the memory contents of Gen 2 RFID Tags. In more detail, the Tag Data Standard covers two broad areas:

- The specification of the EPC, including its representation at various levels of the EPCglobal Architecture and its correspondence to GS1 keys and other existing codes.
- The specification of data that is carried on Gen 2 RFID tags, including the EPC, user memory data, control information, and tag manufacturer information.

4.6.2 UHF Gen 2

Developed by the EPCglobal Industry Group, the EPC Gen 2 standard[5] defines the physical and logical requirements for a passive-backscatter, ITF RFID system operating in the 860–960-MHz frequency range. *EPC Generation 2* is the latest standard for RFID tags, specifying the operation of the tag and the communication protocol for interoperability with EPC readers worldwide (see Table 4.4).

EPC Gen 2 was developed by a collaboration of leading RFID users and vendors, working through EPCglobal, a nonprofit trade group. EPCglobal is part of the UCC/EAN organization, which has long administered bar code and other standards around the world. The frequency used for UHF RFID systems varies between 860 and 960 MHz. UHF RFID systems operate at 915 MHz in the United States and 868 MHz in Europe, and they are being widely deployed due to RFID mandates from several large corporations, including international retailers, and

Table 4.4 UHF Systems Worldwide

	North America	*Europe*	*Singapore*	*Japan*	*Korea*	*Australia*	*Argentina, Brazil, Peru*	*New Zealand*
Band size (MHz)	902–928	866–868	866–869, 923–925	950–956	908.5–914	918–926	902–928	864–929 (parts)
Power	4-W EIRP	2-W ERP	0.5-W ERP, 2W (in upper band)	4-W EIRP	2-W ERP	4-W EIRP	4-W EIRP	0.5–4-W EIRP
Number of channels	50	10	10	12	20	16	50	Varied

4. The latest revision of the EPC Tag Data Standard, Version 1.5, was ratified on August 18, 2010.
5. The latest revision of the standard, V1.2.0, extends the item-level tagging capabilities of UHF Class-1 Generation-2 air interface protocol. In this protocol, three optional features have been added.

the DoD. In addition to retail, UHF systems are employed in various supply chain management applications.

Gen 2 (as well as ISO 18000-6) is a technical standard that specifies the air interface protocol or, in other words, how tags and readers communicate. The leadership of EPCglobal, along with collaborative efforts throughout the industry, has now set the stage for real-world implementation of Gen 2 technologies that are compliant with the newest EPCglobal RFID specifications for the UHF band.

While primarily intended for the supply chain market, the Gen 2 RFID systems may also be used in asset tracking, baggage tagging, manufacturing, and a wide assortment of other applications where long read range is required. The Gen 2 chip is intended for use in the manufacture of passive RFID tag products operating in the 860–960-MHz frequency band. Thus, Gen 2 addresses a very small, although very important, part of the total standards environment.

The Gen 2 protocol takes the best features of the Gen 1 Class 1, Gen 1 Class 2, and ISO protocols to make a new and improved standard. The Gen 2 promises to be a global, open, interoperable standard that incorporates the frequency and performance requirements for worldwide use.

Some of the features of Gen 2 are as follows:

- High read rate of 1,500 tags/second in North America, and 600 tags/second in Europe. This is especially important in countries where the narrow bandwidth limits the data rates to 30% of that can be achieved in the United States.
- Proven air interface with forward link PIE ASK, backscatter link FM0 or Miller-modulated subcarrier.
- Provides an operating mode in dense reader environments.
- Better read algorithms (bit mask filtering) that eliminate duplicate reads, allow tags to enter reader field late and still be read, and allow tags to stay quiet until asked to talk, thus making it faster to find a specific tag.
- Each tag has security enhanced with 32-bit password encryption and permanent kill capability.
- Write schemes enhance write speed function.
- In addition to the required EPC data, there is optional memory to support user-specific data.
- Meets global regulatory compliance standards.

4.6.3 Electronic Product Code Information Services

In September 2007, EPCglobal Inc. announced a new industry standard providing the capability for visibility into the movement, location and disposition of assets, goods and services throughout the world. EPC Information Services (EPCIS)[6] allows for the seamless, secure exchange of data at every point in the life cycle of goods and services.

6. At the time of this writing, the latest revision of this specification was Version 1.0.1, approved on TSC September 21, 2007.

In October 2006, EPCglobal successfully completed interoperability testing of the platform along with 12 other large and small solution providers from Japan, Korea, and North America, including Auto-ID Labs, Avicon, BEA Systems, Bent Systems, IBM, Globe Ranger, IIJ, NEC, Oracle, Polaris Systems, Samsung, and T3Ci. The interoperability test marked a significant milestone in the development of EPCIS, which is the result of years of effort by more than 150 companies and organizations participating in the EPCIS working group. The positive results of this test and solution provider support have led to the ratification of this standard.

EPCIS, by providing a standard set of interfaces for EPC data, enables a single way to capture and share information, while still allowing the flexibility for industry and organization-specific implementations. The specification supports business cases and consumer benefits such as container tracking, product authentication, promotions management, baggage tracking, and electronic proof of delivery, chain of custody, returns management, and operations management.

4.6.4 UHF RFID Tag Example

The EPC tag class structure is often misunderstood; *class* is not the same as *generation*. Class describes a tag's basic functionality, for example, whether it has memory or a battery, whereas generation refers to a tag specification's major release or version number. The full name for what is popularly called EPC Generation 2 is actually EPC Class 1 Generation 2, indicating that the specification refers to the second major release of a specification for a tag with write-once memory. Figure 4.3 and the description of the UHF RFID tag were taken from [6].

Antenna performance at UHF and microwave frequencies is dependent on the substrate, that is, thickness and electromagnetic properties (conductivity, permittivity, and permeability). It also depends on the conductive quality of electrodes. Low-cost constraints and diversity of RFID applications lead to consider and investigate nonstandard materials to be used for both tag and antenna. These investigations concern low-cost substrate material such as paper, plastic, and polymers (Figure 4.4).

The use of conductive inks is an alternative to usual electrodes made with standard conductors such as copper and aluminum. The performance of an antenna made with conductive ink is limited by the conductivity and the thickness of the deposited ink. To avoid excess loss and get good directivity, the thickness of the ink must be larger than skin depth (which is dependent on the conductivity and the frequency) [7].

The tag antenna is generally omnidirectional in order to ensure the identification in all directions. The structure of the tag antenna should also be as small as possible in size. Because of its simplicity and omnidirectionality, the $\lambda/2$ dipole is one of the most preferred forms. At UHF frequencies the typical size is 150 mm (6 inches), which is big. Usually the dipole is folded in order to reduce its size. This usually needs full-wave electromagnetic simulation in order to take into account the capacitive and inductive coupling introduced by the folded form.

Some typical specification parameters are shown in Table 4.5.

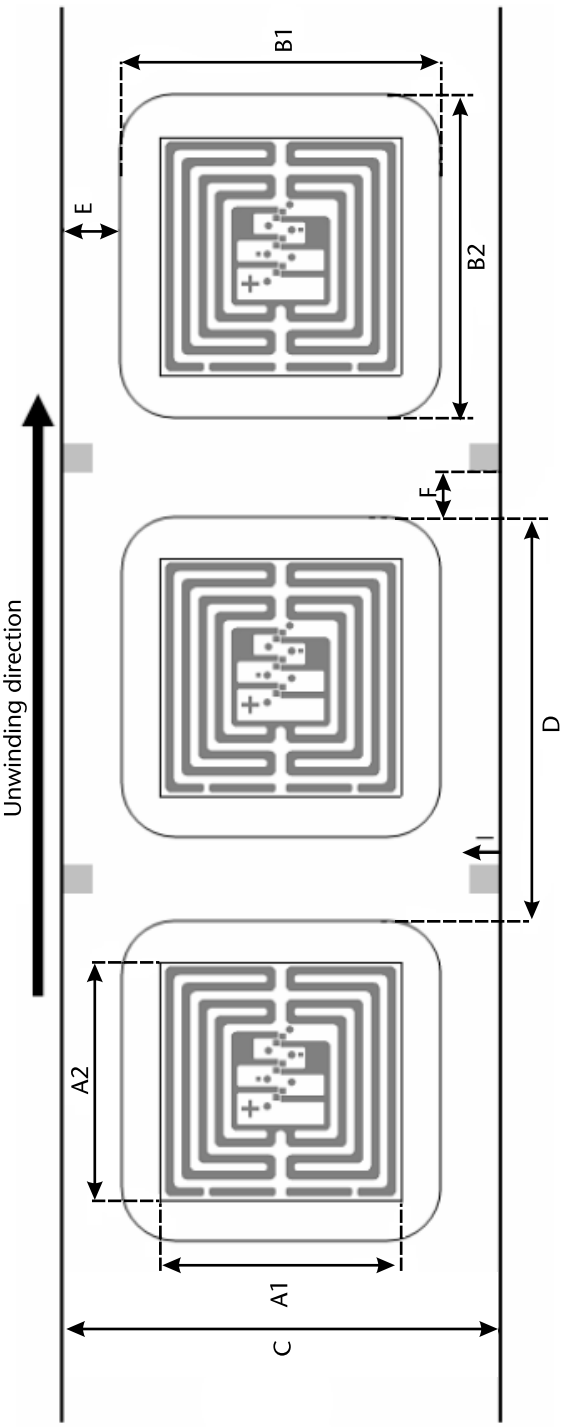


Figure 4.3 UHF RFID tag layout.

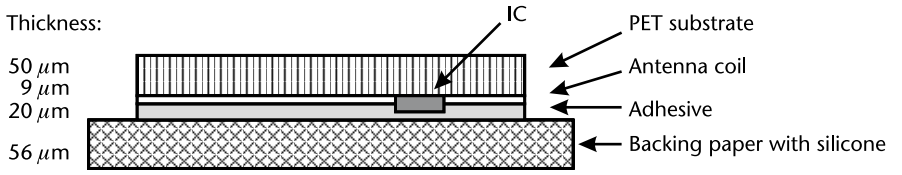


Figure 4.4 UHF RFID tag cross section.

Table 4.5 UHF RFID Specification

Integrated Circuit (IC)	96 bit EPC Class 1 Gen 2
Free air frequency	915 ± 15 MHz, loaded mode
Read Sensitivity	Min. 7.8 V/m
Operating temperature (electronics parts)	–40°C/+65°C
Thermal cycle resistance (electronics parts)	200 cycles – 40°C/+80°C
Temperature humidity resistance (electronics parts)	80°C, 85% relative humidity, 168 hours
ESD voltage immunity	±1 kV peak, HBM
Storage	(15–25)°C, 40–60% relative humidity, max. 2 years
Bending diameter (D)	Greater than 50 mm, tension less than 10 N
Static pressure (P)	Less than 10 MPa (10 N/mm^2)

4.7 Review Questions and Problems

1. Why is standardization very important in the RFID world? What are the most important standardization organizations in this arena?
2. What is a present status of harmonization of the UHF RFID standards?
3. What is the most widely used RFID frequency/system today? Explain why.
4. Explore the latest utilization and standardization status of the 2.4- and 5.8-GHz bands in RFID systems in the United States and the rest of the world. Why are microwave bands not more popular in RFID applications?
5. An antenna has a gain of 16 dBi (the term "dBi" is often used as a reminder that the directivity is with respect to the isotropic radiator), and the power delivered to the antenna is 100 mW. What is the effective isotropic radiated power in dBm and in watts? *(Answer: 36 dBm/4 W.)*
6. What frequency is used by UHF RFID devices in the United States?
   a. 902–928 MHz;
   b. 121–124 kHz;
   c. 2.4 GHz;
   d. 915–928 MHz.
7. The Spanish Post Office has implemented Europe's largest UHF RFID system in sorting centers in 16 cities across Spain. Reusable tags are inserted into an envelope and sent through the system to monitor the movement of letters as well as the system's real-time performance. How would you

conceptually (without going into detail) design a system like that in a country you live in?

8. You want an RFID tag that supports longer-distance communications and does not rely on the reader to provide power to the tag. What kind of tag do you need?
   a. Passive tag
   b. Semipassive tag
   c. Active tag
   d. Powered tag.
9. In 2009, the DASH7 Alliance was formed to create wireless technology that extends the ISO 18000-7 standard for low power wireless data transfer [8]. The goal of the group is to provide interoperability between RFID devices and applications.

   The DoD has standardized on DASH7; recently the DoD awarded a $428 million contract for devices that comply with the ISO 18000-7 (DASH7) standard, making DASH7 the standard for active RFID devices throughout the DoD. DASH7 operates in UHF band at 433 MHz, giving the system good penetration capabilities, global availability, low interference, good range, and acceptable data rates of up to 28 kbps.

   Research and list a few large companies that are members of the DASH7 Alliance. Describe the DASH7 alignment with complementary technologies such as cellular, passive RFID, Wi-Fi, and 2-D bar code. Discuss importance of the DoD embracing this specific technology.
10. In many countries, SRDs are license-exempt. Fears have been expressed that SRDs are increasing in quantity and diversity so fast that existing interference protection rules for licensed radio services may not be adequate. This could eventually lead to tighter restrictions on license-exempt devices. What is your opinion?

References

[1] "Radio Frequency Identification—Opportunities and Challenges in Implementation," Washington, D.C., U.S. Department of Commerce, 2005.

[2] *Dynamic Evolution of RFID Market*, ECO Report 01, European Communications Office (ECO), August 31, 2010.

[3] Loy, M., et al., *ISM-Band and Short Range Device Regulatory Compliance Overview*, Application Report, SWRA048, Texas Instruments, May 2005.

[4] http://www.ero.dk/doc98/official/pdf/rec7003e.pdf (accessed May 8, 2011).

[5] Engals, D. W., and S. E. Sarma, "Standardization Requirements Within the RFID Class Structure Framework," Cambridge, MA: Auto-ID Labs, Massachusetts Institute of Technology, January 2005.

[6] www.upmraflatac.com (accessed May 19, 2011).

[7] Tedjini, S., et al., "Antennas for RFID Tags," *Joint sOc-EUSAI Conference*, Grenoble, October 2005.

[8] Burns, P., "Five Reasons DASH7 Will Transform Logistics," *Defense Transportation Journal*, September 2009.

CHAPTER 5

Components of the RFID System

5.1 RFID Engineering Challenges

An RFID system consists of an RFID reader, an RFID tag, and an information-managing host computer. The reader contains an RF transceiver module (transmitter and receiver), a signal processor and controller unit, a coupling element (antenna), and a serial data interface (RS232, RS485) to a host system.

The tag acts as a programmable data carrying device and consists of a coupling element (resonant tuned circuit) and a low-power CMOS IC. The IC chip contains an analog RF interface, antenna tuning capacitor, RF-to-dc rectifier system, digital control, Electrically Erasable and Programmable Read-Only Memory (EEPROM), and data modulation circuits.

RFID involves contactless reading and writing of data into an RFID tag's nonvolatile memory through an RF signal. The reader emits an RF signal and data is exchanged when the tag comes in proximity to the reader signal.

Tags can be categorized as:

- An *active* tag, which has a battery that supplies power to all functions;
- A *semipassive* tag, which has a battery used only to power the tag IC, and not for communication;
- A *passive* tag, which has no battery on it. The absence of a power supply makes passive tags much cheaper and more reliable than active tags.

Given the increase in RFID usage, many new challenges face design engineers. Currently, these challenges include multiple tag standards, 20% tag failure rate, installation and placement issues, the need for cost-effective management and maintenance of readers, the need for reductions in reader size that allows them to be imbedded into structures and handheld devices, and intellectual property protection and secure access control protocols.

A number of different parameters will influence a quality and reliability of the RFID system: tag size, reader/writer antenna size, tag orientation, tag operating time, tag movement velocity, effect of metallic substances and metal items on operating range, multi-tag operating characteristics, and effect of number of tags on operating success rate, tag overlapping, and so forth.

5.2 Near-Field and Far-Field Propagation

There are two main categories for RFID systems on the market today. These are *near-field* systems that employ inductive (magnetic) coupling of the transponder tag to the reactive energy circulating around the reader antenna, and *far-field* systems that couple to the real power contained in free-space propagating electromagnetic plane waves [1].

Near-field coupling techniques are generally applied to RFID systems operating in the LF and HF bands with relatively short reading distances, whereas far-field coupling is applicable to potentially long-range UHF and microwave RFID systems. Whether or not a tag is in the near or far field depends upon how close it is to the field creation system and the operating frequency or wavelength.

There is a distance, commonly known as the *radian sphere*, inside which one is said to be in the near field and outside of which one is said to be in the far field. Because changes in electromagnetic fields occur gradually, the boundary is not exactly defined; the primary magnetic field begins at the antenna and induces electric field lines in space (the near field).

The zone where the electromagnetic field separates from the antenna and propagates into free space as a plane wave is called the *far field*. In the far field, the ratio of electric field E to magnetic field H has the constant value of 120 or 377Ω. The approximate distance (5.1) where this transition zone happens is given as follows:

$$r = \frac{\lambda}{2\pi} \tag{5.1}$$

It is also important to notice that this expression is valid for small antennas where $D << \lambda$.

The *reactive near-field* region is a region where E- and H-fields are not orthogonal; anything within this region will couple with the antenna and distort the radiation pattern, so the antenna gain is not a meaningful parameter here.

Using (5.1), at 13.56 MHz (λ = 22m), this places the near-field to far-field boundary at about 3.5m (10 feet).

It has been estimated that the far-field distance (5.2) in case when $D > \lambda$ is given as follows:

$$r = \frac{2D^2}{\lambda} \tag{5.2}$$

where D is the maximum dimension of the radiating structure and r is the distance from the antenna. Note that this is only an estimate, and the transition from near to far field is not abrupt.

Typically, D for reader antennas is 0.3m (1 foot.) The far-field distance in UHF ISM band in the United States (λ = 0.33m) can be estimated to be 0.56m.

Generally speaking, the *radiating near-field* or *transition region* is defined as a region between a reactive near field and a far field. In this region, antenna pattern is taking shape but is not fully formed, and the antenna gain will vary with distance.

$$\frac{\lambda}{2\pi} < r < \frac{2D^2}{\lambda} \tag{5.3}$$

The solution of Maxwell's equations for the fields around an antenna consists of three different powers of the range: $1/r$, $1/r^2$, and $1/r^3$. At very short ranges, the higher powers dominate the solution, while the first power dominates at longer ranges. This can be interpreted as the electromagnetic wave breaking free from the antenna.

The near field may be thought of as the transition point where the laws of optics must be replaced by Maxwell's equations of electromagnetism.

5.2.1 Far-Field Propagation and Backscatter Principle

RFID systems based on UHF and higher frequencies use far-field communication and the physical property of backscattering or "reflected" power. Far-field communication is based on radio waves in which the reader sends a continuous base signal frequency that is reflected back by the tag's antenna. During the process, the tag encodes the signal to be reflected with the information from the tag (the ID) using a technique called *modulation* (i.e., changing the amplitude or phase of the waves returned) [2].

The concept of the radian sphere, which has a value for its radius of $r = \lambda/2\pi$, helps in the visualization of whether the tag coupling is in the near or far field. If the tag is inside this sphere, the reactive energy storage fields (dipolar field terms) dominate, and the near-field coupling volume theory is used. If the tag falls outside the sphere, then propagating plane wave EM fields dominate and the familiar antenna engineering concepts of gain, effective area or aperture, and EIRP are used. These often more familiar EM concepts, whereby real power is radiated into free space, are relevant to the cases of UHF and microwave tagging technologies.

Most theoretical analysis, at least in the first approximation, assumes the so-called free-space propagation. *Free space* simply means that there is no material or other physical phenomenon present except the phenomenon under consideration. Free space is considered the baseline (ideal) state of the electromagnetic field. Radiant energy propagates through free space in the form of electromagnetic waves, such as radio waves and visible light (among other electromagnetic spectrum frequencies). Of course, this model rarely describes the actual propagation accurately; phenomena such as reflection, diffraction, and scattering exist that disturb radio propagation.

In the wireless industry, most models and formulas we use today are semi-empirical, that is, based on the well-known radio propagation laws but modified with certain factors and coefficients derived from the field experience. RFID is definitely an area where this practice is required; short distances cluttered with multiple tags

and/or other objects are potential obstacles to radio propagation and will cause serious deviations, predictable or not, from the theoretical calculations.

A backscatter tag operates by modulating the electronics connected to the antenna in order to control the reflection of incident electromagnetic energy. For successful reading of a passive tag, two physical requirements must be met:

1. *Forward power transfer:* Sufficient power must be transferred into the tag to energize the circuitry inside. The power transferred will be proportional to the second power of the distance.
2. *The radar equation:* The reader must be able to detect and resolve the small fraction of energy returned to it. The power received will be reduced proportional to the fourth power of the distance.

5.2.1.1 Forward Power Transfer

A typical RFID tag consists of an antenna and an integrated circuit (chip), both with complex impedances. The chip obtains power from the RF signal transmitted by the RFID reader. RFID tag antenna is loaded with the chip whose impedance switches between two impedance states, usually high and low. At each impedance state, RFID tag presents a certain *radar cross section* (RCS). The tag sends the information back by varying its input impedance and thus modulating the backscattered signal.

In Figure 5.1 $Z_A = R_A + jX_A$ is the complex antenna impedance and $Z_C = R_C + jX_C$ is the complex chip (load) impedance; chip impedance may vary with the frequency and the input power to the chip.

The power scattered back from the loaded antenna can be divided into two parts. One part is called *structural mode* and is due to currents induced on the antenna when it is terminated with complex conjugate impedance. The second part is called *antenna mode* and is a result of a mismatch between antenna impedance and load impedance.

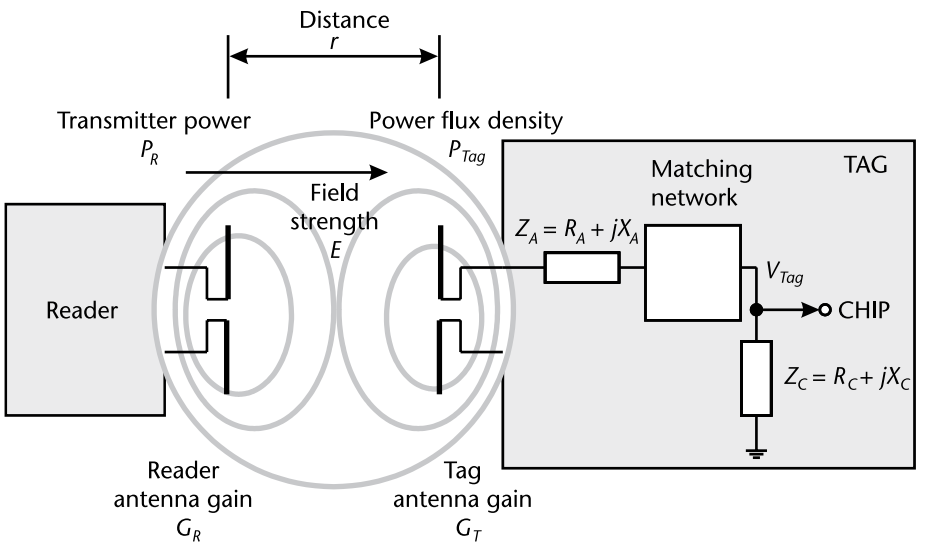


Figure 5.1 Forward power transfer.

The separation between the antennas is r, which is assumed to be large enough for the tag to be in the far field of the reader. The efficiency of the matching network will be taken as unity and ignored (losses in the network may also be accounted for in the value of G_T).

Antenna gains G_R and G_T are expressed relative to an isotropic antenna. Polarization mismatch between antennas is assumed negligible.

Let us assume for a moment that electromagnetic waves propagate under ideal conditions (i.e., without dispersion). If high-frequency energy is emitted by an isotropic radiator, then the energy propagates uniformly in all directions. Areas with the same power flux density therefore form spheres ($A = 4\pi r^2$) around the radiator.

Here $\frac{P_R}{4\pi r^2}$ is a *nondirectional power flux density, S,* from an isotropic antenna. An isotropic radiator is a theoretical, lossless, omnidirectional (spherical) antenna. That is, it radiates uniformly in all directions. The power of a transmitter that is radiated from an isotropic antenna will have a uniform power density (power per unit area) in all directions. In other words, the spreading out of electromagnetic energy in free space is determined by the inverse square law.

The *directional power flux density* at a distant point from a real reader antenna with a gain of G_R is the power flux density from an isotropic antenna multiplied by the reader antenna gain:

$$S_D = \left(\frac{P_R}{4\pi r^2}\right) \cdot G_R \quad \left[W/m^2\right] \tag{5.4}$$

Power flux density can also be expressed using the *electric field strength, E,* of the reader at the tag position, and the *wave impedance in free space*, $Z_0 = 120\pi \approx 377\Omega$[1]:

$$S = \frac{E^2}{Z_0} = \frac{E^2}{120\pi} \tag{5.5}$$

The amount of power that is actually received by a tag placed at distance r from the isotropic antenna is denoted P_{Tag}. The received power will depend on the receiving antenna's aperture, which describes how well an antenna can pick up power from an incoming electromagnetic wave. For an isotropic antenna with an aperture $A_e = \frac{\lambda^2}{4\pi}$, and from the considerations of *power flux density* at the tag, with λ as the wavelength, we get the following:

$$P_{Tag} = S \cdot \left(\frac{\lambda^2}{4\pi}\right) \tag{5.6}$$

1. The symbol η (eta) may be used instead of Z for wave impedance to avoid confusion with electrical impedance.

After combining (5.4) and (5.6), and using real tag antenna with the gain G_T, we obtain the electromagnetic power, P_{Tag}, received by the tag:

$$
\begin{aligned}
P_{Tag} &= S \cdot \left(\frac{\lambda^2}{4\pi}\right) \cdot G_T \\
P_{Tag} &= \left(\frac{P_R G_R}{4\pi r^2}\right)\left(\frac{\lambda^2}{4\pi}\right) \cdot G_T = \frac{P_R G_R G_T \lambda^2}{(4\pi)^2 r^2} \\
P_{Tag} &= \frac{P_R G_R G_T}{\left(\frac{4\pi r}{\lambda}\right)^2} = \left(\frac{\lambda}{4\pi r}\right)^2 P_R G_R G_T
\end{aligned}
\tag{5.7}
$$

Here, the attenuation factor $\left(\frac{4\pi r}{\lambda}\right)^2$ is called the *free-space path loss*. Free-space path loss is proportional to the square of the distance between the transmitter and receiver, and also proportional to the square of the frequency of the radio signal.

The typical maximum reader output power is 500 mW, 2W (ERP, CEPT), and 4W (EIRP, FCC). Converted to dBm, the permitted maximum limits are about 29 dBm (500-mW ERP, 825-mW EIRP), 35 dBm (2-W ERP, 3.3-W EIRP), and 36 dBm (4-W EIRP).

The gain of the transmitter (reader) antenna (typical value) is assumed to be 6 dBi. Therefore, the maximum output power from power amplifier should be 23 dBm, 29 dBm, and 30 dBm, respectively. The tag available power versus distance can be seen in Figure 5.2.

From the industrial experience, the minimum RF input power of 10 μW (–20 dBm) to 50 μW (–13 dBm) is required to power on the tag. The power received by the tag is then divided in two parts: the reflected power and the available power used by the chip. The distribution of these two parts is very critical for obtaining a maximum distance. For dipole antennas presented in the best orientation, G_T may be taken as 2 dBi (gain over isotropic with allowance for losses, approximately 1.6).

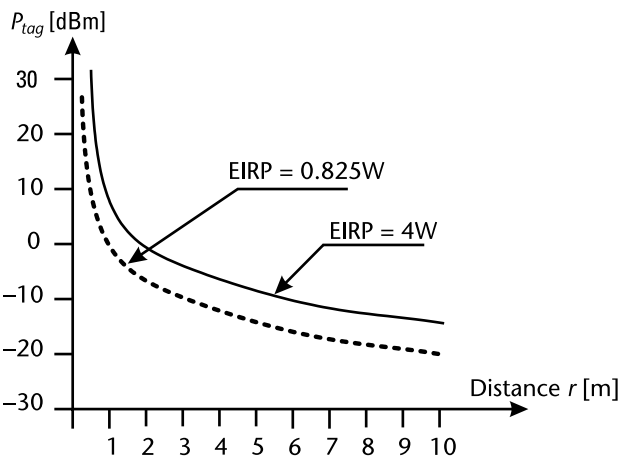


Figure 5.2 Tag received power versus distance.

From $P = \frac{V^2}{R} \Rightarrow V = \sqrt{PR}$ and (5.7), we can say that:

$$V_{Tag} = \sqrt{P_{Tag} R_C} = \sqrt{\left(\frac{\lambda}{4\pi r}\right)^2 P_R G_R G_T R_C}$$
$$V_{Tag} = \frac{\lambda}{4\pi r} \sqrt{P_R G_R G_T R_C} \tag{5.8}$$

Note that $P_R G_R$ is the EIRP of the reader. The maximum practical value of the input resistance, R_C, is 600Ω. The received voltage V_{Tag} must be large enough to be rectified and power the tag; a voltage in excess of 1.2 V_{rms} may be required. This is with the tag presented to the interrogating field in the ideal orientation and with no power margin.

It would be interesting to find out what is a required transmit power of the reader in order to have a large enough tag voltage for reliable operation. If we assume that the tag's RMS voltage is 1.6V, we can calculate P_{Tag}:

$$P_{Tag} = \frac{V_{Tag}^2}{R_C} = \frac{1.6^2}{600} \approx 0.0043\,W \approx 4.3\,mW \tag{5.9}$$

Solving (5.8) for P_R, we get (5.10). At the frequency of 915 MHz (λ = 0.33m), for example, it can be seen that with 1.6 V_{rms} tag voltage (assuming both the gain of the reader and the tag to be 2 dBi or 1.6), the required reader power at 1-m distance is 2.42W.

$$P_R = \left(\frac{4\pi r V_{Tag}}{\lambda}\right)^2 \left(\frac{1}{G_R G_T R_C}\right)$$
$$P_R = \left(\frac{4\pi \cdot 1 \cdot 1.6}{0.33}\right)^2 \left(\frac{1}{1.6 \cdot 1.6 \cdot 600}\right) \approx 2.42\,W \tag{5.10}$$

Since $P_R G_R$ is the EIRP of the reader, we get the required EIRP of the reader:

$$P_{REIRP} = P_R \cdot G_R = 2.42 \cdot 1.64 \approx 3.97\,W \tag{5.11}$$

The relationship between the electrical field strength and the power flux density is the same as between voltage and power in an electrical circuit; from (5.4) and (5.5), we can say that the electric field strength of the reader, in volts per meter, at the tag location is equal to 10.9 V/m:

$$\frac{P_R G_R}{4\pi r^2} = \frac{E^2}{120\pi}$$

Solving for E, we get

$$E = \frac{\sqrt{30 P_R G_R}}{r}$$
$$E = \frac{\sqrt{30 \cdot 3.97}}{1} \approx 10.9 V/m \tag{5.12}$$

Note that the gain of 2 dBi is approximately equivalent to the gain of 1.6 and can be calculated as follows:

$$G_{dBi} = 10 \log G$$
$$G = 10^{\frac{G_{dBi}}{10}} = 10^{\frac{2}{10}} = 10^{0.2} \approx 1.6 \tag{5.13}$$

5.2.1.2 The Radar Equation

Radar principles tell us that the amount of energy reflected by an object is dependent on the reflective area of the object: the larger the area, the greater the reflection. This property is referred to as the RCS. The RCS is an equivalent area from which energy is collected by the target and retransmitted (backscattered) back to the source.

For an RFID system in which the tag changes its reflectivity in order to convey its stored identity and data to the reader, this is referred to as *differential radar cross-section* or ΔRCS. Calculations of the complete return signal path are conveniently conducted in terms of the ΔRCS of the backscatter device.

For the antenna to transfer maximum energy to the chip, the impedance of the chip must be a conjugate of the *antenna impedance,* Z_A. However it is important to remember that the logic circuits of a chip used in a tag draw very little power relative to the amount of power consumed by the chip RF input circuits. As the modulator switches between two states, the load impedance of the chip, Z_C, will switch between two states.

The *reflection due to a mismatch* between antenna and load in a backscatter tag is analogous to the reflection found in transmission lines and may be expressed in terms of *coefficient of reflection*, ρ:

$$\rho = \frac{Z_C - Z_A^*}{Z_C + Z_A} \tag{5.14}$$

In a transmission line theory, we define a voltage reflection coefficient (at the load) as the ratio of reflected voltage to incident voltage, which can, in general, be a complex number. As already mentioned in more detail in Chapter 3, three special cases are possible:

- Matched load, $Z_C = Z_A$, no reflection, $\rho = 0$;
- Open load, $Z_C = \infty$, full in-phase reflection, $\rho = +1$;
- Shorted load, $Z_C = 0$, full out-of-phase reflection, $\rho = -1$.

The coefficient of reflection, ρ, will therefore change as the modulator switches between two states. When the tag modulator is in the *off state*, the chip input impedance will be closely matched to the antenna impedance; therefore, the reflectivity will be low and hence the SWR will approach 1. When the modulator is in the *on state*, the tag antenna impedance will be mismatched and so the reflectivity will be high, and the SWR will tend to infinity, causing the maximum amount of power to be reflected.

The tag varies its RCS by changing the impedance match of the tag antenna between two (or more) states. The ratio between the states is called the *differential coefficient of reflectivity* represented by the symbol $\Delta\rho$ and may be calculated using well-known transmission line theory, a summary of which is provided next.

Signal propagation follows the well-known *Friis transmission formula*[2]; analytical approaches like the Friis equation assume undisturbed near-field conditions (i.e., no proximity of dielectric or metal objects), known antenna characteristics, and no diffraction and reflection effects.

An antenna of gain G_T has an effective aperture (5.15) as shown here:

$$A_e = \frac{A^2}{4\pi} \cdot G_T \quad \left[m^2\right] \tag{5.15}$$

The $\Delta\rho$ is the *differential reflection coefficient* of the tag modulating circuitry and can be calculated as shown:

$$\Delta\rho = p_1\left(1 - |\rho_1|^2\right) + p_2\left(1 - |\rho_2|^2\right) \tag{5.16}$$

where the IC is in states 1 and 2 for a fraction of time, p_1 and p_2, respectively [3].

It is worth mentioning that in case the tag modulator switches from a perfectly matched state ($\rho_2 = 0$) to a short circuit state or to an open circuit state ($\rho_1 = 1$), $\Delta\rho$ will be approximately 0.5. Lower (and more realistic) modulation ratios will result in a $\Delta\rho < 0.5$. However, in those cases where the modulator switches from a state where the chip has an impedance higher than the antenna impedance, to a condition where it is lower than the antenna impedance, $\Delta\rho$ will represent the difference between the two states, in which case $\Delta\rho$ could be greater than 0.5 (but never higher than 1).

In modulation schemes, where one of the two states is active most of the time (e.g., $p_1 << p_2$), this is a good choice in terms of power efficiency, but these schemes require a much larger bandwidth (due to the short gaps) which is often prohibited by national authorities' regulations. Assuming that both states are active an equal amount of time (as it is in ASK with total mismatch in one state), that is, $p_1 = p_2 = 0.5$, and assuming there are no antenna losses, 50% of the available input power is

2. The formula was derived in 1945 by the Danish-American radio engineer Harald T. Friis at Bell Labs.

actually available for rectification, 25% is used as backscattered modulated power, and the remaining 25% is wasted.

The σ is the ΔRCS of the tag and P_{Ret} is the power returned to the reader. The radar cross-section (RCS),[3] ΔRCS, of the tag antenna is equivalent to the antenna effective aperture A_e, when the tag is matched. For the ΔRCS, when the tag antenna is mismatched, theoretically speaking, we can say that:

$$\sigma = \Delta_{RCS} = A_e G_T (\Delta\rho)^2 = \frac{\lambda^2}{4\pi} G_T \cdot G_T (\Delta\rho)^2$$
$$\sigma = \frac{\lambda^2 G_T^2 (\Delta\rho)^2}{4\pi} \left[m^2\right] \tag{5.17}$$

where G_T is the gain of the tag antenna (G_T is squared because the signal is received and reradiated), λ is the wavelength, and $\Delta\rho$ is the differential reflection coefficient of the tag modulator.

First, we assume that electromagnetic waves propagate under ideal conditions, that is, without dispersion. If high-frequency energy is emitted by an isotropic radiator, than the energy propagate uniformly in all directions. Areas with the same power density therefore form spheres ($A = 4\pi r^2$) around the radiator (see Figure 5.3).

As discussed in a previous section, the same amount of energy spreads out on an incremented spherical surface at an incremented spherical radius. That means that the power density on the surface of a sphere is inversely proportional to the radius of the sphere. So we get the formula to calculate the nondirectional power flux density, at the distance r:

$$S = \frac{P_R}{4\pi r^2} \left[W/m^2\right] \tag{5.18}$$

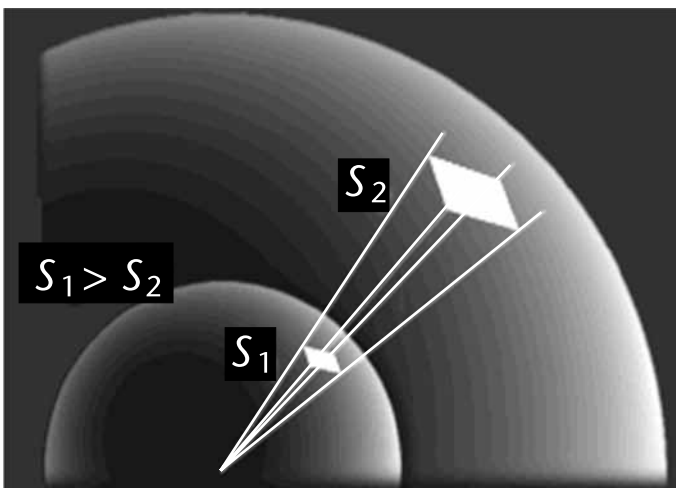


Figure 5.3 Illustration of the power flux density.

3. RCS is the efficiency with which the transponder reflects incoming electromagnetic waves.

where P_R is the power transmitted from the reader.

Because a spherical segment emits radiation equally in all direction (at constant transmit power), if the power radiated is redistributed to provide more radiation in one direction, resulting in an increase of the power density in direction of the radiation. This effect is called *antenna gain* and it is obtained by directional radiation of the power.

So, from the definition, the *directional power flux density,* S_D, is equal to:

$$S_D = S \cdot G_R \tag{5.19}$$

The target (tag in our case) detection is not only dependent on the power density at the tag's position, but also on how much power is reflected in the direction of the radar (reader in our case). In order to determine the useful reflected power, it is necessary to know the RCS, σ. This quantity depends on several factors, but it is a general rule that a bigger area reflects more power than a smaller area. For example, a jumbo jet offers a higher RCS than a sporting aircraft in the same flight situation. Beyond this, the reflecting area depends on design, surface composition, and materials used.

Keeping all this in mind, we can say that the returned (reflected) power P_{Ret} towards the RFID reader depends on the distance, power density S_D, the reader's antenna gain, G_R, and the RCS, σ:

$$P_{\mathrm{Re}t} = \frac{P_R}{4\pi r^2} \cdot G_R \cdot \sigma \ \ [W] \tag{5.20}$$

Because the reflected signal encounters the same conditions as the transmitted signal, the power density yielded at the receiver of the reader is given by:

$$S_{REC} = \frac{P_{\mathrm{Re}t}}{4\pi r^2} = \frac{P_R G_R}{(4\pi)^2 r^4} \cdot \sigma \tag{5.21}$$

The backscatter communication radio link budget, a modification of the monostatic radar equation, describes the amount of modulated power that is scattered from the RF tag to the reader:

$$P_{REC} = S_{REC} \cdot A_e = \frac{P_R G_R \sigma}{(4\pi)^2 r^4} \cdot \frac{\lambda^2 G_R}{4\pi} = \frac{P_R G_R^2 \lambda^2 \sigma}{(4\pi)^3 r^4} \tag{5.22}$$

It could be noticed that the received power at the reader depend heavily on the distance and decays with the fourth power of the distance. For successful operation, it is required both that signal at reader's receiver is above the noise floor and that ratio of the power received and transmitted from the reader is not too small.

Table 5.1 shows the return signal ratio for various situations. Ratios below 100 dB are both manageable in terms of signal processing and ensure the return signal is significantly above the thermal noise floor.

Table 5.1 Received UHF RFID Power for Various Distances

Frequency (MHz)	*Wavelength (m)*	*Distance (m)*	*Reader Power (W)*	*Reader Power (dBm)*	*Reader Antenna Gain*	*Tag Antenna Gain*	$\Delta\rho$	σ (m^2)	*Received Power (W)*	*Received Power (dBm)*	*Power Ratio (dB)*
915.00	0.33	1.00	2.00	33.01	1.60	1.60	0.50	0.0055	1.5185	–28.19	–61.20
915.00	0.33	2.00	2.00	33.01	1.60	1.60	0.50	0.0055	0.0949	–40.23	–73.24
915.00	0.33	4.00	2.00	33.01	1.60	1.60	0.50	0.0055	0.0059	–52.27	–85.28
915.00	0.33	8.00	2.00	33.01	1.60	1.60	0.50	0.0055	0.0004	–64.31	–97.32
433.00	0.69	1.00	2.00	33.01	1.60	1.60	0.50	0.0244	30.2793	–15.19	–48.20
433.00	0.69	2.00	2.00	33.01	1.60	1.60	0.50	0.0244	1.8925	–27.23	–60.24
433.00	0.69	4.00	2.00	33.01	1.60	1.60	0.50	0.0244	0.1183	–39.27	–72.28
433.00	0.69	8.00	2.00	33.01	1.60	1.60	0.50	0.0244	0.0074	–51.31	–84.32

5.2.1.3 Noise

Noise is the major limiting factor in communications system performance. Noise can be divided into four categories: thermal noise, intermodulation noise, crosstalk, and impulse noise. For this analysis, we will only consider thermal noise and neglect other potential sources of noise. Now we have to calculate the power of the reflected signal at the receiver of the reader and compare it with the thermal noise threshold.

Thermal noise results from thermal agitation of electrons; it is present in all electronic devices and transmission media and is function of temperature and the channel bandwidth. Thermal noise is independent of any specific frequency. Thus, the thermal noise power in watts present in a bandwidth of B hertz can be expressed as:

$$N = kTB \tag{5.23}$$

where Boltzmann's constant $k = 1.3803 \times 10^{-23}$ J/K and T is the temperature in kelvin[4], and $T = 273.16 + t$ [°C].

In dBW, (5.23) would look like this:

$$N = -228.6 + 10\log T + 10\log B \tag{5.24}$$

It is also possible to define *rms noise voltage* across some resistance R by applying Ohm's law to (5.23):

$$E_N = \sqrt{4RkTB} \tag{5.25}$$

The *noise figure* or *noise factor* (NF) is a contribution of the device itself to thermal noise. It is commonly defined as the signal-to-noise ratio at the input divided by the signal-to-noise ratio at the output and is usually expressed in decibels. Typical noise figures range from 0.5 dB for very low noise devices to 4 to 8 dB.

4. The *kelvin* is a unit of measurement for temperature and is one of the seven base units in the International System of Units (SI). The Kelvin scale is named after the Belfast-born engineer and physicist William Thomson, first Baron Kelvin (1824–1907), who wrote of the need for an *absolute thermometric scale*.

In the case of 500-kHz bandwidth, at the room temperature, we can use (5.22) and calculate thermal noise of –117 dBm; with addition of the 3-dB receiver noise factor, total noise is –114 dBm, leaving enough reader signal margin to the noise threshold.

We can be seen that the return power ratio conditions are met at relatively longer ranges and that forward power transfer is the limiting factor in UHF backscatter tags. Battery-powered backscatter tags overcome this limitation and can be read at significantly greater ranges.

5.2.2 Near-Field Propagation Systems

5.2.2.1 Magnetic Field Calculations

At low to mid RFID frequencies, RFID systems make use of near-field communication and the physical property of inductive coupling from a magnetic field. The reader creates a magnetic field between the reader and the tag and this induces an electric current in the tag's antenna, which is used to power the integrated circuit and obtain the ID. It is important to remember here that the term *antenna* actually means the magnetically coupled coil, and not an antenna in a radio propagation sense.

The ID is communicated back to the reader by varying the load on the antenna's coil, which changes the current drawn on the reader's communication coil. In the near field, it is possible to have an electric field with very little magnetic field or a magnetic field with very little electric field. The choice between these two alternatives is determined by the design of the interrogation antenna, and RFID systems are generally designed to minimize any incidental electric field generation.

In the near field, the magnetic field strength attenuates according to the relationship $1/r^3$ (i.e., the magnetic field intensity decays rapidly as the inverse cube of the distance between the reader antenna and the tag). In power terms, this equates to a drastic $1/r^6$ reduction with distance (60 dB/decade) of the available power to energize the tag.

The magnetic field strength is high in the immediate vicinity of the transmitting coil, but a very low level exists in the distant far-field; hence, a spatially well-confined interrogation region or localized tag reading zone is created. Note that magnetic loop reader antennas can also be designed that exhibit good electrical symmetry and balance to eliminate stray electric E-field pick-up.

The tag's ability to efficiently draw energy from the reader's field is based on the well-known electrical resonance effect. The coupling or antenna element of the tag is really an inductor coil and capacitor connected together and designed to resonate at the 13.56-MHz system operating frequency (Figure 5.4).

The current passing through the inductor creates a surrounding magnetic field according to Ampere's law. The created magnetic field B (expressed in tesla[5]) is not a propagating wave [4], but rather an attenuating carrier wave (Figure 5.5), with its strength calculated using the following formula:

5. The tesla (symbol T) is the SI-derived unit of magnetic flux density (or magnetic induction). It is used to define the intensity (density) of a magnetic field. It is named in honor of world-renowned inventor, scientist, and electrical engineer Nikola Tesla. The tesla, equal to 1 weber per square meter, was defined in 1960.

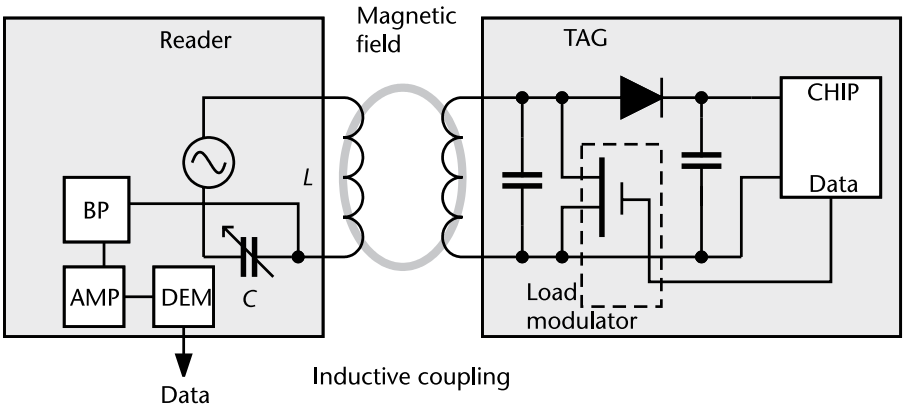


Figure 5.4 Principle of inductive (near-field) coupling.

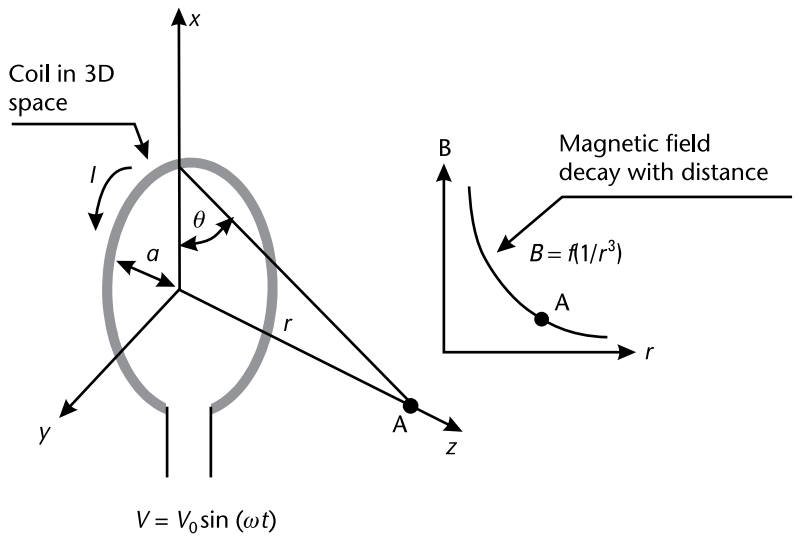


Figure 5.5 Calculation of the magnetic field away from the coil.

$$B = \frac{\mu_0 INa^2}{2r^3} \quad \left[Weber/m^2 \text{ or } tesla\right] \tag{5.26}$$

where:

I = current through the coil;
N = number of windings in the coil;
a = radius of the coil;
μ_0 = permeability of free space ($4\pi \times 10^{-7}$ H/m);
r = perpendicular distance from antenna to point A and $r >> a$.

As one moves away from the source with $r >> a$, the simplified equation (5.26) shows the characteristic $1/r^3$ attenuation. This near-field decaying behavior of the magnetic field is the main limiting factor in the read range of the RFID device.

We use Ohm's law for ac circuits and obtain:

$$I = \frac{V}{Z_L} = \frac{V}{\omega L} \quad \text{where} \quad \omega = 2\pi f \tag{5.27}$$

and assume that L can be approximated as follows:

$$L \approx \mu_0 \pi a N^2 \tag{5.28}$$

We can then rewrite (5.26) as:

$$B = \frac{Va}{2\omega N \pi r^3} \tag{5.29}$$

From (5.29) with a given coil voltage at some distance from the coil, we can now see that B is inversely proportional to N. This is due to the fact that the current increases at the rate of $1/N^2$ with a given coil voltage V. Only the case of an air-coiled inductor has been described, but the ferrite-cored inductor could be used as well. Adding a core has the effect of increasing the effective surface area, enabling one to reduce the physical size of the coil.

To maximize the magnetic field, given fixed antenna dimensions, (5.26) dictates that the current delivered to the antenna must be maximized. Additionally, to maximize current, the antenna must resonate at the excitation frequency provided by the reader circuit. Resonance frequency, f_0, of the reader is determined by the inductance, L, of the antenna (determined by the radius of the coil, the number of windings, the thickness of the windings, and the length of the coil) and a tuning capacitor, C, and is calculated as follows:

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{5.30}$$

The same formula is used to calculate tag's resonant frequency, which is determined by choosing the inductive and total capacitive values, so that the value for the tag's resonant frequency, f_0, is achieved. In the case of a tag's resonant frequency:

L [H] = inductance of tag's antenna coil;

C [F] = capacitance of tag's tuning capacitor.

$$Z(j\omega) = R + j(X_L - X_C) \tag{5.31}$$

In practice, when the tuned circuit is resonating, the sum of its capacitive and inductive reactance is zero ($X_L = X_C$), and the impedance shown in (5.31) becomes purely resistive.

Total resistance is thereby minimized and current through the antenna is maximized, yielding a maximized magnetic field strength. Passive tags utilize the energy provided by the carrier wave through an induced antenna coil voltage. The voltage is proportional to the product of the number of turns in the tag antenna and the total magnetic flux through the antenna. The ASIC within the tag must receive a certain minimum threshold voltage (or power) to operate.

5.2.2.2 Voltages Induced in Antenna Circuits

Faraday's law states that a time-varying magnetic field through a surface bounded by a closed path induces a voltage around the loop. Figure 5.6 shows a simple geometry for an RFID application.

When the tag and reader antennas are in close proximity, the time-varying magnetic field, *B*, produced by a reader antenna coil induces a voltage (called electromotive force or simply EMF) in the closed tag antenna coil. The induced voltage in the coil causes a flow of current through the coil. The induced voltage in the coil is equal to the time rate of change of the magnetic flux Ψ:

$$V = -N\frac{d\psi}{dt} \tag{5.32}$$

where N is the number of turns in the antenna coil and Ψ is the magnetic flux through each turn. The negative sign indicates that the induced voltage acts in such a way as to oppose the magnetic flux producing it. This is known as *Lenz's law*, and it emphasizes the fact that the direction of current flow in the circuit is such that the induced magnetic field produced by the induced current will oppose the original magnetic field.

The magnetic flux Ψ in (5.33) is the total magnetic field B that is passing through the entire surface S of the antenna coil (magnetic field B and surface S are vector quantities), and it is found by:

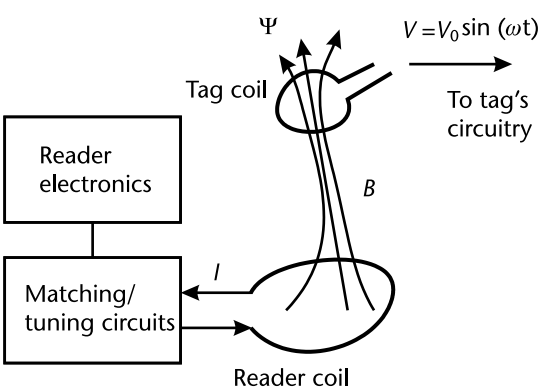


Figure 5.6 Basic reader and tag configuration.

$$\psi = \int B \bullet dS \tag{5.33}$$

where:

B = magnetic field given in (5.22);

S = surface area of the coil;

• = inner product (cosine angle between two vectors) of vectors B and surface area S.

The above formula for the inner product of two vectors suggests that the total magnetic flux ψ that is passing through the antenna coil is affected by an orientation of the antenna coils.

The inner product of two vectors becomes minimized when the cosine angle between the two are 90°, or the two (B field and the surface of coil) are perpendicular to each other and maximized when the cosine angle is 0°. The maximum magnetic flux that is passing through the tag coil is obtained when the reader coil and tag coil are placed in parallel with respect to each other. This condition results in maximum induced voltage in the tag coil and also maximum read range.

The inner product expression also can be expressed in terms of a mutual coupling between the reader and tag coils. The mutual coupling between the two coils is also maximized when they are parallel.

Combining expressions given so far, the voltage across the tag antenna, at the resonant frequency, can be calculated using the following equation:

$$V_{Tag} = 2\pi f N Q B(S\cos\theta) \tag{5.34}$$

where:

f = frequency of the carrier signal;

S = area of the coil in square meters;

Q = quality factor of the resonant circuit;

B = strength of magnetic field at the tag;

θ = angle of the field normal to the tag area.

The $S\cos\theta$ term in (5.34) represents an *effective surface area* of the antenna coil[6] and is defined as an exposed area of the loop to the incoming magnetic field. The effective antenna surface area is maximized when $\cos\theta$ becomes unity ($\theta = 0°$), which occurs when the antennas of the base station and the transponder units are positioned in a face-to-face arrangement. In practical applications, the user might notice the longest detection range when the two antennas are facing each other and the shortest range when they are orthogonally faced.

6. It is important to notice that the antenna coil is not a real antenna in a radio communications sense; we have to keep in mind that this is just a transformer coil.

Voltage is built up in an onboard storage capacitor, and when sufficient charge has accumulated to reach or surpass the circuit operating threshold voltage, the electronics power up and begin transmitting data back to the reader. Both the reader and the tag must use the same transmission method in order to synchronize and successfully exchange data.

Two main methods of communication occur between the reader and tag: full-duplex and half-duplex. In a *full-duplex* configuration, the tag communicates its data by modulating the reader's carrier wave by applying a resistive load. A transistor (load modulator) within the tag shorts the antenna circuit in sequence to the data, removing the antenna from resonance at the excitation frequency, thereby removing its power draw from the reader's carrier wave. At the reader side, the loading and unloading is detected and the data can be reconstructed. In a *half-duplex* RFID system, the carrier wave transmits power and then pauses. Within the pause, the tag transmits the data back to the reader.

For a given tag, the operating voltage obtained at a distance r from the reader is directly proportional to the flux density at that distance. The magnetic field emitted by the reader antenna decreases in power proportional to $1/r^3$ in the near field. Therefore, it can be shown (5.35) that for a circularly coiled antenna, the flux density is maximized at a distance r (in meters) when:

$$a = \sqrt{2} \cdot r \tag{5.35}$$

where a is the radius of the reader's antenna coil. Thus, by increasing a, the communication range of the reader may be increased, and the optimum reader antenna radius a equals 1.41 times the demanded read range r. It is obvious that for some applications very large reader antennas may be required.

The *quality factor*, Q, of the coupling element defines how well the resonating circuit absorbs power over its relatively narrow resonance band. In general, the higher the Q, the higher the power output for a particular sized antenna. Unfortunately, too high a Q may conflict with the bandpass characteristics of the reader and the increased ringing could create problems in the protocol bit timing.

In smart-label RFID applications, the Q value demanded is reasonably high. Because most of the resonant circuit's tuning capacitance is located within the microchip where high capacitor Q can be realized, the effective circuit Q value is determined mainly by the antenna coil losses. The coil Q is usually calculated (without taking into account additional parasitic capacitance losses) according to the following equation:

$$Q = \frac{\omega L}{R_s} = \frac{1}{\omega CR} = \frac{\omega}{\omega_2 - \omega_1} \tag{5.36}$$

In (5.36), R_s is the coil's total effective series loss resistance and takes into account both the dc resistance and the ac resistance due to high frequency current flow concentration caused by skin-effect phenomena in the conductor windings.

Practical smart-label systems usually operate with a coupling element resonator Q within the range of 20 to 80. The Q of the LC circuit is typically around 20 for an air-core inductor and about 40 for a ferrite-core inductor. Higher Q values

than this are generally not feasible because the information-bearing amplitude modulated reply sidebands are undesirably attenuated by the resonator's bandpass frequency response characteristic. At resonance, the induced RF voltage produced across the tuned tag and delivered to the microchip will be Q times greater than for frequencies outside of the resonant bandwidth.

Figure 5.7 shows the frequency response curve for a typical serial resonant tank circuit. A good rule of thumb is to stay within the –3-dB limits; for the individual manufacturing tolerances for capacitance and inductance of 2%, a Q of 30 can be used. Lower tolerance components may be used at the expense of sensitivity and, thus, yield a lower range. The corresponding final design must accommodate a wider bandwidth and will, therefore, have a lower response.

As a resonant application, the smart-label tag can be affected by the *environmental detuning effects* that may cause a reduction in transponder sensitivity and reading distance. Undesirable changes in the tag's parasitic capacitance and effective inductance can happen easily. The presence of metal and different dielectric mediums can cause detuning and introduce damping resulting from dissipative energy losses. Such permeable materials can also distort the magnetic flux lines to weaken the energy coupling to the tag. However, these effects can largely be overcome when they are taken into account during the label and system design phase.

Clusters of tagged objects that sometimes come together in close physical proximity to each other can also exhibit significant detuning effects caused by their mutual inductances. This shift in tuning is called *resonance splitting* and occurs whenever resonant circuits, such as near-field tags, come in close physical proximity to each other. They become coupled tuned circuits, and the degree of coupling (called the *coupling coefficient, k*) determines the amount of frequency shift. The value of k depends on the coil geometry (size and shape) and spacing distance.

Bigger area coils are inherently more susceptible to deleterious mutual coupling effects. When in close proximity, the magnetic flux lines of the individual coils overlap and the coils exhibit mutual inductance. This mutual inductance generally adds to the coil's normal inductance and produces a downward shift in the effective resonant frequency. This, in turn, results in the tag receiving less energy from the reader field and hence the reading distance decreases accordingly. The higher

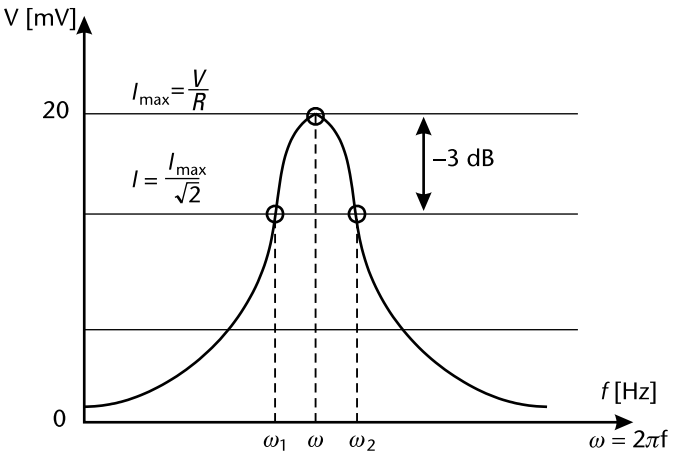


Figure 5.7 Frequency response curve for resonant tank circuit.

the tag Q, the more pronounced is the effect. Closely coupled tags can also have problems with commands signaled from the reader being misinterpreted due to cross-coupling between tags.

This rapid attenuation of the energizing and data communication field with increasing distance is the fundamental reason why 13.56-MHz passive RFID systems have a maximum reading distance of the order of about 1m (3 feet). This is also the reason why well-designed near-field RFID systems have good immunity to environmental noise and electrical interference. All these characteristics are particularly well suited to many smart-label applications.

The efficiency of power transfer between the antenna coil of the reader and the transponder is proportional to the operating frequency, the number of windings, the area enclosed by the transponder coil, the angle of the two coils relative to each other, and the distance between the two coils. As frequency increases, the required coil inductance of the transponder coil, and thus the number of windings, decreases. For 135 kHz it is typically 100 to 1,000 windings, and for 13.56 MHz, typically 3 to 10 windings. Because the voltage induced in the transponder is still proportional to frequency, the reduced number of windings barely affects the efficiency of power transfer at higher frequencies.

5.3 Tags

5.3.1 Tag Considerations

There really is no such thing as a typical RFID tag. The read range is a balancing act between numbers of engineering trade-offs and ultimately depends on many factors. It depends on the frequency of RFID system operation, the power of the reader, and interference from other RF devices.

Several general RFID tag design requirements whose relative importance depends on tag application are discussed here [5]. These requirements largely determine the criteria for selecting an RFID tag antenna:

- *Frequency band:* Desired frequency band of operation depends on the regulations of the country where tag will be used.
- *Size and form:* Tag form and size must be such that it can be embedded or attached to the required objects (cardboard boxes, airline baggage strips, identification cards, and so on) or fit inside a printed label (see Figure 5.8).
- *Read rang:* Minimum required read range is usually specified.
- *EIRP:* EIRP is determined by local country regulations (active versus passive tags).
- *Surrounding:* Tag performance changes when it is placed on different objects (e.g., cardboard boxes with various contents) or when other objects are present in the vicinity of the tagged object. A tag's antenna can be designed or tuned for optimum performance on a particular object or designed to be less sensitive to the content on which the tag is placed.
- *Orientation* (also called *polarization*)*:* The read range depends on antenna orientation. How tags are placed with respect to the polarization of the reader's

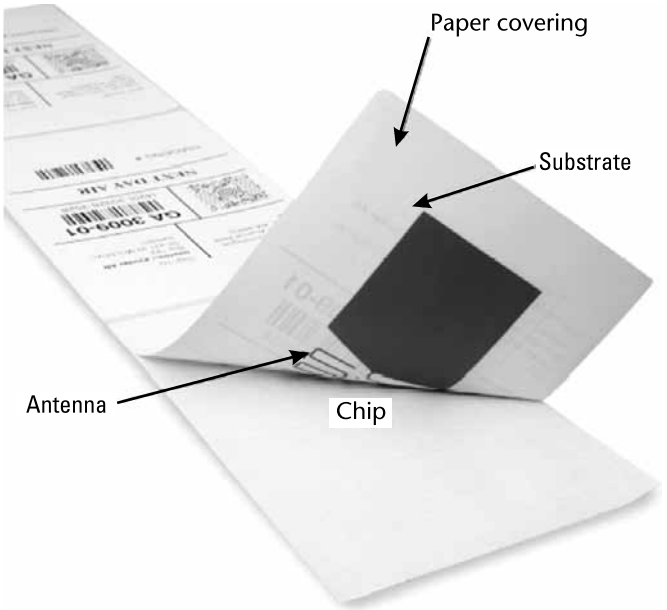


Figure 5.8 RFID label cross-section.

field can have a significant effect on the communication distance for both HF and UHF tags, resulting in a reduced operating range of up to 50%. In the case of the tag being displaced by 90°, the reader may not being able to read the tag at all.

The optimal orientation for HF tags is for the two antenna coils (reader and tag) to be parallel to each other (see Figure 5.9). UHF tags are even more sensitive to polarization due to the directional nature of the dipole fields. Some applications require a tag to have a specific directivity pattern such as omnidirectional or hemispherical coverage.

- *Applications with mobility:* The RFID tag may be used in situations where tagged objects like pallets or boxes travel on a conveyor belt at speeds up to 600 feet/min. The Doppler shift in this case is less than 30 Hz at 915 MHz and does not affect RFID operation. However, the tag spends less time in the

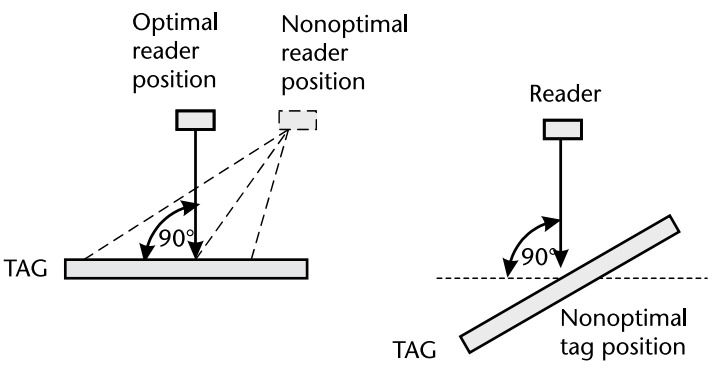


Figure 5.9 Optimal and nonoptimal tag and reader position.

read field of RFID reader, demanding high read-rate capability. In such cases, RFID system must be carefully planned to ensure reliable tag identification.

- *Cost:* RFID tag must be a low-cost device, thus imposing restrictions both on antenna structure and on the choice of materials for its construction including the ASIC used. Typical conductors used in tags are copper, aluminum, and silver ink. The dielectrics include flexible polyester and rigid PCB substrates like FR4.
- *Reliability:* RFID tag must be a reliable device that can sustain variations due to temperature, humidity, and stress and survive such processes as label insertion, printing, and lamination.
- *Power for the tag:* An active tag has its own battery and does not rely on the reader for any function. Its range is greater than passive tags. Passive tags rely on the reader for power to perform all functions, and semipassive tags rely on the reader for powering transmission but the battery for powering their own circuitry. For comparison, see Table 5.2.

5.3.2 Data Content of RFID Tags

5.3.2.1 Read-Only Systems

Read-only systems can be considered low-end; these tags usually only contain an individual serial number that is transmitted when queried by a reader and basically can be used to replace the functionality of bar codes. Due to the structural simplicity of read-only tags, costs and energy consumption can be kept down.

More advanced tags contain logics and memory, so they support writing and information can be updated or changed remotely. High-end tags have microprocessors enabling complex algorithms for encryption and security. More energy is needed for these than for less complex electronics.

One-bit tags can be detected, but they do not contain any other information. They are very useful for protecting items in a shop against shoplifters. A system like this is called Electronic Article Surveillance (EAS) and has been in use since the 1970s. In practice, this system can be identified by the large gates of coils or antennas at the exits of shops.

Read-only tags that contain more than 1 bit of data are simple ones that only contain a unique serial number that it transmits on request. The contents of the read-only chips are usually written during manufacturing. The serial number can, for example, be coded by cutting small bridges on the chip. Usually these simple chips also contain some logic for anticollision, thus allowing multiple tags to be read simultaneously.

Table 5.2 Power for the Tag

Tag Type	*Power Source*	*Memory*	*Communication Range*
Active	Battery	Most	Greatest
Semi-Passive	Battery and Reader	Moderate	Moderate
Passive	Reader	Least	Least

Different principles of operation can be used for the 1-bit tags. For example, the principle of the microwaves tags is quite simple; it uses the generation of harmonics by diodes, that is, frequencies that are an integer multiples of the original frequency.

The tag is a small antenna that has a diode in the middle. Because the diode only lets current pass one way, the oscillations that get trapped behind the diode generate a frequency twice the original frequency. The system sends out a microwave signal, for example, 2.45 GHz, and listens for the first harmonics at 4.90 GHz. If a tag is present, it generates harmonics that can be detected.

However, false alarms may be caused by other sources of this particular frequency. In order to avoid such false alarms, a modulation signal of, for example, 100 kHz is added to the interrogation signal. This means that the same modulating signal can also be found in the reflected signal from the tags.

5.3.2.2 Read-Write Systems

Many read-only tags are factory programmed and carry an ID number (tag ID). Other tags, including read/write devices, can also carry a tag identifier that is used to unambiguously identify a tag. This identifier is distinct from user-introduced identifiers for supporting other application needs.

- *Read-only devices:* They are generally less costly and may be factory programmable read-only or one time programmable (OTP). One-time programmability provides the opportunity to write once then read many times, thus supporting passport-type applications, in which data can be added at key points during the lifetime or usage of an item, and thus providing an incorruptible history or audit trail for the item data.
- *Write once/read many (WORM):* Some chips allow writing only once and read many times. These tags are versatile since they can be written with a serial number when applied to an item, instead of linking a predefined serial number to an item. More advanced chips allow both reading and writing multiple times, and the contents of a tag can be altered remotely by a scanner.
- *Read/write data carriers:* They offer the facility for changing the content of the carrier as and when appropriate within a given application. Some devices will have both a read-only and read-write component that can support both identification and other data carrier needs. The read-write capability can clearly support applications in which an item, such as a container or assembly support, is reusable and requires some means of carrying data about its contents or on what is being physically carried. It is also significant for lifetime applications such as maintenance histories, where a need is seen to add or modify data concerning an item over a period of time. The read/write capability may also be exploited within flexible manufacturing to carry and adjust manufacturing information and item-attendant details, such as component tolerances. A further important use of read/write is for local caching of data as a portable data file, using it as and when required, and selectively modifying it as appropriate to meet process needs.

  Additionally, the chip must be able to resolve who can access it and prevent

wrong people from altering its contents. For secure data transmission, some kind of encryption is added as well.

- *Time to read* is the time it takes to read a tag, which, of course, is related to data transfer rate. For example, a system operating at 1-kbps transfer rate will take approximately 0.1 second to read a 96-bit tag. Various factors can influence read time including competing readers and tags (reader access and multiple tags).

5.3.3 Passive Tags

5.3.3.1 About Passive Tags

Passive RFID devices have no power supply built in, meaning that electrical current transmitted by the RFID reader inductively powers the device, which allows it to transmit its information back.

Because the passive tag has a limited supply of power, its transmission is much more limited than an active tag, typically no more than simply an ID number. Similarly, passive devices have a limited range of broadcast, requiring the reader be significantly closer than an active one would. Applications for passive devices tend to include inventory, product shipping and tracking, hospitals and other medical purposes, and antitheft, where it is practical to have a reader within the few meters or so of the RFID device. Passive devices are ideal in places that prevent the replacement of a battery, such as implanted into the human body or placed under the skin.

Tags consist of a silicon device (chip) and antenna circuit (Figure 5.10). The purpose of the antenna circuit is to induce an energizing signal and to send a modulated RF signal and the read range of a tag largely depends upon the antenna circuit and size. The antenna circuit is made of a LC resonant circuit or E-field dipole antenna, depending on the carrier frequency. The LC resonant circuit is used for frequencies of less than 100 MHz. At these frequencies, the communication between the reader and tag is achieved with magnetic coupling between the two antennas. An antenna utilizing inductive coupling is often called a *magnetic dipole antenna*. The antenna circuits must be designed in such a way to maximize the magnetic coupling between them. This can be achieved with the following parameters:

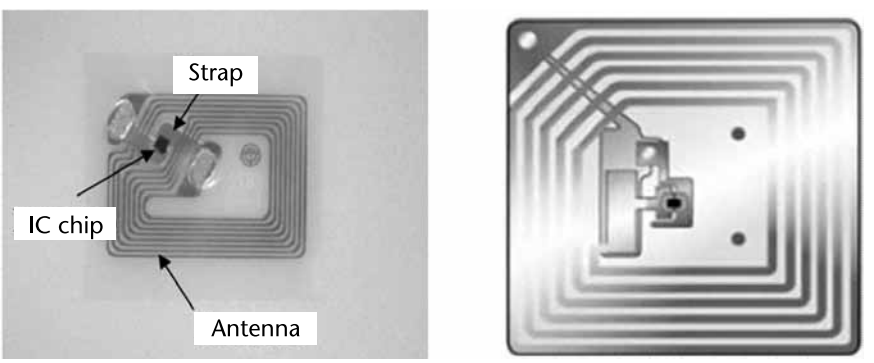


Figure 5.10 13.56-MHz RFID tags.

- The LC circuit must be tuned to the carrier frequency of the reader.
- The Q of the tuned circuit must be maximized.
- The antenna size must be maximized within physical limit of application requirement.

The passive RFID tags sometimes use backscattering of the carrier frequency for sending data from the tag to the reader. The amplitude of backscattering signal is modulated with the data of the tag device. The modulation data can be encoded in the form of ASK (NRZ or Manchester), FSK, or PSK.

During backscatter modulation, the incoming RF carrier signal to the tag is loaded and unloaded, causing amplitude modulation of the carrier, corresponding to the tag data bits. The RF voltage induced in the tag's antenna is amplitude modulated by the modulation signal (data) of the tag device. This amplitude modulation can be achieved by using a modulation transistor across the LC resonant circuit or partially across the resonant circuit. Changes in the voltage amplitude of the tag's antenna can affect the voltage of the reader antenna. By monitoring the changes in the reader antenna voltage (due to the tag's modulation data), the data in the tag can be reconstructed.

The RF voltage link between the reader and tag antennas are often compared to *weakly coupled transformer coils*; as the secondary winding, tag coil, is momentarily shunted, the primary winding, reader coil, experiences a momentary voltage change. Opening and shunting the secondary, tag coil, in sequence with the tag data are seen as amplitude modulation at the primary, reader coil.

5.3.3.2 RFID Chip Description

An RFID tag consists of an RFID chip, an antenna, and tag packaging. The RFID circuitry itself consists of an RF front end, some additional basic signal processing circuits, logic circuitry to implement the algorithms required, and EEPROM for storage. The RFID chip is an integrated circuit implemented in silicon [6].

The major blocks and their functions of the RFID front end are:

- *Rectifier:* From the coupled EM field it generates the power supply voltage for all the electronic circuits;
- *Power (voltage) regulator:* Maintains the power supply at a certain level and at the same time protects the circuit from the excessively large input RF power;
- *Demodulator:* Extracts the data symbols embedded in the carrier waveforms;
- *Clock extraction or generation:* Extracts the clock out of carrier (usually in HF systems) or generates the system clock;
- *Backscattering:* Modulates the return link by alternating the impedance of the chip;
- *Power on reset:* Generates the chip power-on reset (POR) signal;
- *Voltage (current) reference:* Generates some voltage or current reference for the use of front-end and other circuit blocks, usually in terms of bandgap reference;

- *Other circuits:* They include persistent node or short-term memory, electrostatic discharge protection, and so forth.

Figure 5.11 describes a block diagram for RFID IC circuits and lists many of its associated function blocks. The RF front end is connected to the antenna, and typically, at UHF, an electric dipole antenna is used while HF tags use a coil antenna. The front-end circuitry impacts the semiconductor process by requiring a process that allows for mixed mode fabrication. Passive RF tags have no power source and rely on the signal from the reader to power up; thus, the RF front end implements modulators, voltage regulators, resets, and connections to the external antenna.

RFID chips have control logic that typically consists of a few thousand gates. The lowest level chip uses very few gates, in the order of 1,500 gate equivalents. Functions in the logic include the error and parity/CRC checkers, data encoders, anticollision algorithms, controllers, and command decoders. More complex RFID chips may include security primitives and even tamper-proofing hardware. The size of the circuit affects the number of mask, metal, and poly layers required in the semiconductor process, and RFID systems usually use CMOS.

A certain amount of information is stored on-chip in an EEPROM. The size of this EEPROM increases as more information is required to be on the RFID chip.

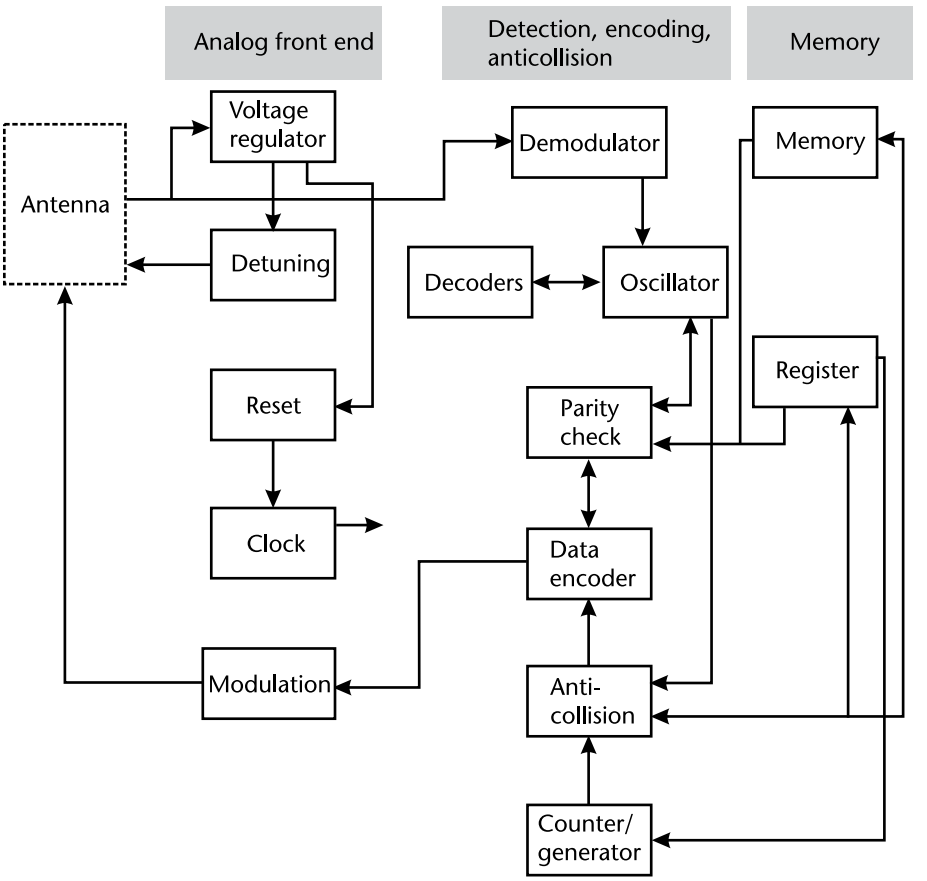


Figure 5.11 RFID tag circuit block diagram.

The size of the required EEPROM is a factor in determining the number of mask, metal, and poly layers required in the semiconductor fabrication process. It is also a factor in the size of the final semiconductor die. Silicon cost is directly proportional to both the die size and the number of mask, poly, and metal layers.

The IC in an RFID tag must be attached to an antenna to operate. The antenna captures and transmits signals to and from the reader. The coupling from the reader to the tag provides both the transmission data and the power to operate the passive RFID tag. Typically antennas for passive RFID systems can be either simple dipole for 915-MHz RFID tags or more complex coiled shapes for 13.56-MHz systems.

The digital anticollision system is one of the major and most important parts of the tag chip, since it not only implements the slotted ALOHA random anticollision algorithm, but also executes read/write operation of memory. As we know, the power consumption of memory is very difficult to reduce. Even more, besides the power consumption, the efficiency of the RF front-end rectifier prefers lower output dc voltage. So it is very important to design a low-power, low-voltage digital anticollision system to achieve maximum operating range.

Currently, antennas are made of metals or metal pastes and typically cost as much as 12 cents per antenna to manufacture. However, new methods that range from conductive inks and new antenna deposition and stamping techniques are expected to reduce costs below 1 cent.

5.3.4 Active Tags

5.3.4.1 Active Tag Description

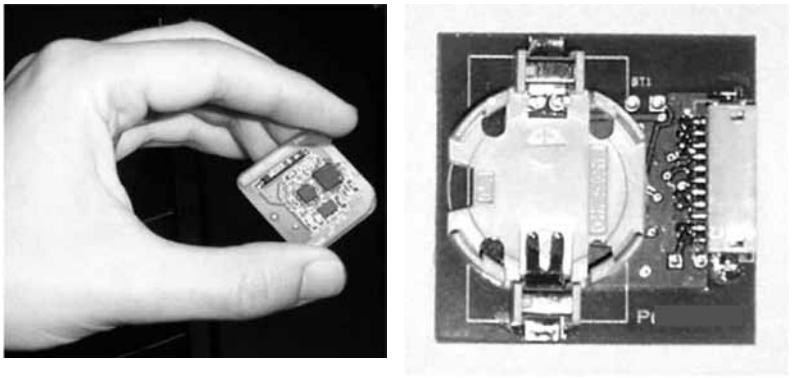
An active tag usually performs a specialized task and has an on-board power source, usually a battery. It does not require inductions to provide current, as is true of the passive tags. The active tag can be designed with a variety of specialized electronics including microprocessors, different types of sensors, or I/O devices. Depending on the target function of the tag, this information can be processed and stored for immediate or later retrieval by a reader.

Active RFID tags, also called *transponders* because they contain a transmitter that is always on, are powered by a battery about the size of a coin, and are designed for communications up to 30m (100 feet) from the RFID reader (Figure 5.12). They are larger and more expensive than passive RFID tags, but can hold more data about the product and are commonly used for high-value asset tracking. A feature that most active tags have and most passive tags do not is the ability to store data received from a transceiver.

Active tags are ideal in environments with electromagnetic interference since they can broadcast a stronger signal in situations that require a greater distance between the tag and the transmitter.

The additional space taken up by a battery in an active device necessitates that the active devices are substantially larger than the passive devices. To date, commercially available passive tags are as small as 0.4 mm square and thinner than a sheet of paper. In contrast, commercially available active tags are still only as small as a coin, which means that active tags are around 50 times the size of passive ones.

For the read-only device, the information that is in the memory cannot be changed by an RF command once it has been written. A device with memory cells



(a) (b)

Figure 5.12 Active tag: (a) front side and (b) reverse side.

that can be reprogrammed by RF commands is called a *read/write device*. The information in the memory can be reprogrammed by interrogator command.

Although passive tags can only respond to an electromagnetic wave signal emitted from a reader, active tags can also spontaneously transmit an ID. There are various types of unscheduled transmission type, such as when there are changes in vibration or temperature.

A *semiactive* or *semipassive tag* (naming convention depends on the manufacturer) also has an on-board battery. The battery in this case is only used to operate the chip. Like the passive tag, it uses the energy in the electromagnetic field to wake up the chip and to transmit the data to the reader. These tags are sometimes called battery-assisted passive (BAP).

5.3.4.2 Active Tag Classification

Two types of tag systems can generally be recognized within active RFID systems.

Wake-up tag systems are deactivated, or asleep, until activated by a coded message from a reader or interrogator. In the sleep mode, limiting the current drain to a low-level alert function conserves the battery energy. Where larger memories are accommodated, there is also generally a need to access data on an object or internal file basis to avoid having to transfer the entire amount of data so held. These are used in toll payment collection, checkpoint control, and in tracking cargo.

Awake tag or *beacon systems* are, as the term suggests, responsive to interrogation without a coded message being required to switch the tag from an energy conservation mode. Because greater switching rate is generally associated with higher energy usage, these tags generally operate at a lower data transfer rates and memory sizes than wake-up tags, so they conserve battery energy in this way. This type of tag is the most widely used of the two, and because of lower component costs it is generally less expensive than wake-up tag systems.

Beacons are used in most real-time locating systems (RTLS), where the precise location of an asset needs to be tracked. In an RTLS, a beacon emits a signal with its unique identifier at preset intervals, every 3 seconds or once a day, depending on how important it is to know the location of an asset at a particular moment in time.

5.3.5 Active And Passive Tags Comparison

Active RFID and passive RFID technologies, while often considered and evaluated together, are fundamentally distinct technologies with substantially different capabilities. In most cases, neither technology provides a complete solution for supply chain asset management applications; rather, the most effective and complete supply chain solutions leverage the advantages of each technology and combine their use in complementary ways.

Passive RFID is most appropriate where the movement of tagged assets is highly consistent and controlled, and little or no security or sensing capability or data storage is required.

Active RFID is best suited where business processes are dynamic or unconstrained, movement of tagged assets is variable, and more sophisticated security, sensing, and/or data storage capabilities are required.

Passive and active tagging systems present very different deployment issues. Active tags contain significantly more sophistication, data management, and security concerns. While passive tags are typically well below $1, active tags generally cost from $10 to $50, depending on the amount of memory, the battery life required, any on-board sensors, and the ruggedness.

5.3.6 Multiple Tag Operation

If many tags are present, then they will all reply at the same time, which at the reader end is seen as a signal collision and an indication of multiple tags. The reader manages this problem by using an anticollision algorithm designed to allow tags to be sorted and individually selected. There are many different types of algorithms (binary tree, ALOHA, and so on), which are defined as part of the protocol standards.

The number of tags that can be identified depends on the frequency and protocol used and can typically range from 50 tags per second for HF and up to 200 tags per second for UHF. Once a tag is selected, the reader is able to perform a number of operations, such as reading the tag's identifier number or, in the case of a read/write tag, writing information to it. After finishing its dialogue with the tag, the reader can then either remove it from the list or put it on standby until a later time. This process continues under control of the anticollision algorithm until all tags have been selected.

When containers or freight are moved on a conveyor or similar equipment in a tag reader system, the reader/writer must read and write data to and from moving tags (Figure 5.13). For successful access, the following conditions must be satisfied:

$$T_c = A_{cn} \frac{D_t}{D_r} \tag{5.37}$$

This formula shows that when the data transfer volume of the tag (D_t) increases, and the data transfer rate (D_r) decreases, the tag-reader/writer operation time (T_c) increases, and operation may fail.

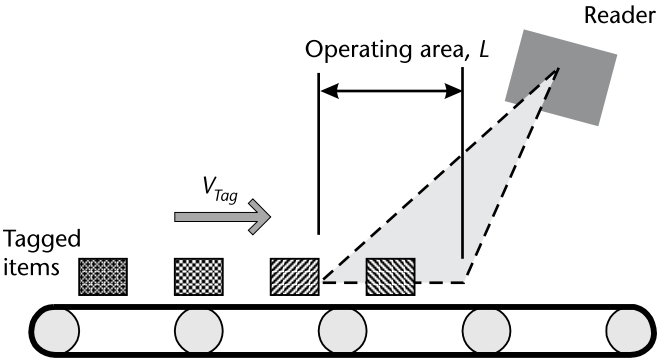


Figure 5.13 Reading moving tags.

$$T_r = \frac{L}{V_{tag}} \tag{5.38}$$

Equation (5.38) shows that when the reader/writer operating area decreases, the distance the tag moves (L) decreases, and the tag movement velocity (V_{tag}) increases, the amount of time the tag is in the operating area (T_r) decreases, and the operation may fail.

$$T_r \geq T_c + T_d \tag{5.39}$$

Finally, (5.39) states that the total amount of time spent in the operating area must be more than the total time taken by the reader/writer and the detection of all tags. If only one type of tag can be used when reading/writing RFID tags attached to freight on a conveyor belt, the reader/writer antenna must have a large operating area to cope with the conveyor belt's speed:

- T_r [seconds] = amount of time tag is in operating area;
- T_c [seconds] = tag-reader/writer operation time;
- D_r [bps] = data transfer rate;
- D_t [bit] = data transfer volume;
- A_{cn} [count] = average number of tag-reader/writer operations;
- V_{tag} [m/second] = tag movement velocity;
- L [m] = distance tag moves within operating area;
- T_d [sec] = amount of time for detecting existence of all tags.

Virtually all high-volume RFID applications require the ability to read multiple tags in the reading field at one time. This is only possible if each RFID tag has a unique ID number. One numbering method is the EPC code that contains both an item ID number and a serial number. A unique number is the basis for implementing anticollision in any RFID technology.

In a multiple-tag operation, in which multiple RFID tags are in the reader/writer's operating area, the reader/writer must detect the presence of these multiple tags, and read/write each of them consecutively (Figure 5.14). This operating method is generally referred to as the *anticollision protocol* and is different from the single-tag operation protocol.

The effects of operating range, tag orientation, tag movement velocity, and the presence of metallic substances on multitag operating characteristics are basically the same as those on single-tag operating characteristics. One additional problem with multitag operating characteristics is that the operating time is several times longer than for single-tag operation. Because the reader/writer must read/write each tag, the time increases in proportion to the number of tags. Also, since multiple tags are used, tags sometimes come into contact or overlap with each other.

When there are N tags in the operating area, and the N_{tag} is the number of tags, the amount of time for which the tags must be in the operating area (T_r) is described by (5.40):

$$T_r \geq \left(T_c + T_{tag}\right) N_{tag} \tag{5.40}$$

Although the reader/writer may sometimes read/write stationary tags, in most cases, the tag will be moving. The reader/writer will generally have to read/write

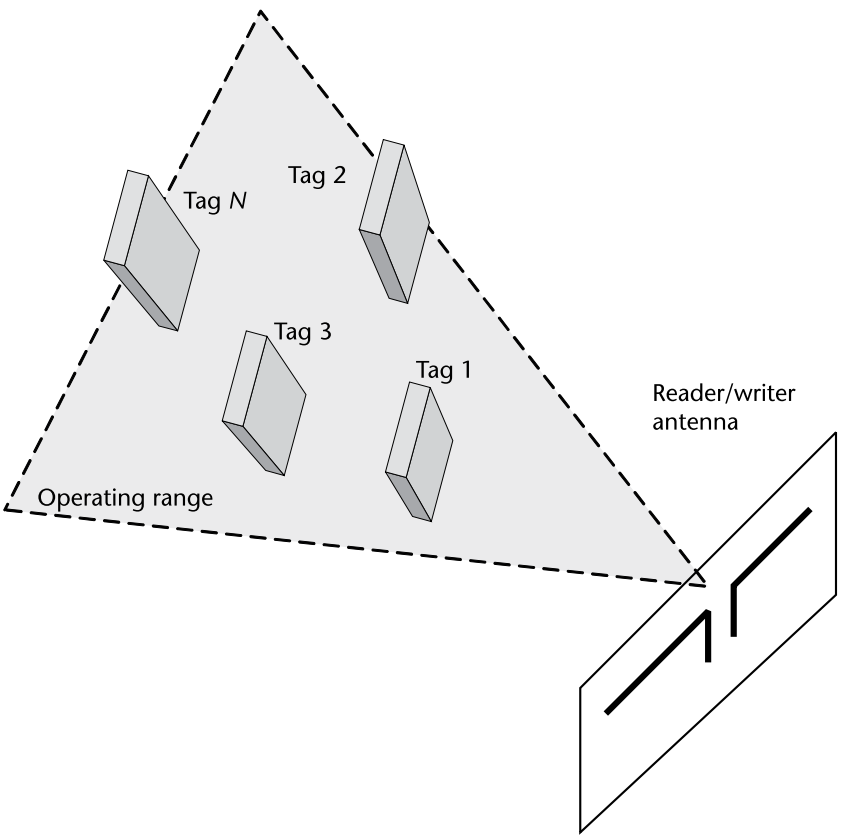


Figure 5.14 Multiple-tag operation.

RFID tags attached to containers or freight being transported on a conveyor belt or trolleys. When T_c is the operating time for a single tag, and T_d is the time required to check for the existence of N multitags in an anticollision protocol, (5.41) gives an approximation of the maximum time required for the reader/writer to read/write all N tags (T_N):

$$T_N = \left(T_c + T_d\right) N_{tag} \tag{5.41}$$

Example:

Tag information volume = 16 bytes;

Data transfer rate (D_r) = 7.8 kbps;

T_c = 0.057 second;

T_d = 0.055 second;

N = 10;

$T_N = (0.057 + 0.055) \times 10 = 1.12$ seconds.

Therefore, roughly 1.1 seconds are needed for the reader/writer to finish reading/writing all 10 tags. When the tag information volume is 100 bytes, T_N becomes roughly 7 seconds. For the reader/writer to read/write all the tags, the time required for the tags to pass through the operating area (T_r) must be greater than T_N.

One unfortunate but real fact about RFID tags is that the quality of tags is currently not consistent, and therefore performance is not consistent. There are considerable variations in performance from one tag to the next, even among tags of the same manufacturer and same model.

5.3.7 Overlapping Tags

In inductive frequency band RFID, the resonance characteristic of the tag antenna coil is used for reader/writer operation. As discussed earlier, tag's resonant frequency, f_0, is calculated by:

$$f_0 = \frac{1}{2\pi\sqrt{LC}} \tag{5.42}$$

where

L [H] = inductance of tag antenna coil;

C [F] = capacitance of tag's tuning capacitor.

If tags overlap, the inductance of their antenna coils is obstructed, and L increases. In this case, the resonant frequency expressed by formula becomes lower ($f_1 < f_0$). As a result, the electromagnetic waves (current i) generated by the tag's coil become smaller, and the operating area decreases (see Figure 5.15).

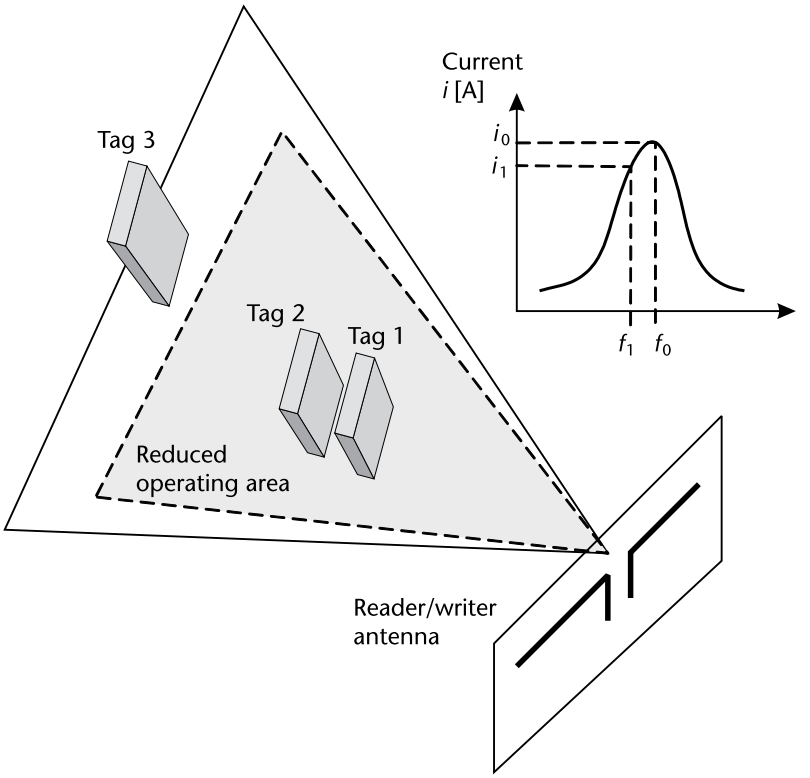


Figure 5.15 Overlapping tags.

5.3.8 Tag Antennas

5.3.8.1 Antenna Selection

An antenna is a conductive structure specifically designed to couple or radiate electromagnetic energy. Antenna structures, often encountered in RFID systems, may be used to both transmit and receive electromagnetic energy, particularly data-modulated electromagnetic energy.

In the low-frequency (LF) range with short read distances, the tag is in the near field of the reader antenna, and the power and signals are transferred by means of a magnetic coupling. In the LF range, the tag antenna therefore comprises a coil (inductive loops) to which the chip is attached. In the UHF range, in cases where the read distances are larger, the tag is located in the far field of the reader antenna. The reader and tag are coupled by the electromagnetic wave in free space, to which the reader and tag are tuned by means of appropriate antenna structures.

Good antenna design is a critical factor in obtaining good range and stable throughput in a wireless application. This is especially true in low-power and compact designs in which antenna space is less than optimal. The tag antenna should be as small as possible and easy to produce. It is important to remember that, in general, the smaller the antenna, the lower the radiation resistance and the lower the efficiency.

Printed antennas are very easy to produce. The antenna is attached as a flat structure to a substrate. The next stage in the production process often involves attaching the chip to the substrate and connecting it to the antenna. This assembly is

called an *inlay*. An inlay becomes a tag or transponder when it is fixed to an adhesive label or a smart card. Note, however, that the electromagnetic properties of the materials surrounding the inlay affect the tag's ability to communicate. In extreme cases, tags cannot be read if unsuitable reader antennas are selected.

Another type of usage involves integration into the object that is to be identified. Parts of the object can be shaped to form an antenna and the antenna can be adjusted optimally to suit the object. This significantly increases readability, while simultaneously protecting against counterfeiting.

The size and shape of the tag antenna have a significant effect on tag read rates, regardless of the coupling used for communication. There are various types of antennas available, among which the most commonly used are dipole, folded dipole, printed dipole, printed patch, squiggle, and log-spiral. Among these, the dipole, folded dipole, and squiggle antennas are omnidirectional, thus allowing them to be read in all possible tag orientations, relative to the base antenna. On the other hand, directional antennas have good read range due to their radiation patterns.

Care must be taken while choosing an antenna because the antenna impedance must match to the ASIC and to free space. The four major considerations when selecting an antenna type are as follows:

1. Antenna impedance;
2. Antenna radiation pattern;
3. Nature of the tagged object;
4. Vicinity of structures around the tagged object.

When individual system performance is not satisfactory, it is advisable to bring redundancy to the system. Low read rates of RFID systems make the deployment of redundant antennas and tags to identify the same object an imperative.

Redundant tags are those tags that carry identical information performing identical functions. *Dual tags* are tags connected to each other having one or two antennas and with/without individual or shared memory; *n* tags serving the same purpose as that of dual tags can be used for beneficial use of multiple tags in product identification.

It is observed that both the inductive coupling and backscatter-based tags have a dependency on the angle of orientation of tag relative to the reader. The placement of two tags in two flat planes, three tags in the three-dimensional axes, four tags along the faces of a regular tetrahedron, and so on, can help in achieving the above mentioned goals.

The choice of etched, printed, or stamped antenna is a trade-off between cost and performance. For a 13.56-MHz tag, the Q factor of the antenna is very important for long read range applications. The Q factor is inversely proportional to the resistance of the antenna trace. It has been determined that the etched antenna is less resistive and inexpensive than the printed antenna with conductive material. However, for a very large antenna size (greater than 4×4 inches), both etching and stamping processes waste too much unwanted material. Therefore, printed or wired antennas should be considered as an alternative.

As previously stated, reducing antenna size results in reduced performance. Some of the parameters affected are reduced efficiency (or gain), shorter range, smaller useful bandwidth, more critical tuning, increased sensitivity to component

and PCB spread, and increased sensitivity to external factors. Several performance factors deteriorate with miniaturization, but some antenna types tolerate miniaturization better than others. How much a given antenna can be reduced in size depends on the actual requirements for range, bandwidth, and repeatability. In general, an antenna can be reduced to half its natural size with moderate impact on performance.

5.3.8.2 Loop Antennas

Passive RFID tags extract all of their power, to both operate and communicate, from the reader's magnetic field. Coupling between the tag and reader is via the mutual inductance of the two loop antennas, and the efficient transfer of energy from the reader to the tag directly affects operational reliability and read/write range.

Loop antennas have the same desirable characteristics as dipoles and monopoles; they are inexpensive and simple to construct. Loop antennas are usually classified as either electrically small or electrically large based on the circumference of the loop. Electrically small loop antennas have circumference $C \leq \lambda/10$, and electrically large loop antennas have circumference $C \approx \lambda$.

Generally, both 13.56-MHz and 125-kHz RFID tags use parallel resonant LC loop antennas tuned to the carrier frequency. The RFID circuit is similar to a transformer in which loop inductors magnetically couple when one of the loops, in the case of the reader antenna, is energized with an alternating current, thus creating an alternating magnetic field. The tag loop antenna acts like the secondary of a transformer, where an alternating current is induced in the antenna, extracting energy from the magnetic field.

Generally, the larger the diameter (and circumference) of the tag's antenna loop, the more magnetic flux lines will be passing through the coil, therefore increasing the transfer of energy from the reader to the tag.

Loop antennas can be divided in three groups:

1. Half-wave antennas;
2. Full-wave antennas;
3. Series-loaded, short loop antennas.

where wave refers to the approximate circumference of the loop.

The *half-wave loop* consists of a loop approximately one-half wavelength in circumference with a gap cut in the ring. It is very similar to a half-wave dipole that has been folded into a ring, and most of the information about the dipole applies to the half-wave loop. Because the ends are very close together, some capacitive loading exists, and resonance is obtained at a somewhat smaller circumference than expected. The feedpoint impedance is also somewhat lower than the usual dipole, but all the usual feeding techniques can be applied to the half-wave loop.

By increasing the capacitive loading across the gap, the loop can be made much smaller than one-half wavelength. At heavy loading, the loop closely resembles a single winding, LC-tuned circuit. The actual shape of the loop is not critical, and typically, the efficiency is determined by the area enclosed by the loop. The half-wave loop is popular at lower frequencies, but at higher frequencies, the tuning capacitance across the gap becomes very small and critical.

As the name implies, the *full-wave loop* is approximately one wavelength in circumference. Resonance is obtained when the loop is slightly longer than one wavelength. The full-wave loop can be thought of as two end-connected dipoles. Like the half-wave loop, the shape of the full wave loop is not critical, but efficiency is determined mainly by the enclosed area. The feed impedance is somewhat higher (approximately 120Ω) than the half-wave loop.

Loading of the full-wave loop is accomplished by inserting small coils or hairpins in the loop, which reduces the size. Like the dipole and half-wave loop, numerous impedance-matching methods exist, including gamma matching and tapering across a loading coil or hairpin. The main advantage of the full-wave loop is it does not have the air gap in the loop, which is very sensitive to load and PCB capacitance spread.

Loaded-loop antennas are commonly used in remote control and remote keyless entry (RKE) applications. The loop is placed in series with an inductor, which reduces efficiency of the antenna but shortens the physical length.

5.3.8.3 UHF Antennas

A typical inductively coupled feeding structure is shown in Figure 5.16(a) where antenna consists of a feeding loop and a radiating body. Two terminals of the loop are connected to the chip, and the feed is combined with the antenna body with mutual coupling. For example, if the measured impedance of the selected IC is 73-j113, the load antenna impedance should be 73+j113 for conjugate matching. To achieve this, the proposed antenna structure is shown in Figure 5.16(b), with the dipole arms bent into arc shape [7].

Another way to achieve high resistance with inductively coupled feeding structure is to introduce extra radiating elements. A dual-body configuration is presented in Figure 5.16(c). Two meandering line arms are placed in each side of the feeding loop. The slight decrease of mutual coupling is due to the shorter coupling length. However, strong mutual coupling is now introduced between the two radiating bodies, which can be similarly regarded as in a parallel connection seen from the feeding loop. In this way, resistance of the radiating body is significantly reduced, resulting in high resistance with meandering line arms.

This antenna can be easily tuned by trimming. Lengths of meander trace and loading bar can be varied to obtain optimum reactance and resistance matching. The trimming is realized by punching holes through the antenna trace at defined locations. Such a tunable design is desirable when a solution is needed for a particular application with a minimum lead time.

5.3.8.4 Fractal Antennas

Short reading distances and the fact that the cost per tag is still too high are the major reasons that passive RFID systems have not made their breakthrough yet. One key to increase reading distances is the tag antenna improvement. Because a passive tag does not have its own power supply, it is important that the tag antenna is able to absorb as much of the energy, radiated from the reader, as possible.

Another important thing is to minimize the size of the tags. Small tags and hence small-tag antennas will expand the applicable areas of RFIDs. The challenge

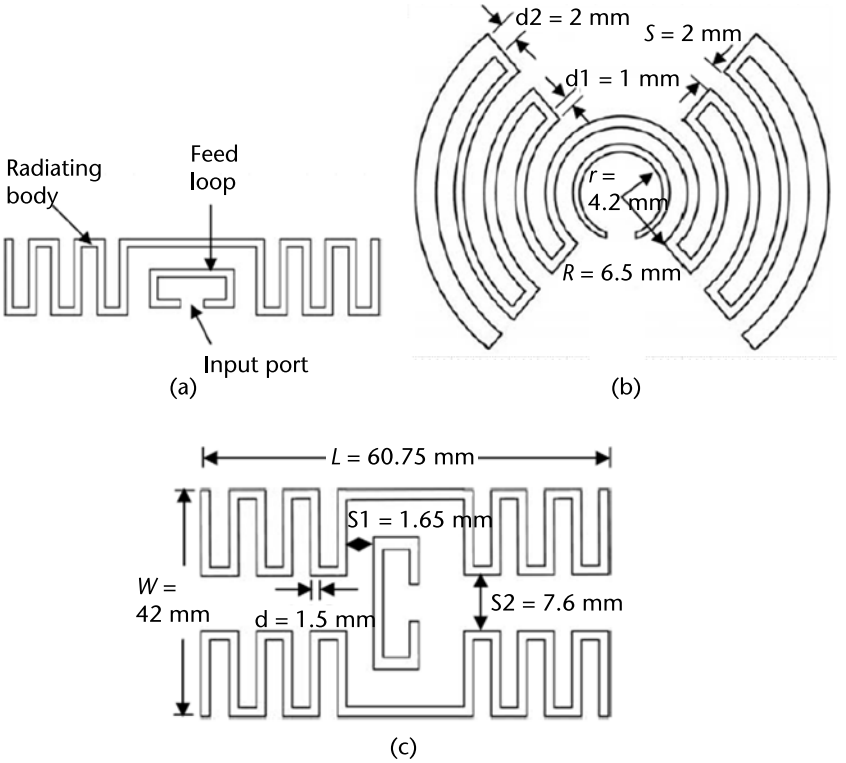


Figure 5.16 (a–c) UHF antennas.

is that small, effective antennas are in general poor radiators. A factor that affects the size of the tags is the frequency that is used. Different frequency bands are allocated for RFID, and these bands differ in different regions of the world. From an economic point of view it is highly desirable to be able to use only one type of tag in all of the different regions.

In the study of antennas, fractal antenna theory is a relatively new area. However, fractal antennas and its superset fractal electrodynamics are a hotbed of research activity these days. Fractals are geometrical shapes, which are self-similar, repeating themselves at different scales [8]. Many mathematical structures that are fractals, for example, Sierpinski's gasket, Cantor's comb, von Koch's snowflake, the Mandelbrot set[7], and the Lorenz attractor. Fractals also describe many real-world objects that do not correspond to simple geometric shapes, for example, clouds, mountains, turbulence, and coastlines [9].

Fractal antennas do not have any characteristic size; fractal structures with a self-similar geometric shape consisting of multiple copies of themselves on many different scales have the potential to be frequency-independent or at least multi-frequency antennas. For example, it has been shown that a bow-tie antenna can operate efficiently over different frequencies and that the bands can be chosen by modifying the structure. Examples of wideband antennas are the classical spiral

7. The terms *fractal* and *fractal dimension* are due to Mandelbrot, who is the person most often associated with the mathematics of fractals [9]. The term fractal means linguistically "broken" or "fractured" from the Latin *fractus*.

antennas and the classical log-periodic antennas, which can also be classified as fractal antennas.

Fractal antennas are convoluted, uneven shapes, and sharp edges, corners, and discontinuities tend to enhance radiation of electromagnetic energy from electric systems. Fractal antennas, therefore, have the potential to be efficient. This is particularly important when small antennas are to be designed, because small antennas are not generally good at radiating electromagnetic energy.

Some fractals have the property that they can be very long, but still fit in to a certain volume or area. Because fractals do not have a dimension that is an integer, they can more effectively fill a given volume or area at deposal. Small antennas generally have a very small input resistance and a very significant negative input reactance, meaning that they are poor radiators. It has been shown that many small fractal antennas have greater input resistance and smaller input reactance than small traditional antennas.

Also, the Q factor of small antennas depends on how effectively the antenna occupies a certain radian sphere. Small fractal antennas can thus be expected to have lower Q factors than their regular counterparts and, hence, higher bandwidth.

Small input resistance and large input reactance also means that it is difficult and expensive to match the antenna input impedance with a matching network. It has been shown that many fractal antennas can even resonate with a size much smaller than the regular ones. Hence, it is possible to reduce or even eliminate the cost associated with input impedance matching.

By shaping antennas in certain ways, the directivity can be improved. Fractal antennas are shaped antennas. In some RFID applications in which the tag orientation can be controlled, it is possible that antennas with high directivity are preferred to achieve long reading distances and/or to avoid problems associated with scanning several tags simultaneously. It is also possible that in some applications a small handheld reader with a high directivity antenna is desirable.

Because frequency-independent antennas and wideband antennas tend to be insensitive to deformations like cutting in the structure and bending of the structure, one might expect some fractal antennas to be resistant against deformations too.

5.3.9 UHF Tags Circuits

5.3.9.1 Tag dc Supply Voltage Circuitry

The voltage multiplier converts a part of the incoming RF signal power to dc for power supply for all active circuits on the chip (Figure 5.17). The specially designed Schottky diodes with low series resistance allow for a high-efficiency conversion of the received RF input signal energy to dc supply voltage.

The voltage multiplier circuit shown here is sometimes also called *charge pump* in the context of memory ICs. A charge pump is a circuit that when given an ac input, is able to output a dc voltage typically larger than a simple rectifier would generate. It can be thought of as an *ac-to-dc converter* that both rectifies the ac signal and increases the dc level. It is the foundation of power converters such as the ones that are used for many electronic devices today.

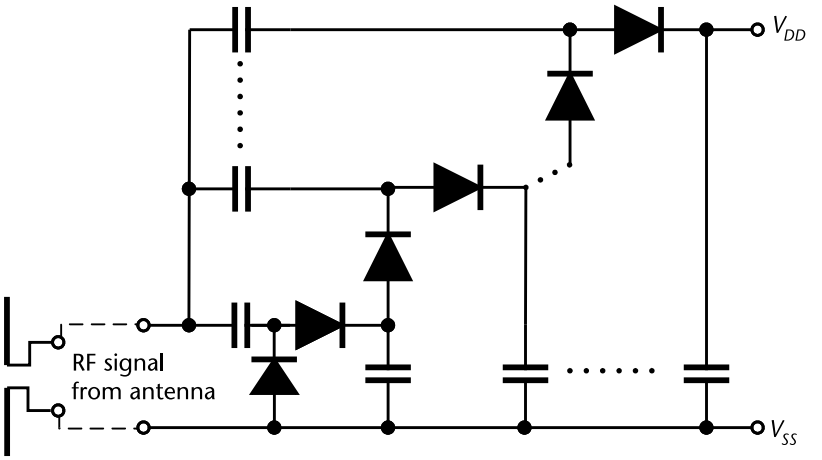


Figure 5.17 Input signal conversion to dc supply voltage.

In this case, for the RF signal, all the diodes are connected in parallel (or anti-parallel) by the capacitors. For dc, however, they are connected in series to allow a dc current flowing between terminals. The voltage generated between these nodes is approximately equal to:

$$V_{DD} = n\left(V_{RF} - V_D\right) \tag{5.43}$$

where n is the number of diodes, V_{RF} is the amplitude of the RF input signal, and V_D is the forward voltage of the Schottky diodes (approximately 200 mV at 7 μA).

The input impedance is mainly determined by the junction and substrate capacitances of the Schottky diodes. The real part of the impedance is much lower than the imaginary part and is strongly dependent on the dc current taken from the output. For a typical operating point, the real part of the impedance is approximately 30 times lower than the imaginary (capacitive) part. In other words, the IC's input capacitance has a quality factor of 30, placing high demands on the antenna, which needs to be matched to the IC's input impedance for sufficiently good power efficiency.

The design parameters of the voltage multiplier are a trade-off between power efficiency, useful impedance, and operating point (load). Optimization parameters include the number of stages, the size of the Schottky diodes, and the size of coupling capacitors.

5.3.9.2 Tag Wake-Up Circuit Principles

The interrogation of active RFID tags will inevitably involve the development of a mechanism for turning on the tags. The power conservation is an important factor that requires the tags to be turned off when not being interrogated. This will also be true for active sensors and sensor networks [10]. There are two practical options for turn-on circuits design:

- Rectifier circuits that can produce, from the RF field, a rectified voltage of the order of 1V, which will turn a CMOS transistor from fully off to fully on;
- Rectifier circuits that can produce, from the RF field, a rectified voltage of the order of 5 mV, which when compared to an internal reference voltage can be used to trigger a transistor from fully off to the fully on state.

Schottky diodes have been proven to perform very well in microwave networks for many years because of their excellent high-frequency behavior. For microwave applications, these Schottky diodes are usually fabricated in specialized processes where barrier heights, capacitances, and other parameters can be fully controlled. The RFID applications demand low-cost solution, therefore requiring the standard CMOS processes. However, the problem is that most of the standard CMOS processes do not support the Schottky diode, so the processes has to be modified in order to incorporate the Schottky diodes.

A Schottky junction is relatively delicate and sensitive to excessive RF power, and RFID applications may work in poorly controlled environments where high power may cause the diode to burn out. Hence, in an application it is important to use power limiters to protect the sensitive Schottky diode.

The battery powering active transponders must last for an acceptable time, so the electronics of the label must have very low current consumption in order to prolong the life of the battery. However, due to circuit complexity or the desired operating range, the electronics may drain the battery more rapidly than desired, so the use of a turn-on circuit allows the battery to be connected only when communication is needed, thus lengthening the life of the battery. The circuit shown in Figure 5.18 is adequate and cost-effective for a backscattering active tag.

Here, a *p-channel FET* was used as a switch to control the power supply to a label control circuits and can be triggered by the incident 915-MHz radiation on the antenna. Thus, the power generated and amplified by the diode resonance can be utilized to turn a *p*-channel FET from an off state to an on state. The turn-on

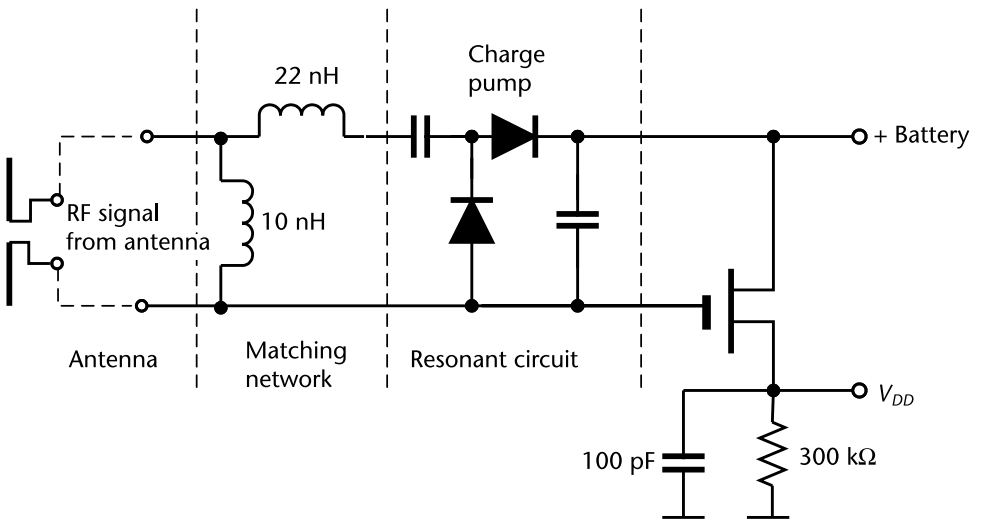


Figure 5.18 Turn-on circuit for the active UHF tag.

circuit performs adequately at a minimum power of –43 dBW at the resonant frequency of the 915 MHz.

5.3.10 Tag Manufacturing Process

The major processes involved in the manufacturing of RFID tags are antenna production process and chip assembly. Both of these activities have gone through heavy development in the last few years with the purpose of increasing the throughput and reducing the final cost of the tag.

5.3.10.1 Antenna Production Process

The most popular processes currently on the market for the production of bidimensional antennas are chemical etching and conductive ink printing [11].

Chemical etching is an established technology used for the realization of printed electronic circuit boards. To reach the necessary volume requirement, the technology has been modified adopting a roll-to-roll process where copper tape roll, typically 20 to 40 microns thick, is glued to a polymeric roll substrate, typically PET (polyethylene terphthalate is a thermoplastic polymer resin of the polyester family). The copper film is then masked with a photoresist mask and inserted in a chemical bath where the exposed copper is chemically attacked and removed, thereby creating the desired pattern. The photoresist mask is then removed using a standard process.

The strongest advantage of this process, originally invented for the development of flexible circuits, is its availability and well-known manufacturing cost parameters. The currently estimated cost for copper etching is $10 per square meter of produced material that, for 600 antennas per square meter, is equivalent to 1.6 cents per antenna.

Conductive ink printing is the process based on flexographic equipment, in which the antennas can be printed on a polymer roll in a single step with no need for masking. The major problem with this technology is the inherent cost of the material used in the process, usually a conductive ink loaded with 30% (or more) of silver flake particles, and the higher surface resistivity of the conductive layer, a feature inversely proportional to the final performance of the antenna.

Another concern is the inherent environmental impact of using silver; as RFID tags are used and disposed of, their increasing density within landfills around the world will eventually jeopardize ground water supplies, leading to requirements for recycling as is presently done for electronics.

5.3.10.2 Chip Assembly

The major challenge in assembling the active component onto the antenna circuit is represented by the small size of the chip itself, the need for low temperature attachment and the required throughput capacity. Chips for passive RFID tags had their size reduced to submillimeter dimensions in order to optimize the cost of the component, thus increasing the number of chips per wafer.

This, of course, creates technical problems when the chip, whose pads are now below the 100-μm size, needs to be connected to the antenna at high speed. To

overcome this problem, Pick-and-Place (PnP) equipment manufacturers have developed innovative technology enabling the attachment of the chip to the antenna roll using standard flip-chip technology and dispensing fast curing conductive adhesive at high speed. The process is quite simple and is capable of attaching about 10,000 components per hour, equivalent to about 70 million tags per year.

Some tag manufacturers have approached the problem from a different point of view, focusing their attention on the production capacity and thus purposefully adopting another process, the *strap attach*. In this solution, the chip is attached to a polymeric carrier film in roll form at high density and high velocity using roll-to-roll equipment.

This polymeric carrier has previously been prepared with small caves on the surface to receive the chip, which corresponds to large conductive pads. Proper geometry of the cave and chip die guarantees the proper positioning of the chip on the film. Once this step is completed, it is possible to couple the roll containing the straps with the roll containing the antennas and accomplish the final assembly by using dispensed adhesive. Of course, due to the different density and location of the chip on the strap carrier, each strap needs to be singulated before attachment.

The advantage of this process, as stated by its major supporters, is the possibility to assemble chips at a very high velocity. The disadvantage is that the process consists of two-step phases and that the singulation of the strap causes in some way a reduction in the speed of the process. While it is possible to load the straps at very high speed, the final assembly of the strap on the antenna is still a process regulated in its throughput by the curing of the adhesive used for attachment.

5.4 Readers

5.4.1 Principles of Operation

RFID tags are interrogated by readers, which, in turn, are connected to a host computer. In a passive system, the RFID reader transmits an energy field that wakes up the tag and powers its chip, enabling it to transmit or store data. Active tags may periodically transmit a signal, much like a lighthouse beacon, so that data may be captured by multiple readers distributed throughout a facility.

The reader is equipped with antennas for sending and receiving signals, a transceiver and a processor to decode data. Companies may need many readers to cover all their factories, warehouses, and stores. Readers typically operate at one radio frequency, so if tags from three different manufacturers used three different frequencies, a retailer might have to have multiple readers in some locations, increasing the cost further.

Readers may be portable handheld terminals, or fixed devices positioned at strategic points, such as a store entrance, assembly line or toll booth (gate readers.) In addition, readers/interrogators could also be mobile.

Fixed readers are designed for large scale deployments that need to process a large volume of assets at certain points and primary read zones, such as dock doors. Permanently installed in a defined location, these devices are constantly reading and always listening in order to detect any RFID tag that passes within the reader's active zone.

Handheld or *portable readers* are a very useful resource to supplement fixed readers. Handheld readers can be used instead of a portal reader to record boxes loaded and identify boxes as they are removed; little efficiency is gained relative to bar-coded labels, but customer mandates can be accommodated with minimal initial expense.

Handheld readers are also very useful for exception handling of boxes that fail to read at a portal or on a conveyor or that have misplaced or misoriented labels identifying boxes of unknown provenance, and so forth. Handheld readers can be useful for inventory cycle count in storage areas or temporary staging locations, for locating specific cartons in storage, for verifying manifests during assembly, and for specialized applications such as tail-to-tail baggage transfer (moving baggage from one airplane to another in an airport without routing it through the terminal).

The newest category of RFID readers are the *mobile RFID readers*. This cable-free device is completely self-contained, with integrated battery, antennas, and wireless communications capabilities. The mobile RFID reader can be used on material handling equipment, such as forklifts, clamp trucks and skate wheels or on other moving equipment like mobile carts. It can also be installed as a stationery device in hard to cable areas. Where the fixed RFID reader requires product to be brought to the reader, mobile readers are brought to the product—like the handheld RFID reader. However, while the handheld reader requires user intervention to read, the mobile RFID reader offers hands-free operation.

RFID readers are used to activate passive tags with RF energy and to extract information from the tag. For this function, the reader includes RF transmission, receiving, and data decoding sections. In addition, the reader often includes a serial communication (RS-232, USB, and so on) capability to communicate with a host computer. Depending on the complexity and purpose of applications, the reader's price range can vary from tens to thousands of dollars' worth of components and packaging. Typically, the reader is a read-only device, while the reader for a read-and-write device is often called *interrogator*. Unlike the reader for a read-only device, the interrogator uses command pulses to communicate with a tag for reading and writing data.

The *carrier* is the transmitted radio signal of the reader (interrogator). This RF carrier provides energy to the tag device and is used to detect modulation data from the tag using a backscattering. In read/write devices, the carrier is also used to deliver the interrogator's commands and data to the tag.

The RF transmission section includes an RF carrier generator, an antenna, and a tuning circuit. The antenna and its tuning circuit must be properly designed and tuned for the best performance. Data decoding for the received signal is accomplished using a microcontroller. The firmware algorithm in the microcontroller is written in such a way to transmit the RF signal, decode the incoming data, and communicate with the host computer.

The main criteria for readers include the following:

- Operating frequency: They could be LF, HF, UHF, and some companies are starting to develop multifrequency readers;
- Protocol agility: support for different tag protocols (ISO, EPC, proprietary);

- Different regional regulations (UHF example):
  - UHF frequency agility 902 to 930 MHz in the United States and 869 MHz in Europe;
  - Power limitation of 4W in the United States and 500 mW some other countries;
  - Manage frequency hopping in the United States and duty cycle requirements elsewhere.
- Networking to host capability:
  - TCP/IP;
  - Wireless LAN (802.11);
  - Ethernet LAN (10base T);
  - RS 485.
- Networking capabilities: Ability to network many readers together (via concentrators or via middleware);
- Upgrades: Ability to upgrade the reader firmware in the field;
- Management of multiple antennas:
  - Typically four antennas per reader;
  - How antennas are polled or multiplexed.
- Adapting to antenna conditions (dynamic auto-tuning):
  - Interface to middleware products;
  - Digital I/O for external sensors and control circuits.

Certain readers also provide connection options to enable simple process control mechanisms to be implemented, such as digital inputs and outputs with 24V, which can be used to control traffic lights or gates that are released once the tag data has been checked at the goods issue/receipt point.

Simple PLC couplings can also be realized using this technology. The higher protocol layers have not been standardized yet, resulting in additional time and effort when it comes to integrating readers across different manufacturers. In addition, readers and antennas at loading gates must be highly tolerant as regards temperature and must be protected against dust and damp.

Up until the recent surge in developments for the supply chain and EPC tags, readers were mainly used in access control systems and other low-volume RFID applications, which meant that the problem of treating very large numbers of tags and high volumes of data was not such a serious issue. This is all changing, and many reader manufacturers are developing next generation products to handle the application problems that will be specific to the supply chain and EPC/ISO infrastructure.

5.4.2 Reader Antenna

The reader antenna establishes a connection between the reader electronics and the electromagnetic wave in the space. In the HF range, the reader antenna is a coil (like the tag antenna) designed to produce as strong a coupling as possible with the tag antenna.

In the UHF range, reader antennas (like tag antennas) come in a variety of designs. Highly directional, high-gain antennas are used for large read distances. Regulatory authorities usually limit the maximum power emitted in a given direction; as a result, the transmission power emitted from the reader to the antenna must also be regulated accordingly. One advantage of highly directional antennas is that the reader power often has to be emitted only to the spaces in which the tags that are to read are located.

Generally speaking, physical interdependencies mean that the antenna gain is linked to the antenna size. The higher the gain (or the smaller the solid angle into which the antenna emits), the larger the mechanical design of the antenna will be. It follows, therefore, that highly directional antennas are not used for handheld readers. Antennas typically used for handheld readers include patch antennas, half-wave dipoles, and helix antennas. Larger antenna structures can be used for stationary readers; in the UHF range, they usually take the form of arrays.

All other things being equal, a high-gain antenna will transmit and receive weaker signals farther than a low-gain antenna. Omnidirectional antennas, such as dipole antennas, will have lower gain than directional antennas because they distribute their power over a wider area. Parabolic antennas usually have the highest gain of any type of antenna but not really useable in typical RFID applications, except maybe for microwave RFID readers where a long-range and narrow radiation pattern is required. A half-wave dipole antenna will have a gain of near unity or nearly equal the isotropic antenna.

Reader antennas may have different requirements depending on whether they are fixed, portable, or mobile readers. For example, selection of antennas for portable devices is dominated by size and weight constraints. Read range and polarization are generally less significant than in the case of fixed readers. Efficient use of RF power to maximize battery life is critical. Highly directive antennas are useful to reduce power consumption, but are generally physically large and thus may not fit portable applications. The fact that handheld reader antennas are small and light constrains the antenna gain.

Antennas that are less than about one-quarter of a wavelength in all dimensions (a quarter-wave is about 80 mm or 3.2 inches for UHF operation in the United States) cannot achieve more than about 4 dBi of gain. Slightly larger antennas allow up to about 6 dBi of gain, but make the reader somewhat bulky and awkward to carry. The trade-off is important because handheld and portable applications benefit from high antenna gain.

The reader is likely to employ less than the maximum allowed transmitting power to improve battery life, so read range is impacted if antenna gain is low. A narrow antenna beam will improve the ability of the user to locate the tag being read by changing the reader orientation and noting the results. The narrow beam of a high-gain antenna, which is undesirable in a stationary-reader application, is often beneficial for a handheld reader, since the user can readily move the beam to cover the area of interest.

5.4.3 Software Defined Radios in RFID Systems

The problem of continuous change in the EPC market is a vitally important for all RFID users and especially those responsible for buying and installing RFID reader

infrastructure. While tags are the consumables of the RFID systems, constantly varying, iterating and regenerating, the RFID reader infrastructure is a deployed capital expense that cannot easily or cost-effectively be replaced every time a new tag variant appears. All this change is good for the RFID user as it will deliver ever-improving performance, and decreasing costs.

Software-defined radio (SDR) uses software for the modulation and demodulation of radio signals. In other words, SDR is a radio communication system where components that have been typically implemented in hardware (mixers, filters, amplifiers, modulators/demodulators, detectors, and so on) are instead implemented by means of software.

An SDR performs the majority of its signal processing in the digital domain, most commonly in a digital signal processor (DSP), which is a type of microprocessor specifically optimized for signal processing functions. The advantage of an SDR-based RFID reader is that it can receive and transmit a new form of RFID communication protocol simply by running new software on existing SDR hardware.

A software-defined RFID reader consists of an RF analog front end that converts RF signals to and from the reader's antennas into an analog baseband or intermediate frequency signal, and analog-to-digital converters and digital-to-analog converters that are used to convert these signals to and from a digital representation that can be processed in software running on the reader's digital signal processor.

SDR technology has long been important in the military context, where new radio equipment must interoperate with legacy equipment, much of which is used for many years beyond its design lifetime. Additionally, the U.S. military is often called upon to work together with allies that have old, outdated equipment that is incompatible with the more modern U.S. communication hardware. This is exactly analogous to the Generation 1 to Generation 2 (and beyond) transition in RFID. Military SDR projects date back to the early 1990s, and several were fielded in that time frame. Aware of these developments, in 1999 the MIT Auto-ID Center began exploring the idea of using SDR in RFID readers.

5.4.4 Data Transfer Between a Tag and a Reader

5.4.4.1 Signal Transmission

For a RFID system to work, three processes are required: energy transfer, downlink, and uplink. According to this we can divide RFID systems into three groups, *full-duplex*, *half-duplex*, and *sequential*.

During full-duplex and half-duplex operation, the energy is transferred constantly, compared to sequential when energy is first transferred by the reader and then the tag responds.

In half-duplex systems the information is sent in turns either transferred inductively through load modulation or as electromagnetic backscatter. In full-duplex systems uplink information is sent on a separate frequency, either a subharmonic or not, so the flow of information can be bidirectional and continuous.

Sequential transfer consists of two phases: first, energy is sent to the tag that stores it in a capacitor, and then, utilizing the power received, it can function for some time and send its reply. This method has the advantage; by extending the

charging time and enlarging the capacitor, it is possible to accumulate more energy for the electronics.

5.4.4.2 Data Transfer Rate

A further influence of carrier frequency is with respect to data transfer, for which it is very important is to understand bit rate (data rate) concept.

Whereas in theory it is possible to transfer binary data at twice the carrier frequency, in practice it is usual to use many cycles of carrier to represent a binary digit or group of digits. However, in general terms, the higher the carrier frequency the higher the rate for data transfer that can be achieved. So, a low-frequency system operating at 125 kHz may transfer data at a rate of between 200 bps and 4,000 bps, depending upon the type of system. Rates up to greater than 100 kbps (but typically less than 1 Mbps) for are possible for microwave systems.

It should also be appreciated that a finite bandwidth is required in practice to transfer data, this being a consequence of the modulation that is used. Consequential to transfer capability is the data capacity of the tag. Loosely speaking, the lower the frequency, the lower the data capacity of the tags, simply because of the data required to be transferred in a defined time period. Keep in mind that the capacity can also be determined by the manner in which the tag is designed to be read or written to (for read/write tags).

The choice of data transfer rate has to be considered in relation to system transfer requirements; this is determined by the maximum number of tags that may be expected to be read in a unit interval of time multiplied by the amount of data that is required to be read from each tag. Where a write function is also involved, the number of tags and write requirements must also be considered.

ISO 15693 is an ISO standard for *vicinity cards*, that is, cards that can be read from a greater distance as compared to *proximity cards*. ISO 15693 systems operate at the 13.56-MHz frequency, and offer maximum read distance of 3 to 4 feet. In ISO 15693 chips, the subcarrier frequency is equal to 423.75 kHz (RF/32) with FSK or OOK modulation and Manchester data coding. The achievable label data transfer rate is up to a relatively fast 26.48 kbps. Most typical bit rate values in bps are RF/8, RF/16, RF/32, RF/40, RF/50, RF/64, RF/80, RF/100, and RF/128.

Every tag sends back information with some predefined, usually fixed bit rate, and once the manufacturer programs the data rate, it usually cannot be changed. This data rate is clocked by internal tag frequency. For LF transponders the range is from 100 to 150 kHz, depending on the manufacturer.

Consider, for example, transponder type that bit rate is RF/32. It means that data rate is 32 field clocks (FC) per logic 1 or 0 data bit. Data (bit) rate is a bit time duration and it is defined as field clocks per bit. Taking field clock equal to 125 kHz and tag bit rate equal to RF/32, data rate is 125 kHz/32 =3.9062 kbps, so receiving 64 bits of information would take $8\ \mu s \times 32 \times 64 = 16.384$ ms.

The manner in which the tags are interrogated is also important. It can be done by singulation (one at a time in the interrogation zone) or as a batch (a number of tags in the interrogation zone at the same time). The latter requires that the tags and associated system has *anticontention* (or *anticollision*) capabilities so that collisions between responses from tags in the zone at the same time can be resolved and contention avoided. Various anticontention protocols have been devised and

applied with various levels of performance in respect of the number of tags that can be handled and the time required handling them. So, the anticontention performance may be an important consideration in many applications.

5.4.4.3 Read/Write Range

The read/write range is the communication distance between the reader (interrogator) and tag. Specifically, the read range is the maximum distance to read data out from the tag and the write range is the maximum distance to write data from interrogator to the tag.

The read/write range is, among other effects, mainly related to:

- Electromagnetic coupling of the reader (interrogator) and tag antennas;
- The RF output power level of reader (interrogator);
- Carrier frequency bands;
- The power consumption of the device;
- Antenna orientation;
- The distance between the interrogator and the tag;
- Operating environment conditions (metal, electric noise, multiple tags, multiple readers, and so on);
- The tag and the tag's dwell time.

The tag's dwell time is the time a tag is in the interrogator's RF field. An RFID interrogator's read range is the distance between the interrogator and the RFID tag at which the signals from the tag can be read properly. Similarly, an RFID interrogator's write range is the maximum distance at which information within the RF signal from the interrogator can be received correctly and stored within the memory of the tag's microchip.

More power is needed to write to a tag than to read it; as a result, the tags need to be closer to the antenna to write than to read. The general rule is that the write range is 50% to 70% of the read range of a particular interrogation zone.

Power limitations, as listed in Table 5.3, are imposed by local authority and cannot be chosen arbitrarily. The standardization of RFID technology, as well as the requirements of the local governing bodies are still in progress and change constantly. For that reason, some of the information provided in this book that was correct during the preparation of the manuscript might change by the time it reaches the reader.

The electromagnetic coupling of the reader and tag antennas increases using a similar size of antenna with high Q on both sides. The read range is improved by increasing the carrier frequency. This is due to the gain in the radiation efficiency of the antenna as the frequency increases. However, the disadvantage of a high-frequency (900-MHz to 2.4-GHz) application is shallow skin depth and narrower antenna beamwidth, causing less penetration and more directional problems, respectively.

Low-frequency application, on the other hand, has an advantage in the penetration and directivity, but a disadvantage in the antenna performance. Read range

Table 5.3 RFID Power Limitations Based on the Region and Frequency

Frequency Band	*Power, Limitations, Region*
125 kHz	Inductively coupled RF tags
1.95, 3.25, and 8.2 MHz	Inductively coupled theft tags, worldwide
13.56 MHz	Inductively coupled RFID tags, worldwide
27 MHz and 40 MHz	0.1-W ERP, Europe
138 MHz	0.05-W ERP, duty cycle < 1%, Europe
402–405 MHz	Medical implants, 25 μW ERP (–16 dBm)
433.05–434.79 MHz	25-mW ERP, duty cycle < 10%, Europe
468.200 MHz	0.5-W ERP, Europe
869.40–869.65 MHz	0.5-W ERP, duty cycle < 10%, Europe*
902–928 MHz	4-W EIRP, United States
2,400–2,483.5 MHz	ISM band, 0.5-W EIRP Europe, 4-W United States, Bluetooth
5,725–5,875 MHz	25-mW EIRP

*To accommodate concerns over Gen 2 RFID systems ability to perform under the European regulations, Europe has already increased its available frequency spectrum from 2 to 8 MHz, allowable power output level from half a watt to 2W, and replaced its 10% duty cycle restriction with a listen-before-talk requirement. Even with these improvements, work is still underway to further alleviate European regulatory constraints.

increases by reducing the current consumption in the silicon device. This is because the LC antenna circuit couples less energy from the reader at further distances. A lower power device can make use of less energy for the operation.

For LF and HF (near-field) systems, to increase the magnetic field at the tag's position, the reader/writer antenna coil's radius must be increased, or the current in the antenna coil must be increased, or both. The strength of the magnetic field attenuated in proportion to the inverse of the cube of distance. By increasing the diameter of the RFID tag's antenna coil, the signal induced in the tag's coil could be increased.

Accordingly, for applications that require long-range operation, the reader/writer antenna coil's radius and tag antenna coil's dimensions must be increased. The primary goal is to design an antenna that maximizes RFID read range. In tests comparing coin-sized and IC card-sized tags using the same reader/writer, the IC card-sized tag had an operating range several times larger than the coin-sized tag.

The read range of UHF-based RFID (propagation) system can be calculated by the Friis free space equation as follows:

$$r = \frac{\lambda \cos\theta}{4\pi}\sqrt{\frac{P_R G_R G_T\left(1-(\Delta\rho)^2\right)}{P_{th}}} \quad \text{for} \quad 0 \le (\Delta\rho)^2 \le 1 \tag{5.44}$$

where G_T is the gain of the tag antenna, $P_R G_R$ is an EIRP of the reader, λ is the wavelength, P_{th} is the minimum threshold power required to power an RFID tag, θ is the angle made by the tag with the reader plane, and $(\Delta\rho)^2$ is the power reflection coefficient, which is the ratio of reflected power to incident power by the tag. Factor $1 - (\Delta r)^2$ is also called *mismatch factor*.

Note that the power received by the tag is inversely proportional to the square of the distance between the tag and the reader's antenna. Studies reveal that the orientation of tag in the RF field affects its read range. In the specific context of a

directivity pattern, a perfectly parallel tag, relative to the reader's antenna, yields maximum read range, while a tag perpendicular to the base station antenna's field has minimum to zero read range. Thus, efforts are made to make the tag parallel to the reader antenna by deploying one or more of the following measures:

- Change in orientation of the reader antenna to suit the orientation of the tag antenna;
- Use of redundant antennas for ensuring proper alignment of at least one reader antenna to the tag antenna;
- Increase reader antenna power (of course, within the limits allowed by the local authorities) to reduce the effect of tag orientation;
- Increase the polling rate of the antenna to make more reads in the same sampling time.

The far-field formula is correct, assuming that polarization of reader antenna and the polarization of tag antenna are perfectly matched. However, in fact, the polarization mismatch is essential and required in most RFID applications. The point is that in the majority of applications the tag is allowed to appear in almost arbitrary position in the field of the reader antenna while the polarization of the tag antenna is usually linear because of required small size of the tag.

In such situation, the only way to fulfill a system requirement is to use circularly polarized reader antenna. Thus, a sacrifice of 3-dB power loss (at least, although it can be even much higher, see the discussion on polarization mismatch in Chapter 2) due to polarization mismatch between circularly polarized reader antenna and linearly polarized tag antenna overcomes the problem of tag orientation. This is why, nowadays, the major vendors offer mainly circularly polarized reader antennas (except for the handheld readers).

At the same time, the linearly polarized antennas are also available in the market for limited RFID applications. In the case of linearly polarized both reader and tag antennas, the substantial polarization misalignment may cause severe power loss. In the case of circular-to-linear polarization mismatch, the read range, r, will be $\sqrt{2}$ times shorter than the one calculated by (5.43).

As we can see from the Table 5.4, systems operating in the 915-MHz band may achieve read ranges of 20 feet (6m) or more under current FCC regulations.

5.4.4.4 Environment and Proximity to Other Objects

Up to now, our considerations have focused upon data transfer across an uncluttered air interface. However, free-space propagation where reader and tag are distanced from any obstructions or other tags, and perfectly aligned relative to each other, is not a realistic situation. In practice, the region between the tag and interrogator may contain obstacles and materials that can influence the performance of the system.

The carrier frequency is one of significance with respect to the effects that the prevailing conditions and clutter factors (obstacles and physical structures) can have. In low- and high-frequency inductive RFID systems, the magnetic field is effectively used to couple data, and such fields are largely unaffected by dielectric or

Table 5.4 Read Range for Different UHF Reader Powers and Reflection Coefficients

Frequency (MHz)	*Wavelength (m)*	*Reader Power (W)*	*Reader Power (dBm)*	*Reader Antenna-Gain*	*Tag Antenna Gain*	*Power Reflection Coefficient Δρ*	*Angle (°)*	*Tag Threshold Power (dBm)*	*Tag Threshold Power (mW)*	*Read Rage (m)*	*Read Range (feet)*
915.00	0.33	0.50	26.99	1.64	1.64	0.40	0.00	−10.00	0.10	2.77	9.10
915.00	0.33	0.50	26.99	1.64	1.64	0.50	0.00	−10.00	0.10	2.62	8.59
915.00	0.33	0.50	26.99	1.64	1.64	0.60	0.00	−10.00	0.10	2.42	7.94
915.00	0.33	2.44	33.87	1.64	1.64	0.40	0.00	−10.00	0.10	6.13	20.09
915.00	0.33	2.44	33.87	1.64	1.64	0.50	0.00	−10.00	0.10	5.79	18.99
915.00	0.33	2.44	33.87	1.64	1.64	0.60	0.00	−10.00	0.10	5.35	17.54

insulator materials (papers, plastics, masonry, and ceramics, for example); the field simply penetrates the materials.

Metals, on the other hand, can distort the field, depending upon how ferrous they are. This will weaken the field strength in regions of the interrogation zone, in some cases to the extent that the system performance is impaired. The range capability may be impaired or ability to read or write to a tag may be impaired. For uncompensated tag designs operating at resonant frequencies, the presence of metals can often detune the device, in some cases preventing it to operate.

At higher frequencies (UHF and above), for propagation RFID, the electric component of field becomes more significant. The higher the frequency, the more easily they can penetrate dielectric materials. However, for some materials where energy exchange mechanisms can be identified at or near the carrier frequency, this can result in energy absorption from the propagating wave, hence causing an impairment of range performance.

One of the challenges with UHF RFID tags is working well in the presence of water or metal. Unfortunately, the human body is made up of mostly water. Thus, if the RFID tag is placed close to (or implanted inside) the human body, performance will suffer.

As far as metals are concerned, they reflect or scatter higher-frequency signals, depending on the size of the metal object in relation to the wavelength of the incident signal. Such effects can impair the range that can be achieved and, in some cases, can screen the reader from the tag and prevent it from being read. Any metal near the tag, such as keys or coins, may also cause the tag to be undetectable.

The *proximity of tags* may also exhibit a similar effect. Reflections and diffraction effects can often allow pathways around metal objects within an interrogation zone. Because it is difficult to generalize on the effects of clutter within the interrogation zone, it is expedient where possible to avoid clutter and choose a carrier frequency that is appropriate to the conditions to be expected.

Passive RF tags in the UHF and microwave bands have drawn considerable attention because of their great potential for use in many RFID applications [12]. However, more basic research is needed to increase the range and reliability of a passive RF tag's radio link, particularly when the RF tag is placed onto any lossy dielectric or metallic surface. This radio link budget is dependent upon the *gain penalty losses* (L_{GP}), a term which quantifies the reduction in RF tag antenna gain due to material attachment.

After combining (5.17) and (5.22), we get the following expression:

$$P_{REC} = \frac{P_R G_R^2 \lambda^2 \sigma}{(4\pi)^3 r^4} = \frac{P_R G_R^2 G_T^2 \lambda^4 (\Delta\rho)^2}{(4\pi)^4 r^4} \tag{5.45}$$

The assumption in this case is that the gain of the reader's transmit and receive antennas are the same, which may not always be the case; the reason is that the single-antenna readers are inexpensive and compact but need excellent matching circuits, high-isolation coupler with extremely high isolation between ports, and electronic circuitry with wide dynamic range.

In the logarithmic form, the same expression for the backscattered power received at the reader looks much simpler:

$$R_{REC} = P_R + 2G_R + 2G_T + 20\log(\Delta\rho) + 40\log\left(\frac{\lambda}{2\pi}\right) - 40\log r \tag{5.46}$$

where $\Delta\rho$ is a reflection change between switched loads.

Now, we can include in (5.45) additional losses due to the antenna being attached to different types of materials in a form of an adjustment for on-object degradation. As a result, we get:

$$P_{REC} = P_R + 2G_R + 2G_T + 20\log(\Delta\rho) + 40\log\left(\frac{\lambda}{2\pi}\right) - 40\log r - 2L_{GP} \tag{5.47}$$

A series of measurements was used to measure the far-field gain pattern, and gain penalty of several flexible 915-MHz antennas when attached to cardboard, pine plywood, acrylic, deionized water, ethylene glycol, ground beef, and an aluminum slab. It is shown that the gain penalty due to material attachment can result in more than 20 dB of excess loss in the backscatter communication link.

From the reader's prospective, handheld and portable antennas are very likely to operate in proximity to people's hands and arms, as well as other obstacles. The amount of power reflected from the antenna, measured by its reflection coefficient or return loss, should ideally be unaffected by such obstacles unless they are actually within the antenna beam.

The best return loss performance in presence of near-field objects is usually obtained from balanced antennas, in which the two halves of an antenna are driven by precisely opposed currents, and there is no large ground plane. However, such antennas are relatively large compared to single-ended (nonbalanced) antennas and require a balanced-unbalanced transformer (*balun*) to connect them to ground-referenced antenna cables or circuit board connectors. With a balun, the antenna is less sensitive to the presence of near-field objects.

5.4.5 UHF Reader Electronic Circuitry

To shrink the size of the RF portion of an RFID reader, it is necessary to increase the functions in each element. Figure 5.19 shows a typical block diagram of an RFID reader and shows one possible way of integrating elements into a chipset. Each module will be briefly described in the following sections.

5.4.5.1 UHF Reader Source Module

The purpose of the *source module* is to provide a synthesized *local oscillator* (LO) for transmitting (Tx) and receiving (Rx) paths in an RFID reader. The updated FCC standard requires frequencies to be within 10 ppm over the operating temperature ranges. It is necessary to amplify the signal after the synthesizer, in order to provide adequate LO input to the Tx and Rx signal paths due to typical synthesizer output powers and the loss of the power divider.

For a source module, it is critical that a single PC board footprint can be used to handle all the different bands. Using an integrated synthesizer/voltage-controlled

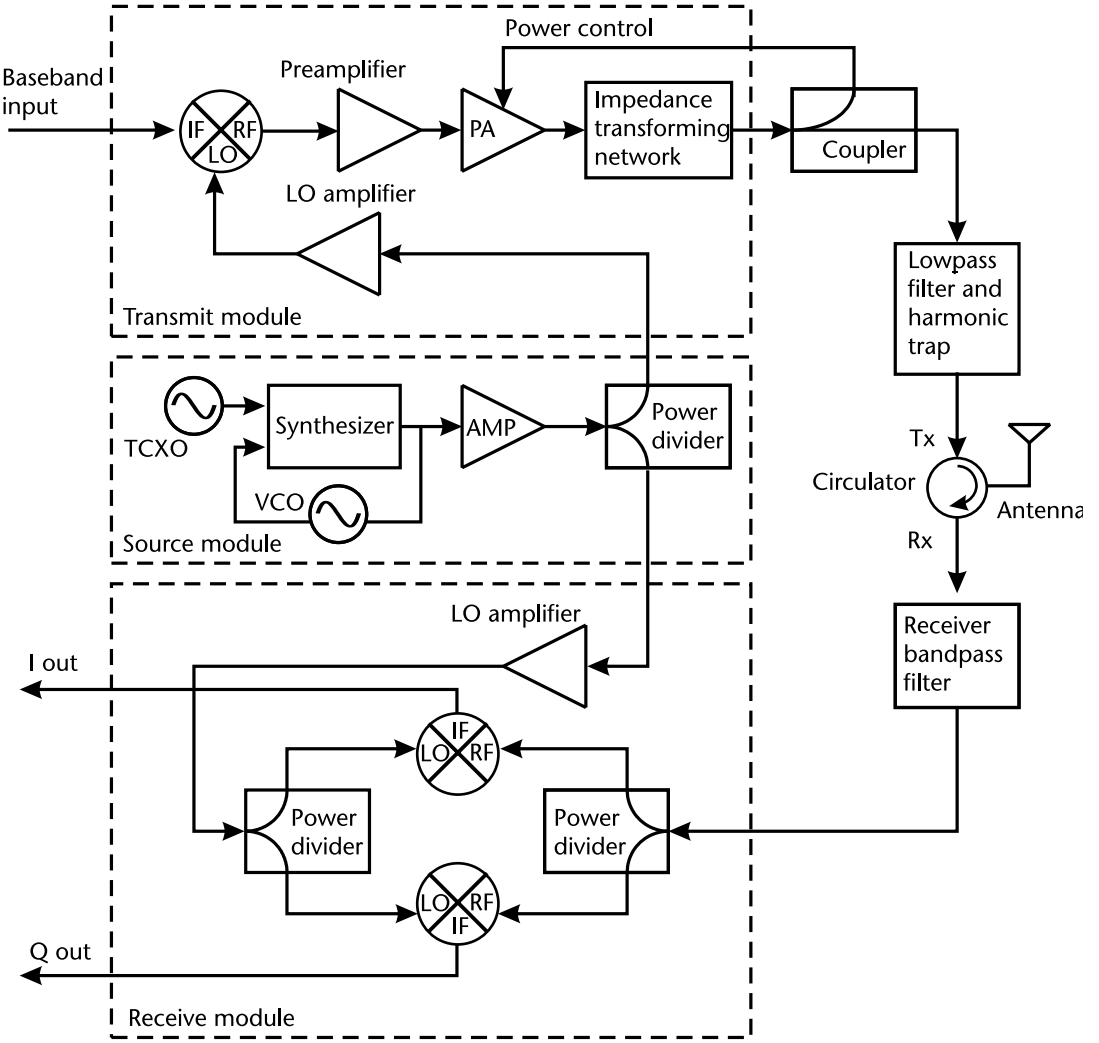


Figure 5.19 UHF RFID chipset block diagram.

oscillator (VCO) IC, it is possible to center the VCO bands by using different inductor values. The Japanese band requires a faster switching speed than the U.S. and European bands, which can still be realized with a 5-kHz bandwidth loop filter, but with different component values. It is desirable to have isolation on the order of 20 dB in the power divider.

For cost reasons, monolithic narrowband power dividers are generally used and are not optimal for covering 850 to 960 MHz. To optimize the isolation for each, tuning inductors and/or capacitors are used to recenter the power divider isolation. To shrink the size and reduce overall component count, it is necessary to combine somewhat diverse parts to create a source module. An additional requirement, that is typical of synthesizer/source modules, is that shielding is required for loop stability and minimization of phase noise.

5.4.5.2 UHF Reader Transmitting Module

As depicted in Figure 5.19, a typical transmitting module would include a double balanced modulator (DBM), LO amplifier, preamplifier, power amplifier, and impedance transforming network (ITN).

The high level of integration, with over 50 dB of available small signal gain, requires careful module layout. To maintain stability, it is necessary to keep the preamplifier located as far as possible from the power amplifier. The DBM provides a means to modulate the carrier signal. An LO amplifier is included to raise the signal available from the source module to a level sufficient to drive the mixers. Additionally, having the LO amplifier provide a 50Ω interface allows for simple interconnection to the source module. The modulated RF output from the mixer goes to a preamplifier and then to a power amplifier. The preamplifier has a gain of 17 dB and the power amplifier, implemented as a three-stage device, provides a small signal gain of 35 dB.

The purpose of ITN is to transform the 50Ω load impedance to a level that the power amplifier needs to drive in order to produce the desired output power at the available supply voltage. For a typical supply of 3.6 V, this impedance is only a few ohms, creating large circulating currents. These low impedances necessitate proper handling of circuit parasitics. This circuit requires careful design and implementation from performance and reliability perspectives.

To be able to provide the desired 1-W RF level at the antenna terminals typical for UHF readers, the power amplifier needs to be capable of providing sufficient power output capability to overcome the signal losses introduced between the transmitting module output and the antenna. These losses would include any coupler, filter, circulator, connector, and cabling used in the path to the antenna. It is desirable to control the power in order to both set the output level to various requirements and to implement a commonly used form of carrier amplitude modulation called a *pulse-interval modulation*, which is used to interrogate tags. The modulation bandwidth must be sufficient for the intended data rates without significant distortion, but as much circuitry as possible should be broadband.

The transmitter transmits encoding data with ASK modulation, including DSB-ASK, and SSB-ASK for forward link, and send an unmodulated carrier for return link. The maximum output power from PA is restricted to 1 W (30 dBm).

5.4.5.3 UHF Reader Receiving Module

Most modern wireless communication systems use digital modulation/demodulation techniques, and there is a good reason for this. They provide increased channel capacity and a greater accuracy in transmitted and received messages in the presence of noise and distortion.

In digital communication systems, a finite number of electrical waveforms or symbols are transmitted, where each symbol can represent 1 or more bits. It is the job of the receiver to identify which symbol was sent by the transmitter even after the addition of noise and distortion. Distortion in wireless communication can be caused by several mechanisms, such as passing a signal through filters having insufficient bandwidth or inefficient switching of nonlinear elements. Ultimately,

the effects of such events within communication systems are termed intersymbol interference (ISI).

In addition to ISI, there are other types of distortion more notably termed *delay spread* and *noise*. Delay spread occurs when multiple versions of the same signal are received at different times. This occurs when the transmitted signal reflects off multiple objects on its way to the receiver (*multipath*).

System designers are focusing their attention on their transceivers in search of a method or components that might help them achieve a superior signal-to-noise ratio, resulting in a lower bit-error rate (BER). It is widely projected that one of the reasons for the delay in a wide scale adoption of RFID systems has been the unacceptable BER of RFID tag reading.

In addition, RFID systems operating in the UHF band have unique attributes; in operation, the reader antenna emits electromagnetic energy in the form of radio waves that are directed toward an RFID tag. The tag absorbs energy, and through its built-in microchip/diode, uses it to change the load on the antenna, which in turn reflects back an altered signal to the reader. This method is known as backscatter and is the basis by which a passive RFID tag identifies its presence. These backscattered signals are essentially at the same frequency as that of the transmitted signal.

The backscattered signal antenna received is sent to the receiver through a directional coupler. The receiver front end must be designed to withstand high-interference signal levels without introducing significant distortion spurs. The receiver noise needs to be low enough that the system has sufficient dynamic range to allow error-free detection of low-level responding tag signals.

Homodyne detection, whereby a sample of the transmitted signal prior to modulation is used as the LO source for the receive I/Q demodulator, is utilized. Having both the transmitted and received signals at the same frequency exacerbates the difficulty of recovering the weak reflected signal, because it has to be identified in the presence of the higher powered carrier frequency. Consequently, it is an advantageous to choose transceiver components that help improve the overall signal-to-noise ratio as well as minimize LO carrier leakage.

The I/Q demodulator is a key element that can be used to maximize the signal-to-noise ratio and to minimize LO carrier leakage. Direct conversion to baseband frequency with the lowest BER and the highest sensitivity possible is crucial, not only for reader accuracy, but also to its range of usage.

5.5 RFID Power Sources

RFID tags need power to sense, compute, and communicate, which is further classified into three categories: *storage* (batteries, capacitors), *energy harvesting mechanisms* (vibrations/movement, photovoltaic, thermal gradient, and so on), and *energy transfer* (inductive coupling, capacitive coupling, backscatter).

Because many of these devices are expected to operate with minimum of human intervention, optimizing power consumption is a very important research area. RFID tags may derive the energy to operate either from an on-tag battery or by scavenging power from the electromagnetic radiation emitted by tag readers.

Storage refers to the way devices store power for their operation done either by using batteries or by using capacitors. Batteries are used when a longer life is

required and capacitors are used in applications that require energy bursts for very short durations [13].

5.5.1 Power-Harvesting Systems

Power harvesting (sometimes termed *energy scavenging*) is the process of acquiring energy from the surrounding environment (ambient energy) and converting it into usable electrical energy; the self-winding watch is a historical example of a power-harvesting device. The watches were wound by cleverly extracting mechanical energy from the wearer's arm movements.

Some of the other common energy conversion techniques are piezoelectric, thermoelectric, and so on. The energy transfer mechanisms are inductive coupling, capacitive coupling, and passive backscattering.

In medical devices, for example, a patient's normal daily activities could power an implantable pump that delivers insulin to a diabetic. The use of piezoelectric materials to harvest power has already become popular; piezoelectric materials[8] have the ability to transform mechanical strain energy into electrical charge. Piezo elements are being imbedded in walkways to recover the "people energy" of footsteps. They can also be embedded in shoes to recover *walking energy*.

Energy transfer is the way by which passive RF devices are powered. Inductive coupling is the transfer of energy between two electronic circuits due to the mutual inductance between the two circuits. Similarly, capacitive coupling is the transfer of energy between two circuits due to the mutual capacitance between the two circuits. Passive backscattering is a way of reflecting back the energy from one circuit to another.

Passive RFID tags obtain their operating power by harvesting energy from the electromagnetic field of the reader's communication signal. The limited resources of a passive tag require it to both harvest its energy and communicate with a reader within a narrow frequency band as permitted by regulatory agencies. A passive tag's power comes from the communication signal either through inductive coupling or far-field energy harvesting.

Inductive coupling uses the magnetic field generated by the communication signal to induce a current in its coupling element (usually a coiled antenna and a capacitor). The current induced in the coupling element charges the on-tag capacitor that provides the operating voltage, and power, for the tag. In this way, inductively coupled systems behave like loosely coupled transformers. Consequently, inductive coupling works only in the near field of the communication signal.

Far-field energy harvesting uses the energy from the interrogation signal's far-field signal to power the tag. The signal incident upon the tag antenna induces a voltage at the input terminals of the tag. This voltage is detected by the RF front-end circuitry of the tag and is used to charge a capacitor that provides the operating voltage for the tag. In the far field, tag-to-reader communication is achieved by modulating the RCS of the tag antenna (backscatter modulation).

8. The most common devices today that turn mechanical stress into electricity utilize piezoelectric materials. The phenomenon was first observed in 1880 by the brothers Pierre and Jacques Curie, but did not find a practical application until World War I when it was used by the French to develop an early form of sonar.

There is a fundamental limitation on the power detected away from a reader antenna. In a lossless medium, the power transmitted by the reader decreases as a function of the inverse square of the distance from the reader antenna in the far field. A reader communicates with and powers a passive tag using the same signal. The fact that the same signal is used to transmit power and communicate data creates some challenging trade-offs.

First, any modulation of the signal causes a reduction in power to the tag. Second, modulating information onto an otherwise spectrally pure sinusoid spreads the signal in the frequency domain. This spread, referred to as a *sideband*, along with the maximum power transmitted at any frequency, is regulated by local government bodies in most parts of the world. These regulations limit the rate of information that can be sent from the reader to the tag. RFID systems usually operate in license-exempt ISM bands, where the emitted power levels and the side band limits tend to be especially stringent.

The signaling from the tag to the reader in passive RFID systems is not achieved by active transmission. Because passive tags do not actively transmit a signal, they do not have a regulated limit on the rate of information that can be sent from the passive tag to the reader. Passive tags obtain impinging energy during reader interrogation periods, and this energy is used to power tag IC. In the near-field, tag to reader communication is achieved by modulating the impedance (load modulation) of the tag as seen by the reader [14]. For the maximum reading range, one has to ensure the maximum power transfer efficiency from the reader to the tag.

What makes the problem challenging is that in the case of inductively coupled reader and tag, the reader must deal with a changing effective load due to the location-dependent mutual coupling effect between the reader and tag as well as unpredictable number of tags in the read zone of the reader.

The powering of and communication with passive tags with the same communication signal place restrictions on the functionality and transactions of which the tags are capable. Governmental regulations can further limit communication timings. In the U.S. 915-MHz ISM band, regulations require that, under certain operating conditions, the communication frequency change every 400 ms. Because every change in frequency may cause loss of communication with a tag, transponders must not be assumed to communicate effectively for longer than 400 ms.

5.5.2 Active Power Sources

5.5.2.1 Batteries

Battery-assisted backscatter tags have their own power source to preenergize the silicon chip. The data is sent and received from the reader otherwise in the same way as a passive tag. This is a benefit where many tags are present in an interrogation zone; if they are all passive, they all need a lot of energy initially to reach sufficient voltage to turn on. With metals and fluids near tags, this is even harder due to interference and blind spots in the field. An on-board power source on each tag helps to overcome this.

Primary lithium has been a favorite option in this market, as the chemistry offers several positive factors including high energy-density, long life (approximately

10 years), and long storage life. Additionally, this chemistry is ideal for RFID tag applications because it is lightweight.

For RFID tag systems, primary lithium/manganese dioxide (Li-MnO_2) and lithium-thionyl chloride (Li-$SOCL_2$) are the two types of batteries that are most common. Lithium batteries offer a set of performance and safety characteristics that are optimal for RFID tag applications. Li-MnO_2 is relatively safe, compared to volatile lithium batteries, such as lithium-sulfur dioxide (Li-SO_2) and lithium-thionyl chloride (Li-$SOCL_2$), and does not develop any gas or pressure during battery operation.

However, one main disadvantage is that a single Li-MnO_2 cell cannot operate at voltages greater than 3V. These are typical in high-pulse applications that Li-SO_2 and Li-$SOCL_2$ can satisfy. Li-MnO_2 cells are best suited for applications that have relatively high continuous or pulse current requirements. However, because most electronic components used in RFID tags require a minimum operating voltage of 3V, at least two Li-MnO_2 cells must be connected in series to ensure a proper margin of safety for system reliability. This requirement adds weight and cost while potentially decreasing reliability due to increased part count.

Overall, the Li-MnO_2 chemistry has a high energy density, and has the ability to maintain a high rate of discharge for long periods of time. It can be stored for a long time (typically between 5 and 10 years) due to its low self-discharge rates. It also has the capability to supply both pulse loads and maintain a constant discharge voltage. Li-MnO_2 cells can operate in temperatures ranging from –20° to +70°C, although storage in temperatures exceeding +55°C is not highly recommended, and operation will be below full energy capacity at low temperatures. Their nominal voltage is typically 3V, which is twice the amount of that found in alkaline manganese batteries.

Li-$SOCL_2$ is a low-pressure system that is considered superior to lithium-sulfur dioxide systems in terms of high temperature and/or unusual form factor applications. Due to its low self-discharge rate, Li-$SOCL_2$ has a shelf life of a maximum of 10 to 15 years. This service life is the same for all construction, whether it is cylindrical, coin, or wafer. This chemistry also has the highest open-circuit voltage of 3.6V.

For most applications, only one cell of Li-$SOCL_2$ is required to maintain sufficient operating voltage. This is true as long as one cell can provide enough current to uphold the operating lifetime. RFID tag applications require very low continuous current and moderate pulse current, which Li-$SOCL_2$ batteries have no problem providing.

5.5.2.2 Other Power Sources

In the design of mobile electronics, power is one of the most difficult restrictions to overcome, and current trends indicate that it will continue to be an issue in the future. Designers must weigh wireless connectivity, CPU speed, size, and other functionality versus battery life in the creation of any mobile device.

Power generation from the user may alleviate such design restrictions and may enable new products such as batteryless on-body sensors. Power may be recovered passively from body heat, arm motion, typing, and walking or actively through user actions such as winding or pedaling. In cases where the devices are not actively

driven, only limited power can generally be scavenged (with the possible exception of tapping into heel-strike energy) without inconveniencing or annoying the user.

Clever power management techniques combined with new fabrication and device technologies are steadily decreasing the energy needed for electronics to perform useful functions, providing an increasingly relevant niche for power harvesting. Current and historical devices have shown that such mechanisms can be practical and desirable, yet much work remains in the creation and exploitation of these microgenerators.

RFID technology was thought to be a passive technology because the tags had no batteries; they just collected energy from the reader and send back their information. New advancement in the technology allowed the development of enhanced tags (active RFID) whose function fills the gap between the RFID traditional field and wireless sensor networks field.

Recent research has revealed *nuclear power*[9] as possible source of power for wireless sensor and RFID networks [15]. While certainly a little bit frightening topic at the first thought, note that the isotopes used in the actual prototypes penetrate no more than 25 μm in most solids and liquids, so in a battery they could be safely contained by a simple plastic package.

The huge amount of energy these devices can produce is illustrated by the following numbers: the energy density measured in mWh/mg is 0.3 for a lithium-ion battery, 3 for a methanol-based fuel cell, 850 for a tritium-based nuclear battery, and 57.000 for a polonium-210 nuclear battery. The current efficiency of a nuclear battery is around 4%, and current research projects (e.g., as part of the new DARPA program called Radio-Isotope Micropower Sources) aim at 20%.

To make a little more sense out of these figures, for example, with 10 mg of polonium-210 (contained in about 1 mm^3), a nuclear battery could produce 50 mW of electric power for more than 4 months.

A novel power supply for implantable biosensors has been described by Goto et al in [16]. In this power supply, near infrared light (NIR) transmission recharges a lithium secondary battery wirelessly through the skin. The Sun is a good source of NIR light, and its use requires no other external device to deliver energy to the recharging system. A photovoltaic cell array and the rectifier using Schottky diodes implanted beneath the skin can receive NIR light through the skin and charge the battery that is directly powering an implanted biosensor(s).

In 2008, a U.K.-based consortium of companies has successfully designed and clinically tested an in-body microgenerator that converts energy from the heartbeat into power for implanted medical devices. The microgenerator could help power implanted medical devices by augmenting the existing battery for devices such as cardiac pacemakers and implanted defibrillators. In preclinical testing, the microgenerator successfully produced one-third of the energy required to power a conventional cardiac pacemaker [17].

9. Most smoke detectors and even some emergency exit signs already contain radioactive material.

5.6 Review Questions and Problems

1. The reader produces a magnetic field that triggers the tag as shown in Figure 5.20. When the reader receives the transmitted data, it interprets the data and takes appropriate action. When the transponder enters the field produced by the reader, the coil produces a voltage inside the tag. In a passive transponder, this voltage can be used to power the tag. In an active transponder, the voltage is used to wake the tag and use its internal battery.

   Active transponders generally have longer read distances and shorter operational life and are larger and more costly to manufacture. Passive transponders are generally smaller, have a longer life, and are less expensive to manufacture. For optimum performance, the transponder coil is used in a parallel LC circuit designed to resonate at the operating frequency of the reader.

   - Calculate the capacitor value for a 4.9-mH transponder coil operating at 125 kHz. *(Answer: C = 331 pF.)*
   - What would be the resonant frequency f_1 if the overlapped tags have a new total inductance of 5.5 mH? *(Answer: f_1 = 117.96 kHz.)*
2. What do you think about the idea of passive RFID devices for locating small children?
3. You want a RFID tag that supports longer distance communications and does not rely on the reader to provide power to the tag. What kind of tag(s) do you need? Discuss advantages and disadvantages of these tags.
4. Are there any health risks associated with RFID and its radio waves? Discuss.
5. Formula for EIRP in dBm is as follows:

$$EIRP = \frac{E^2 r^2}{30}$$

   where EIRP is power in watts, E is in V/m, and r is distance in meters.
   a. Show the EIRP in dBm, using E in dBμV/m and r in meters.
   (Answer: $EIRP_{[dBm]} = E_{[dB\mu V/m]} + 20\log r_{[meters]} - (10\log 30 + 90)_{[dB]}$.)

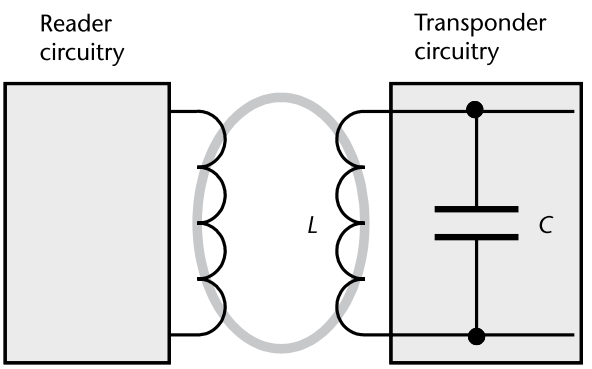


Figure 5.20 RFID system and a resonant frequency calculation.

b. In standard test setups, the electrical field strength is often measured at a distance of 3m. Show that in this case we can use the simplified formula:

$$EIRP_{[dBm]} = E_{[dB\mu V/m]} - 95.23_{[dB]}$$

6. The chip device turns on when the antenna coil develops 4 V_{PP} across it. This voltage is rectified and the device starts to operate when it reaches 2.4 V_{DC}.
   a. Calculate the magnetic field to induce a 4 V_{PP} coil voltage with an ISO standard 7810 card size (85.6 × 54 × 0.76 mm) using the coil voltage equation. The frequency is 13.56 MHz, number of turns is 4, the Q of the tag antenna coil is 40, and the coils are perfectly parallel to each other.
   b. Calculate the induced voltage, assuming that the frequency of the reader was 1 kHz off from the resonant frequency of the tag. What conclusion can you make from this calculation?

$$V_{Tag} = 2\pi fNSQB\cos\theta$$

From here we have:

$$B = \frac{V_{tag}}{2\pi fNSQ\cos\theta}$$

Tag coil size = (85.6 × 54) mm^2 (ISO card size) = 0.0046224 m^2

$$B = \frac{4/\sqrt{2}}{2\pi \cdot 13.56 \cdot 10^6 \cdot 4 \cdot 4.6 \cdot 10^{-3} \cdot 40 \cdot 1}$$
$$B = 0.045\,\mu T$$

For the reader frequency offset, instead of frequency f, we will use the following expression to calculate f_1:

$$f_1 = \frac{f_0}{1+\Delta f} = \frac{13.56MHz}{1+10^{-3}} = 13.546[MHz] \qquad (5.48)$$
$$V_{Tag} = 2\pi fNSQB\cos\theta = 2.83V$$

7. The use of the electromagnetic field for energy scavenging has been considered [18]. Research and calculate how far you have to be from a cellular station in order to achieve successful energy scavenging. Are there any

other areas of urban living offering similar levels of electromagnetic field sufficient for energy scavenging?

8. Although GPS is today quite common in many electronics devices and gadgets, even in combination with RFID, miniaturization of a rice-grain-sized GPS implant is still a very interesting and challenging technological issue. Is a submicron-sized "spychip" implant connected to GPS achievable using the nowadays technology? Discuss the topic.

References

[1] Lehpamer, H., *Microwave Transmission Networks: Planning, Design, and Deployment*, New York: McGraw-Hill, 2004.

[2] Jiang, B., "Energy Scavenging for Inductively Coupled Passive RFID Systems," *IMTC-Instrumentation and Measurement, Technology Conference*, Ottawa, Canada, May 2005.

[3] Karthaus, U., and M. Fischer, "Fully Integrated Passive UHF RFID Transponder IC with 16.7-μW Minimum RF Input Power," *IEEE Journal of Solid-States Circuits*, Vol. 38, No. 10, October 2003.

[4] Microchip, "13.56 MHz RFID System Design Guide," 2004.

[5] Rao, S. K. V., et al., "Antenna Design for UHF RFID Tags: A Review and a Practical Application," *IEEE Transactions on Antennas and Propagation*, Vol. 53, No. 12, December 2005.

[6] Swamy, G., and S. Sarma, "Manufacturing Cost Simulations for Low Cost RFID Systems," White Paper, Auto-ID Center, Massachusetts Institute of Technology, February 2003.

[7] Yang, L., et al., *Design and Development of Novel Inductively Coupled RFID Antennas,* School of ECE, Georgia Institute of Technology, Atlanta, GA, 2006.

[8] Felber, P., "Fractal Antennas," A literature study as a project for ECE 576, Illinois Institute of Technology, December 12, 2000.

[9] Mandelbrot, B., *The Fractral Geometry of Nature*, New York: W. H. Freeman and Company, 1983.

[10] Hall, D., et al., "Turn-On Circuits Based on Standard CMOS Technology for Active RFID Labels," School of Electrical & Electronic Engineering, University of Adelaide, SA, Australia, 2005.

[11] Montauti, F., "High Volume, Low Cost Production of RFID Tags Operating at 900 MHz," White Paper, WaveZero, Inc., June 2006.

[12] Griffin, J. D., et al., "RF Tag Antenna Performance on Various Materials Using Radio Link Budget," *IEEE Antennas and Wireless Propagation Letters*, December 2006.

[13] Cheekiralla, S., and D. W. Engels, *A Functional Taxonomy of Wireless Sensor Network Devices,* Auto ID Laboratory, Massachusetts Institute of Technology, Cambridge, 2005.

[14] Jiang B., et al., "Energy Scavenging for Inductively Coupled Passive RFID Tags," *IMTC 2005, Instrumentation and Measurement Technology Conference*, Ottawa, Canada, May 2005.

[15] Dulman, S., "Data-Centric Architecture for Wireless Sensor Networks," Ph.D. Dissertation, University of Twente, 2005.

[16] Goto, K., et al., "An Implantable Power Supply with an Optically Rechargeable Lithium Battery," *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 7, 2001, pp. 830–833.

[17] http://www.zarlink.com/zarlink/hs/press_releases_15776.htm (accessed August 24, 2010).

[18] Yang, G. Z., *Body Sensor Networks*, New York: Springer-Verlag, 2006.

CHAPTER 6

RFID System Design Considerations

6.1 RFID System Main Considerations

6.1.1 Configuration Design

In practice, determining the number, type, and placement of readers and the manner in which they are connected to other sensors (e.g., motion detectors) and actuators (e.g., conveyor belt speed controls) is part of a large design challenge.

As an example, suppose we wish to use RFID tags to keep track of rare books in a large bookstore; perhaps the most straightforward design is to assign a reader to each bookshelf in order to determine the books in its vicinity. However, the number of readers required by this design, and the implied size of higher-level infrastructure to support the data rate from them, may not be economically feasible.

An alternate design is to assign readers to the points of entry and exit from aisles between bookshelves; in this case, we can infer the current location of a book based on the location of the reader that read its tag most recently. In the former case, tag readers provide *state information* (book x is at location y), whereas in the latter case, readers provide *change-of-state* (event) information (book x just entered aisle z).

This design choice at the lower layers of the architecture would affect the amount and nature of data that must be stored at other layers, as well as the complexity, and therefore cost of the system. In the state-based design, if all past sensor readings for book x are somehow lost (perhaps due to a system malfunction), the book can still be very easily located by simply issuing a query for its EPC. In the event-based design, this option may not be available because the current location of x is out of the range of all sensors [1].

Although hardware configurations (placement of readers, interconnections, and so on) are difficult to change on a frequent basis, the software configurations, which handle how readings are interpreted and routed, can be changed without much effort. This possibility provides the opportunity to rapidly incorporate new business processes into the RFID infrastructure.

During the system design stages, we have to keep in mind some of the basic constraints of the RFID systems and incorporate those in our approach (Figure 6.1).

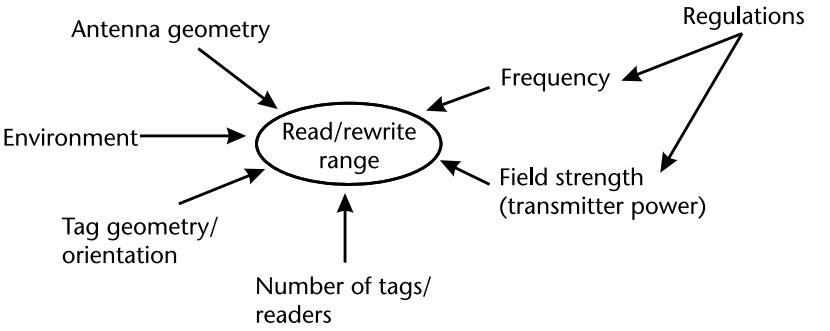


Figure 6.1 Constraints on read/write range.

When designing and inductively coupled RFID system (125 kHz and 13.56 MHz) for optimum read range, we should primarily consider the reader's power, the tag's power consumption, and the tag's quality factor, *Q*, the tag's tuning, the reader's antenna aperture, and the tag's antenna aperture.

Secondary considerations include the tag's modulation depth, the reader's signal-to-noise ratio (SNR), the tag's power-conversion efficiency, the reader's antenna tuning and carrier accuracy, the reader's filter quality, how well the reader's driver matches the antenna, the microcontroller's speed and code efficiency, and the tag's data rate.

Sometimes, the modulation type also affects read range. phase-shift-keying (PSK) and frequency-shift-keying (FSK) systems are inherently more immune to noise than amplitude-shift-keying (ASK) systems, because PSK and FSK systems use a subcarrier that noise cannot easily duplicate. In ASK systems, any sufficiently wide noise spike can look like data and corrupt a bit, so we must use checksums, parity schemes, or cyclic redundancy check (CRC) to counteract the noise. In PSK systems, 0° or 180° phase shift represents a binary bit (1 or 0) during the entire bit time; in FSK systems, two different subcarrier frequencies represent 1 or 0.

However, in a passive system, the tag does not transmit anything, so there is no true subcarrier, only variations of AM. The use of *checksums* or CRCs and the range factors mentioned earlier affect read range so dramatically that any benefits of using FSK or PSK is usually insignificant.

The application environment can also affect read/write range. Key factors include the proximity of the metal to the tag or reader antennas, the presence of in-band noise sources, whether the tag and reader are stationary or moving, and the angle of the tag with regard to the reader's H-field. Another environmental factor is whether the system is enclosed; a system in a shielded tunnel, for example, can use more power than one in the open air. Some of the additional challenges for RFID systems are large population of tags, dynamic tag population, random orientation of tagged objects, and a very high-speed reading.

Tag power consumption, turn-on voltage, and modulation depth vary dramatically from model to model and manufacturer to manufacturer. In addition, chips for different bands typically have very different power requirements; for example, typical 13.56-MHz chip power-up at 4 Vpp and typically draw 7 μA, whereas the 125-kHz chip powers up at 9 Vpp and draws 10 μA.

Power consumption differs widely for systems operating at 13.56 MHz, because CMOS devices consume more current proportionally as their clocking frequency increases. This frequency-dependent consumption is not a problem in synchronous tags operating at 125 kHz; however, a tag that is deriving its clock from a 13.56-MHz carrier has at least one gate that consumes 100 times more current than its counterpart in the 125-kHz tag. The rest of the divider chain draws as much or more than the fastest gate.

As a summary, we can say that the range of passive RFID systems is limited by such factors as tag characteristics, propagation environment, and RFID reader parameters. For example, high-frequency systems have better propagation characteristics, but poorer range in clear air. Pallets, for example, may use UHF tags, but a box of strawberries may need an HF tag.

Typically, reader sensitivity is high, and the tag limitation prevails. Tag range can be maximized by designing a high-gain antenna that is well matched to the chip impedance, but this is a task for electronics circuit design engineers and not system designers and/or integrators.

6.1.2 System Design Checklist

Recognizing opportunities for applying any technology, in this case RFID, is largely a matter of being aware of its capabilities and being able to see how those capabilities relate to your own business operations, processes, services, and products. For an RFID project to be successful, it is necessary to approach the business problem and potential RFID solution using a systems approach.

During the design process, it is required to look at all the processes, plan for the future, and think creatively on how you to improve on each operation. RFID systems should be conceived, designed, and implemented using a systematic development process in which end users and specialists work together to design RFID systems based on the analysis of the business requirements of the organization. Implementing an RFID-based system is like implementing any system; following a checklist will help define requirements:

Systems:

- Why are you implementing RFID?
- Are you being mandated or are you looking at improving your internal operation?
- Is there a requirement or preference for standards?
- Is your market domestic, international, or both?

Tags:

- Do you require disposable tags or are reusable tags acceptable?
- Type of tag required (read-only, R/W, WORM)?
- What is the maximum amount of data to be stored in the tag (data capacity)?
- What data format will be used?

- How and where will the tags be applied?
- What do you do when a tag is read?
- What do you do if a tag is not read?
- What are the tag redundancy requirements?

Reader:

- What is the required read zone (width, height, and depth)?
- How may tags will the reader read or write to at one time?
- What are the possible location(s) for the tag?
- What is the orientation of the tags and distance between tags?
- At what speed and direction will the tags be traveling?
- What error control and correction will be required?
- Do you require any data security?
- What will the required distance be between different reader antennas?
- What is the distance between antenna location and the reader?
- Is portability a requirement?
- What are the data interface and protocol: reader/interrogator (batch, online, wireless, Ethernet) requirements?

Environment:

- What is the proximity of tags and reader antenna proximity to metals, liquids, and so forth?
- What temperature and humidity will the equipment normally be exposed to? What about exposure to chemicals, UV and X-rays, mechanical stress, splash conditions, dust, and so forth?

Business:

- What is an average cost per tag?
- How is RFID implementation going to affect the bottom line?
- What is the return on investment (ROI)?

Note that the larger the coverage area in the environment the greater the implementation challenges. Therefore, the longer reading distance between the tag and the reader, the more noise and interference that has to be contended with.

Although problems with electrical noise are rare, it is a good practice to perform a site survey before commencing antenna design, than struggle to solve a problem later. In general, electrical noise tends to influence the receive performance and results in reduced reading ranges. Slight changes in antenna orientation to the noise source, additional grounding, or shielding can all help to reduce the effects of noise.

Not many of the off-the-shelf interrogators will survive all the extreme conditions to which they may be subjected. Making the choice of a proper interrogator

for the specific environment is critical in reducing the costs involved with replacing frequently damaged equipment and the downtime associated with hardware failure. A thorough environmental study is always recommended, even if the conditions seem to be readily apparent.

6.1.3 Carrier Frequency and Bandwidth

The carrier frequency and channel bandwidth are key considerations in RFID systems for a number of reasons, practical and legislative. Data is carried on a carrier frequency within a regulatory defined channel. The channel is characterized by the carrier frequency and the associated bandwidth or range of spectral allocation to accommodate the frequency spread relating to the data-modulated signal.

The process of modulation invariably generates symmetric so-called sidebands, represented in stylized form as the shaded triangles in the diagram in Figure 6.2.

Depending on the type of modulation, subcarrier components are used or produced as a result of the modulation process. For example, a technique often used for 13.56-MHz, high-frequency, RFID systems, uses a 212-kHz subcarrier to accommodate the baseband coded data, resulting in two subcarrier modulation products, close to the reader carrier frequency but sufficiently distant to allow more effective detection and separation from the reader carrier.

The bandwidth and the associated sensitivity to frequency components within the band, characterized for both tag and reader, are important for a number of reasons. They largely determine the performance of the transfer system, including susceptibility to interference. These quantities also have to be appropriately controlled to meet regulatory requirements, to ensure they do not interfere with other spectrum users.

A number of channels may be specified for use within a regulatory directive. Where this is so, the channels are sufficiently separated to avoid interference but also require protocols to allow access to these channels without contention. Where the bandwidth and sensitivity are specified, the density of readers for realizing coband operation (i.e., the minimum distance between readers) also becomes a consideration.

From a practical standpoint the choice of frequency, together with strength or power of the carrier, has a bearing upon the range of communication that can be

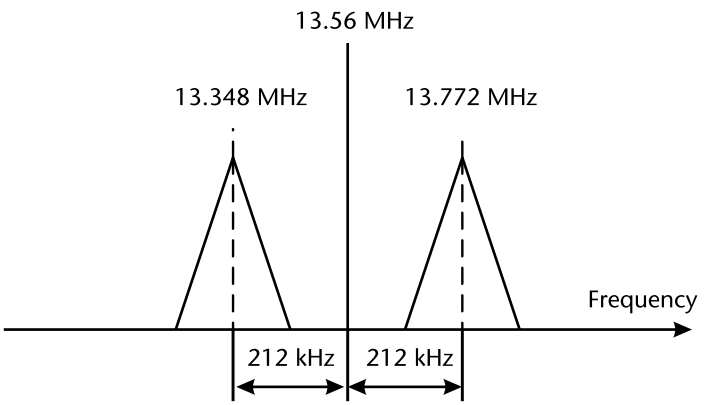


Figure 6.2 13.56-MHz carrier frequency with subcarriers.

achieved between tag and reader. To work, the tag has to receive a signal of sufficient magnitude and the reader must be sufficiently sensitive to pick-up the tag response. Any carrier is subject to a reduction in strength the further it is detected from the source.

Other factors can also influence the magnitude of the signal over distance, including objects and materials in the region between the tag and reader and mechanisms that add noise or interference to the signal being communicated, making it difficult at the receiver end to distinguish the data carrying signal from the noise and interference signals.

6.1.4 Frequency Band Selection

In practice, the region between the tag and interrogator may contain obstacles and materials that can influence the performance of the system (Figure 6.3). The carrier frequency is one of significance with respect to the effects that the prevailing conditions and clutter factors (obstacles and physical structures) can have.

In LF and HF inductive RFID systems, the magnetic field is effectively used to couple data and such fields are largely unaffected by dielectric or insulator materials (papers, plastics masonry, and ceramics, for example), the field simply penetrates the materials.

Metals can distort the field dependent upon how ferrous they are. This will weaken the field strength in regions of the interrogation zone, in some cases to the extent that the system performance is impaired. The range may be impaired or ability to read or write to a tag may be impaired. For uncompensated tag designs operating at resonant frequencies the presence of metals can often detune the device, in some cases preventing it to operate.

At higher frequencies (UHF and above) where, for propagation RFID, the electric component of field becomes more significant the higher the frequency and the more easily that they can penetrate dielectric materials. However, for some materials where energy exchange mechanisms can be identified at or near the carrier

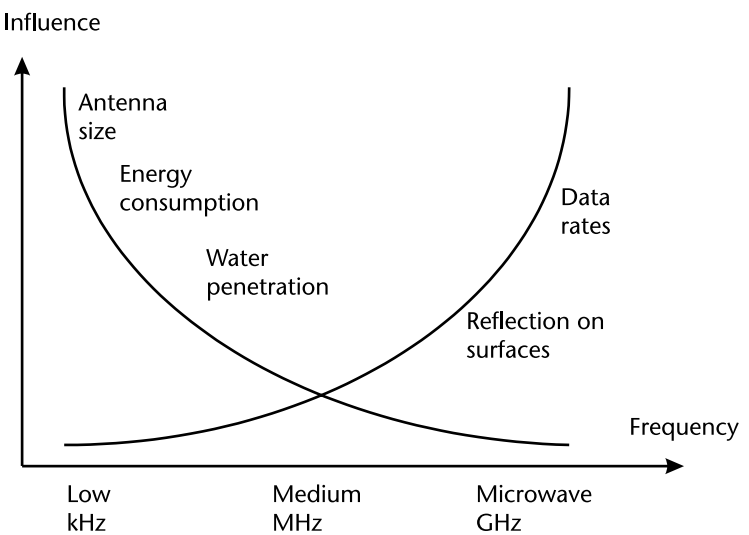


Figure 6.3 Influence of frequency on RFID performance.

frequency, this can result in energy absorption from the propagating wave, so causing an impairment of range performance. Water molecules, for example, can have a significant effect upon microwave transmissions.

As far as metals are concerned, they reflect or scatter these higher-frequency signals depending on the size of the metal object in relation to the wavelength of the incident signal. Such effects can impair the range that can be achieved and in some cases can screen the reader from the tag and prevent it from being read.

Close proximity of tags may also exhibit a similar effect. Reflections and diffraction effects can often allow pathways around metal objects within an interrogation zone. Because it is difficult to generalize on the effects of clutter within the interrogation zone, it is expedient where possible to avoid clutter and choose a frequency band that is appropriate to the expected conditions.

6.1.5 Power and Range

From what has been said so far, the simple method to extending the range would be to increase the power to interrogate the tag and/or the power available within the tag to affect a response. Indeed this can be done, but only within specified and regulated limits.

The extent to which a tag or reader is subject to noise and interference is essentially governed by their respective bandwidth and sensitivity ratings. The greater the sensitivity, the smaller the signals it can respond to, providing they are within the bandwidth of the receiving device. The greater the bandwidth of the receiving device, the greater is the susceptibility to noise and interference. However, mitigation techniques may be used to help improve the performance in avoiding or rejecting unwanted signals.

The sensitivity, together with channel selection, can also have a bearing upon the relative positioning and density of readers. Where readers are operating within the same channel, without any access or anticontention management facilities, the allowable distance between readers is determined by the reader's transmission signal strength and the receiver's sensitivity.

For given power or operational field strength, the greater the receiver sensitivity, the greater the separation has to be between readers. This, in turn, sets the limit on the density of readers that can be accommodated within a particular application environment. To achieve effective functionality where readers are in range of each other, it is necessary for the readers to operate using appropriate mitigation or communication management techniques. These techniques include channel selection to avoid coband coincidence, operational duty cycles, or access management algorithms.

In the United States, FCC regulations limit the amount of power that can be transmitted between 902 and 928 MHz to 30-dBm maximum transmitter power output and a maximum of 36-dBm ERP. A 6-dBi gain antenna (typical for RFID antennas) is added to 30-dBm transmitter power output and yields 36-dBm ERP. Sometimes we can find that reader is actually transmitting 32.5 dBm of power, not 30 dBm as required by the FCC. The typical loss in antenna cables is about 2.5 dB, so starting with 32.5 dBm coming out of the reader, and subtracting a 2.5 dB loss in the cables, means that 30.0 dBm of power arrives at the antenna. Generally speaking, we can say:

$$EIRP_{dBm} = Tx_{dBm} - Transmission\,Line\,Loss_{dB} + Antenna\,Gain_{dBi} \quad (6.1)$$

It is possible in the United States to use a higher-gain antenna, such as an 8-dBi antenna, as long as one reduces the transmitter power output by 2 dBm so that the ERP stays under the 36-dBm ERP limit. Generally, an 8-dBi gain antenna has a narrower beamwidth than a 6 dBi antenna, so doing this may be useful in specific situations where one wants a longer but narrower read field.

In general, users tend to want the largest possible read field, and, given FCC constraints, that is accomplished with a 6-dBi antenna. It is not a good idea to change the power settings, cabling, or antenna that come with your reader, because this may violate FCC or other local regulations. Check your reader's documentation or ask the manufacturer about changing power settings, cabling, and antennas that comply with relevant regulations.

6.1.6 Link Budget

Due to the indirect power supply it is essential to make a careful calculation of the power budget. Here, we are going to use a simplified approach and calculation using decibels (more detailed formulae and their derivations are presented in Chapter 5).

Contributions from the transmitter and the reader antenna are relatively easy to evaluate; however, in the operational environment, the transmission setup and antenna are subject to strong variations due to the strongly variable environment. Modern systems are targeted for communication up to a few hundred tags, thus requiring good reliability (i.e., safety margins have to be included). In modeling a real situation with a reader, many tags and other objects in between is an important part, of the RFID system design, and the use of RF simulation tools in all parts of the system helps to predict the range of reliable operation. Typical UHF operating parameters are as follows:

- Reader transmit power $P_R = 33$ dBm (2W);
- System operating wavelength $\lambda = 0.33$m (915 MHz);
- Reader receiver sensitivity $S_R = -80$ dBm (10^{-11} W);
- Reader antenna gain $G_R = 4$ (6 dBi);
- Tag power (sensitivity) requirement $P_T = -14$ dBm (40 μW);
- Tag antenna gain $G_T = 1.26$ (1 dBi);
- Tag backscatter efficiency $E_T = 0.01$ or 1% (–20 dB) calculated as $(\Delta\rho)^2$.

The differential reflection coefficient, $\Delta\rho$, is described in Chapter 5 in more detail.

6.1.6.1 Forward Link Budget

The signal received at the tag (Figure 6.4) has to be bigger or, in the worst case, equal to the tag sensitivity threshold. In case the tag is at the far edge of the interrogation zone, we can say that:

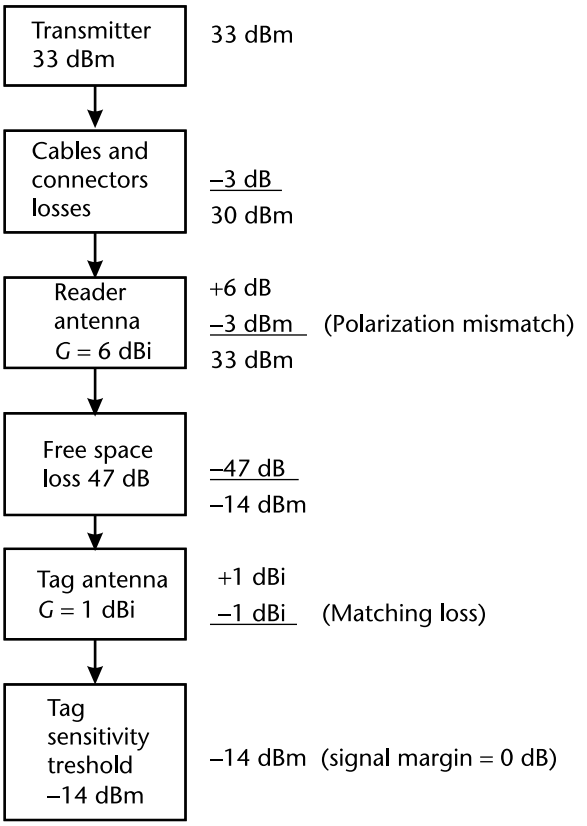


Figure 6.4 UHF RFID forward link budget.

$$\begin{aligned} P_{Tag} &= P_R + G_R + G_T - \Sigma_{Losses} - FSL \\ FSL &= P_R + G_R + G_T - \Sigma_{Losses} - P_{Tag} \\ FSL &= (33+6+1) - (3+3+1) - (-14) = 47\,dB \end{aligned} \tag{6.2}$$

$$r = 10^{\frac{FSL-31.75}{20}} = 10^{\frac{47-31.75}{20}} \approx 5.8m \tag{6.3}$$

The maximum free-space loss (FSL) allowed for this case is 47 dB. From the Friis formula for FSL, we can calculate the maximum (in ideal cases) distance at 915 MHz to be 5.8m (or approximately 19 feet).

6.1.6.2 Reverse Link Budget

The backscatter communication radio link budget, describes the amount of modulated power that is scattered from the RF tag to the reader. The antenna gains include losses due to mismatch.

$$P_{REC} = \frac{P_R G_R^2 G_T^2 \lambda^4 E_T}{(4\pi)^4 r^4}$$

$$P_{REC} = \frac{10^3 (2)^2 (1)^2 (0.33)^4 (0.01)}{(4\pi)^4 (5.8)^4} = 0.0000168\,\mu W \tag{6.4}$$

$$P_{REC} \approx -78\,dBm$$

Using decibels (Figure 6.5), we can write:

$$P_{REC} = P_{Tag} + G_R + G_T - \Sigma_{Losses} - FSL - E_T$$

$$P_{REC} = -14 + 6 + 1 - (3 + 1) - 47 - 20$$

$$P_{REC} = -78\,dBm \text{ (the same result as before)}$$

Thus, because $S_R = -80$ dBm and $P_{REC} > S_R$, we still have around a 2-dB signal margin at the reader's receiver. The question is whether the UHF read range is tag sensitivity limited or reader sensitivity limited. Well-designed passive systems are always limited by the tag's sensitivity.

In practice, the maximum theoretical activation range is decreased by four types of additional losses:

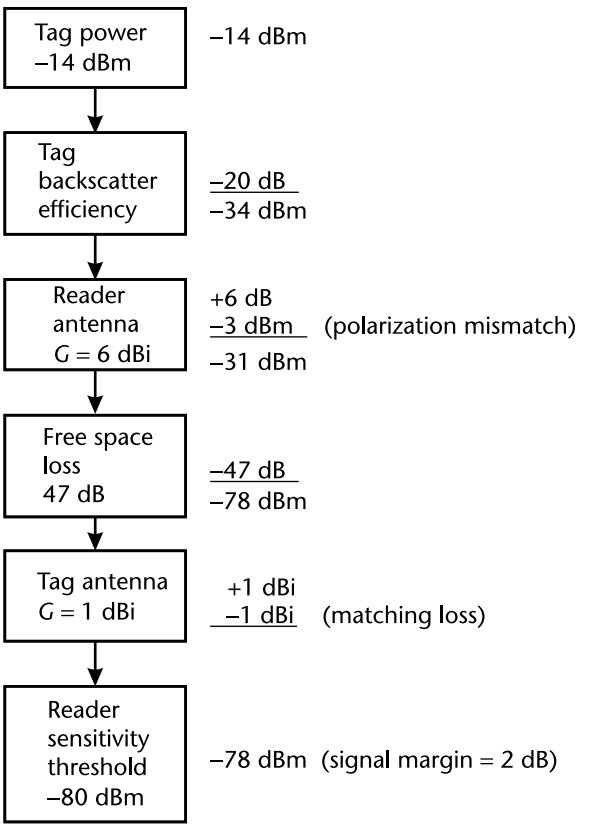


Figure 6.5 UHF RFID reverse link budget.

- *Absorption:* Because most RFID systems are deployed indoors and there is not always a line-of-sight path between tag and reader, the free-space assumption is usually not valid. The electromagnetic wave supplying tags with power is, for example, completely reflected by perfect conductors and partially reflected by perfect dielectrics. Real-world lossy dielectrics between the reader antenna and the tag antenna will also absorb some of the incident radiation, resulting in the read range significantly shorter than predicted theoretically.
- *Multipath fading:* Even if there is line of sight between reader antenna and tag, small-scale fading effects can increase and decrease the read range. Multipath fading is caused by interference between two or more versions of the transmitted reader signal, which arrive at the receiver at slightly different times. These multipath waves combine at the receiver to result in a signal that can vary widely in amplitude and phase. Due to the constructive and destructive effects of multipath waves, a tag moving past a reader antenna can pass through several fades in a small period of time. If the tag passes through such a field null, it will lose power and possibly also its state.
- *Polarization losses:* The activation range is further significantly reduced by polarization losses, since the precise orientation of tags relative to the reader antenna is usually not known. Even when the reader is transmitting with a circularly polarized antenna, the transponder fails to be adequately powered when the axis of the tag dipole antenna is aligned with the propagation direction of the emitted electromagnetic wave. Circularly polarized antennas also introduce an additional loss of 3 dB. A promising approach to alleviate this orientation dependence is the use of two tag antennas that are orthogonally polarized and attached to the same microchip.
- *Impedance mismatch:* The activation range predicted by the Friis transmission equation could, in practice, be further reduced by impedance mismatch between tag antenna and microchip. In most calculations, this fact is neglected and perfect match is assumed.

The *EIRP* determines the power of the signal transmitted by the reader in the direction of the tag. The maximum allowed EIRP is limited by national regulations.

Chip sensitivity threshold is the most important tag limitation. It is the minimum received RF power necessary to turn on an RFID chip [2]. The lower the required RF power, the longer the distance at which the tag can be detected. Chip sensitivity is primarily determined by RF front end architecture and fabrication process; RFID chips may also have several RF inputs connected to different antenna ports. Antenna gain is another important limitation; tag range is highest in the direction of maximum gain that is fundamentally limited by the frequency of operation and the tag size.

Reader sensitivity is another important parameter that defines the minimum level of the tag signal that the reader can detect and resolve. The sensitivity is usually defined with respect to a certain SNR or error probability at the receiver. Factors that can affect reader sensitivity include receiver implementation details,

communication protocol specifics, and interference, including signals from other readers and tags.

Tag detuning is due to the fact that antenna characteristics change when the tag is placed on different objects or when other objects are present in the vicinity of the tag. Tag detuning degrades antenna gain (due to changes in the radiation pattern) and impedance match, thus affecting the tag range.

Figure 6.6 shows that, although based on calculations and even RF measurements, correct tag readings should be achieved without problem, at increased distances they will become increasingly unreliable. After a certain distance, the number of correct readings of the 60-tag pallet will decrease sharply with increasing distance.

6.1.7 Collision Avoidance

A major problem with RFID systems is that a tag might not be read, in spite of being in the reader's range, due to collisions. A collision is said to have occurred when various devices interfere with each other's operations or their simultaneous operations lead to loss of data.

The reading process is not efficient due to various types of collisions, which are classified as follows:

- *Single reader-multiple tags collision:* Multiple tags are present to communicate with the reader. They respond simultaneously and reader is not able to interpret the signal.
- *Single tag-multiple readers collision:* Single tag is in the range of two or more readers. Tags are mainly passive entities, so they do not have enough power to differentiate between frequency ranges of the readers. Tag interference is more common among active tags, in that they have a greater range and are more likely to interact with multiple readers at a given instant in time. When this problem exists in isolation, it is said to be a *resource-constrained scheduling problem* and solved using optimization methods.

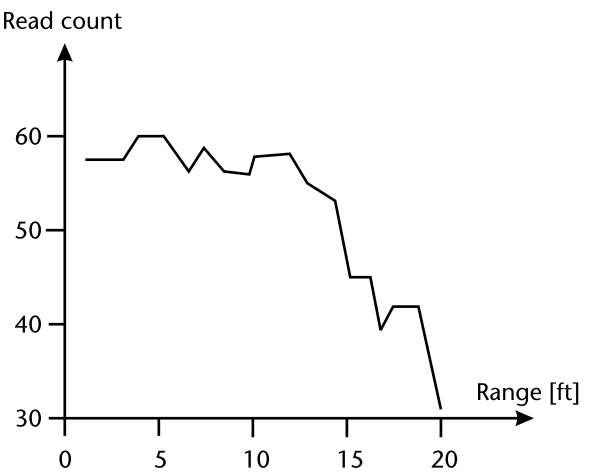


Figure 6.6 The tag sensitivity limitation in practice.

- *Reader-reader collision:* Two or more readers within the same frequency range interfere with each other's operations.

These problems need to be resolved to provide efficient solutions for tag identification, and these are the major research areas where work needs to be done to practically implement RFID systems. Several metrics may be used to judge the quality of anticollision algorithms: performance, range, bandwidth requirements, implementation costs, noise and error tolerance, and security.

6.1.7.1 Tag-Tag Collision

In many existing applications, a single-read RFID tag is sufficient; animal tagging and access control are examples. However, in a growing number of new applications, the simultaneous reading of several tags in the same RF field is absolutely critical; library books, airline baggage, garments, and retail applications are a few. In order to read multiple tags simultaneously, the tag and reader must be designed to detect the condition that more than one tag is active. Otherwise, the tags will all backscatter the carrier at the same time, and the amplitude-modulated waveforms would be garbled. This is referred to as a collision, and no data would transfer to the reader.

With several entities communicating on a same channel, it is necessary to define some rules to avoid collisions and therefore to avoid information loss. The required rules are known as the *collision avoidance protocol*. The tags' computational power is very limited and they are unable to communicate with each other. Therefore, the readers must deal with the collision avoidance themselves, without the help from the tags. Usually, readers query the tags until all identifiers are obtained. The process of addressing and isolating a single tag is referred to as *singulation*. We say that the reader performs the *singulation* of the tags because it can then request them selectively, without collision, by indicating the identifier of the queried tag in its request [3].

The collision avoidance protocols that are used in the current RFID systems are often proprietary (i.e., not open-source) algorithms. Therefore, obtaining information on them is difficult. Currently, several open standards appear, and they are used more instead of proprietary solutions.

We distinguish the EPC family from the ISO family; regardless of whether they are EPC or ISO, there are several collision avoidance protocols. Choosing one of them depends, in part, on the frequency used. EPC proposes standards for the most used frequency, that is, 13.56 MHz and 860 to 930 MHz. ISO proposes standards from 18000-1 to 18000-6, where 18000-3 corresponds to the frequency 13.56 MHz and 18000-6 corresponds to the 860–960-MHz frequency.

There are two main classes of collision avoidance protocols: the *deterministic protocols* and the *probabilistic protocols*. Usually, we use the probabilistic protocols for systems that use 13.56-MHz frequency (the U.S. regulations in this band offer significantly less bandwidth), and the deterministic protocols for systems using the 860–960-MHz frequency because they are more efficient in this case.

The *deterministic protocols* rely on the fact that each tag has a unique identifier. If we want the singulation process to succeed, the identifiers must stay unchanged until the end of the process. In the current tags, the identifiers are set by

the manufacturer of the tag and written in the tag's ROM. In normal RFID systems, there is no exchange after the singulation because the reader has obtained the expected information, that is, the identifiers of the tags which are in its field.

The *probabilistic protocols* are usually based on a time-division multiple access (TDMA) protocol, called ALOHA. The ALOHA[1] protocol is a simple protocol originally developed for use in radio communication systems, but can be applied in every system where uncoordinated information is send over the same channel. The original protocol has two rules:

- Whenever you have something to send, send it.
- If there is a collision when transmitting (i.e., another entity is trying to send at the same time), try to resend later. This also applies in case of transmission failure.

In the tag-reader context, tags avoid collisions with other tags by randomly delaying their responses. If a collision does occur, the reader will inform all nearby tags and the culprits will wait another, usually longer, random interval before continuing. Higher densities of tags will result in a higher collision rate and degraded performance. The ISO 15693 standard for RFID supports a slotted ALOHA mode of anticollision.

Slotted ALOHA is a more advanced, but still simple, protocol, where the receiving entity sends out a signal (called a beacon) at equally spaced intervals, thus dividing time into slots, and requiring synchronization. The beacon announces the start of a new slot and thereby the time to start sending the next packet for any entity having one ready.

Several researchers have examined the issue of collision in RFID-tagged systems. The approaches taken in existing literature in this area fall into two broad categories: tree-based algorithms and variants of the ALOHA protocol [4]. The version of slotted ALOHA applied in RFID collects a number of consecutive slots into groups. At the beginning of each group the reader announces that only transponders with IDs starting with a specified substring are to answer now. Each tag thus activated picks a random number and waits for that many slots before transmitting.

In general, the number of slots is chosen randomly by the reader, which informs the tags that they will have n slots to answer to its singulation request. Each tag randomly chooses one slot among the n and responds to the reader when its slot arrives. If n is not sufficiently large with regard to the number of tags which are present, then some collisions occur. In order to recover the missing information, the reader interrogates the tags one more time. It can mute the tags that have not brought out collisions (switched-off technique) by indicating their identifiers or the time slots during which they transmitted. Also, according to the number of collisions, it can choose a more appropriate n.

1. ALOHA is a random (or contention) access protocol developed at the University of Hawaii for sharing broadcast channel access among a number of users with relatively low-throughput demand. ALOHA protocols are often used in satellite communications systems and cellular radio systems and are a precursor to the popular Ethernet protocol.

Tag-tag collision occurs when multiple tags respond to the same reader simultaneously and due to multiple signals arriving at the same time, the reader may not be able to detect any tag. This problem prevents the reader from detecting all tags in its interrogation zone. A simple deterministic algorithm used to solve this problem is the *tree-walking algorithm* (TWA), which is generally used in UHF readers (Figure 6.7).

In this protocol, the reader splits the entire ID space into two subsets and tries to identify the tags belonging to one of the subsets, recursing along the way until a subset has exactly one tag or no tags at all. To describe how a tree walk is performed, a simple example is given in which the transponders IDs only consist of 3 bits. Three transponders with the IDs 001, 011, and 110 are introduced into the reader's scanning area.

The reader first asks if any transponders have a 0 as the first bit. The 110 transponder does not and goes into a sleep state, while the other two transponders answer. The reader then asks if any transponders have a 0 as the second bit. Again, this is confirmed by the 001 transponder, but the 011 transponder goes into a sleep state. Then the reader asks for transponders with a 0 as the third bit. Nobody answers, and 001 goes into the sleep state. Because nobody answers, the reader backs up one step and asks all transponders that confirmed their presence at the second bit to wake up. This reactivates 001. The reader now asks for transponders with a 1 as the third bit; 001 answers and is now fully identified. By continuing this back-up-one-step and forward-one-step sequence a number of times, all three transponders are identified.

Due to larger turnaround times at lower frequencies, TWA is not deemed suitable for HF readers. Instead, the HF readers use a Slotted Termination Adaptive Collection (STAC) protocol somewhat similar to the framed ALOHA protocol.

The binary tree walking anticollision algorithm discussed here has an inherent security problem due to the asymmetry between forward and backward channel strengths. Every bit of every singulated tag is broadcast by the reader on the forward channel. At certain operating frequencies, a long-range eavesdropper could monitor these transmissions from a range of up to 300 feet (100m) and recover the contents of every tag. A variant of binary tree walking, which does not broadcast insecure tag IDs on the forward channel and does not adversely affect performance, originally appeared under the name *silent tree-walking*.

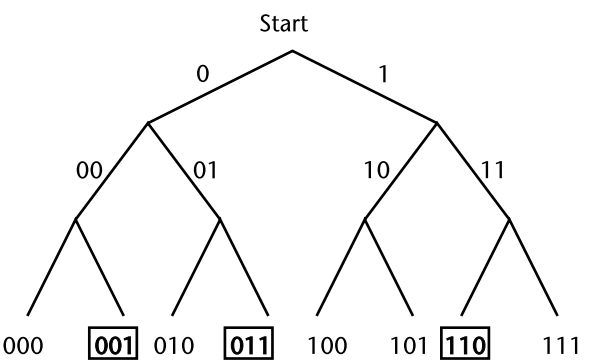


Figure 6.7 The tree-walking algorithm.

Assume that a population of tags share some common ID prefix, such as a product code or manufacturer ID. To singulate tags, the reader requests all tags to broadcast their next bit. If there is no collision, then all tags share the same value in that bit. A long-range eavesdropper can only monitor the forward channel and will not hear the tag response. Thus, the reader and the tags effectively share a secret bit value. When a collision does occur, the reader needs to specify which portion of the tag population should proceed. If no collisions occur, the reader may simply ask for the next bit, since all tags share the same value for the previous bit.

Since we assumed the tags shared some common prefix, the reader may obtain it as a shared secret on the uplink channel [5]. The shared secret prefix may be used to conceal the value of the unique portion of the IDs. Suppose we have two tags with ID values b_1b_2 and $b_1\overline{b_2}$. The reader will receive b_1 from both tags without a collision and then will detect a collision on the next bit. Since b_1 is hidden from long-range eavesdroppers, the reader may send either $b_1 \oplus b_2$ or $b_1 \oplus \overline{b_2}$ to singulate the desired tag without revealing either bit. Eavesdroppers within the range of the backward channel will obviously obtain the entire ID. However, this blinded tree-walking scheme does effectively protect against long-range eavesdropping of the forward channel with little added complexity. Performance is identical to regular tree-walking, since a tag will be singulated when it has broadcast its entire ID on the backward channel.

A number of other variants of the same idea have been described in the literature.

6.1.7.2 Reader-Tag Collision

Reader-tag collision occurs when the signal from a neighboring reader interferes with tag responses being received at another reader. This problem has been studied in the EPCglobal Class1 Gen1 and Gen2 standards for UHF readers. In Gen 1 standard, the reader-tag collision problem is mitigated by allowing frequency hopping in the UHF band or by TDMA. In Gen 2 the readers and tags operate on different frequencies so that the tag response does not interfere or collide with reader signals. Either solution requires fairly sophisticated technology [6].

6.1.7.3 Reader-Reader Collision

Large-scale RFID deployments in future will most likely involve multiple readers due to the fact that each RFID reader has a limited interrogation range. The reader can communicate with any tag within its interrogation range. The size and shape of the interrogation range of a particular reader are determined by many factors, including antenna characteristics, radio transmit power, radio obstructions, and wireless interferences. It also depends on characteristics of the tag.

In many applications such as warehousing or manufacturing, a large area must be perfectly covered. This motivates the use of multiple RFID readers geographically dispersed and networked in some fashion (in an ad hoc network, for example) and performing tag reading concurrently. Use of multiple readers not only improves coverage, but also improves read throughput by virtue of concurrent operation. However, several collision problems might occur when multiple readers are used in close proximity to each other.

Colorwave is a distributed algorithm based on TDMA and one of the first works to address reader-reader collisions [7]. In particular, it considers an interference graph over the readers, wherein there is an edge between two readers if they could lead to a reader-reader collision when transmitting simultaneously. It then tries to randomly color the readers such that each pair of interfering readers has different colors. If each color represents a time slot, then the above coloring should eliminate reader-reader collisions. If conflicts arise (i.e., two interfering readers pick the same color or time slot), only one of them wins (i.e., sticks to the chosen color); the others pick another color again randomly.

Some authors suggest coloring of the interference graph using k colors, where k is the number of available channels. If the graph is not k-colorable using their suggested heuristic, then the authors suggest removal of certain edges and nodes from the interference graph using other methods that consider the size of the common interference regions between neighboring readers.

A query is said to be successfully sent if it is sent by a reader and is successfully received by all the tags in the read range, that is, if it does not collide with any other query in the network. Hence, if the reader does not receive any offline message for a query, the query is considered as being sent successfully. We define the system throughput as follows:

$$System\,Throughput = \frac{Total\,Successful\,Query(All\,Readers)}{Total\,Time} \tag{6.5}$$

In general, the tag identification is through a query-response protocol where the reader sends a query and the tag responds with its unique identification number. The higher the number of queries sent successfully, the higher the throughput and, hence, the higher would be the number of tags identified by the readers. Thus, throughput and efficiency together define the effectiveness of the anticollision protocol.

Engels et al. presented two algorithms called Distributed Color Selection (DCS) and Variable-maximum Distributed Color (VDCS) [8].

6.1.8 Tag Reading Reliability

Ghost reads occur when an RFID reader gathers information from a noisy environment and reports on a tag that does not exist. Reporting of phantom tags consumes processing and network time that may impact system reliability and performance.

Statistically, there is always a chance of a ghost read; however, Gen 2 was specifically designed to address and minimize the occurrence of ghost reads. Statistics reveal that roughly one ghost read occurs per 1,000 tags. Gen 2 virtually eliminates the potential for these phantom reads, even in a noisy environment, by providing five sequential mechanisms that serve as checks for a tag's validity. Only tags that pass all of these five tests are designated as valid tag reads and entered into the system:

1. *Tag response time:* Tags must respond to a reader within a very short, defined time frame. If the tag response is not timed exactly from the beginning

of a response to the end of the response, the probability that the tag is a phantom is high, so the reader will ignore the tag. The reader may try to reread the tag at a later time, possibly under different conditions.

2. *Preamble:* For each and every response, tags first send a signal called a preamble. When a reader receives a valid preamble, it knows the signal is valid (from a real tag), and not simply noise. If the preamble is not valid, the reader discontinues communication and moves on to the next tag signal.
3. *EPC format check:* If the preamble is validated, the reader then examines the bit stream transmission to ensure it is in a valid EPC format. If the EPC format is validated, communication between tag and reader continues. If the EPC format is not validated, the reader begins communication with another tag.
4. *Bit match:* The reader compares the number of bits the Gen 2 tag reported that it would be sending to the number of bits received; if it is not a match, the information is discarded.
5. *CRC:* The CRC checks for bit errors in transmission by comparing the number of bits the tag stated it would send with the number of bits actually sent. If the correct number of bits was received, the transmission is verified as accurate. If the correct number of bits is not received, the data is rejected and the reader moves on to begin communication with the next tag.

6.2 RFID Reader-Tag Communication Channel

Irrespective of the mode of coupling (inductive or propagation), the means of effectively transferring data between tag and reader relies upon a dialogue between the two, based upon command data within the reader and the tag response signals, the dialogue generally being initiated by the reader (reader talk first).

However, in some RFID systems the tag may operate in a beacon type mode (active tags) and effectively talk first. Suffice is to say at this stage that the objective in either case is to transfer data, so the dialogue essentially requires the tag to be identified and the data to be requested, acknowledged, and sent. To achieve this, it is necessary for data to be written or encoded into a tag in a particular way:

1. With other data elements added to facilitate identification, description of the source data (the data required to be carried and used at the receiver end of communications) components (so-called *metadata*), and, as appropriate, elements used for error detection and correction, contention management, and communications dialogue are included. This is generally known as *source encoding*.
2. The source-encoded message is structured by certain baseband techniques to better match the signal form of the message to be sent to the characteristics of the transmission channel or medium through which it is to be transmitted. This is process is often referred to as *channel encoding*.

The data communicated between tags and readers must be sent in a reliable manner. With data encoded in this way, the final conditioning that is used to facilitate transmission is the modulation of the encoded data using a suitable

frequency-defined carrier signal. At the receiver end the reverse of these processing elements are performed (demodulation, together with channel and source decoding) to recover the source data. A variety of techniques are used for channel encoding and modulation distinguished by particular performance and cost characteristics.

The combination of coding and modulation schemes determines the bandwidth, integrity, and tag power consumption. The coding and modulation used in RFID communications is limited by the power and modulation/demodulation capabilities of the tags.

Another limiting factor is the bandwidth occupied by the signal; RFID tags that are passive do not transmit signals actively and therefore can use more bandwidth than a reader. A reader has its own power source and therefore is required by regulations to use less bandwidth.

6.2.1 Data Content and Encoding

Line coding involves converting a sequence of 1s and 0s to a time-domain signal (a sequence of pulses) suitable for transmission over a channel. The following primary factors should be considered when choosing or designing a line code:

1. *Self-synchronization*: Timing information should be built into the time-domain signal so that the timing information can be extracted for clock synchronization. A long string of consecutive 1s and 0s should not cause a problem in clock recovery.
2. *Transmission power and bandwidth efficiency*: The transmitted power should be as small as possible, and the transmission bandwidth needs to be sufficiently small compared to the channel bandwidth so that intersymbol interference will not be a problem.
3. *Favorable power spectral density*: The spectrum of the time-domain signal should be suitable for the transmission channel. For example, if a channel is ac coupled, it is desirable to have zero power spectral density near dc to avoid dc wandering in the pulse stream.
4. *Low probability of error*: When the received signal is corrupted by noise, the receiver can easily recover the uncoded signal with low error probability.
5. *Error detection and correction capability*: The line code should have error detection capability, and preferably have error correction capability.
6. *Transparency*: It should be possible to transmit every signal sequence correctly regardless of the patterns of 1s and 0s. If the data are coded so that the coded signal is received correctly, the code is transparent.

There are two broad categories of codes used in RFID: level codes and transition codes. *Level codes* represent the bit with their voltage level. *Transition codes* capture the bit as a change in level. Level codes, such as nonreturn-to-zero (NRZ) and return-to-zero (RZ), tend to be history independent; however, they are not very robust. Transition codes can be history-dependent, and they can be robust.

Transmitter (tag) is responsible for encoding, that is, inserting clocks into the data stream according to a select coding scheme while the receiver (reader) is responsible for decoding (separating) clocks and data from the incoming embedded data stream.

Readers are capable of transmitting at high power, but are limited to narrow communication bands by communications regulations; therefore, the encoding used from reader to tag usually needs to occupy a low bandwidth. Passive tags, however, do not actively transmit a signal; therefore, the encoding used for tag to reader communication can occupy a high bandwidth. There are many types of line codes, but we will only discuss a few of them that are important for our RFID discussion (see Figure 6.8). More general information about data encoding can be found in [9].

6.2.1.1 Nonreturn-to-Zero Coding

The simplest channel code is the one known as nonreturn-to-zero (NRZ). Simple, combinational logic, signals are a good example of NRZ, where a logical 1 is coded as one DC level and logic 0 as another. NRZ requires time coordination, but longs strings of 0s and 1s do not produce any transitions that may create problems in error detection and recovery. NRZ produces a high dc level (an average of 0.5V). NRZ is rarely used for serial transmission, except for relatively low-speed operations such as those associated with modem transfers.

Serial transmission in wired systems generally consists of at least two transmission lines: one carrying the data, and the other the clock to which the data is synchronized. Wireless transmission, however, is an entirely different situation; wireless data has only one medium to travel through, the air, and, as such, cannot support separate transmission of data and clock. Therefore, traditional NRZ data, in which a logic one is a high signal for one clock period and a logic zero is a low signal for one clock period, cannot be used.

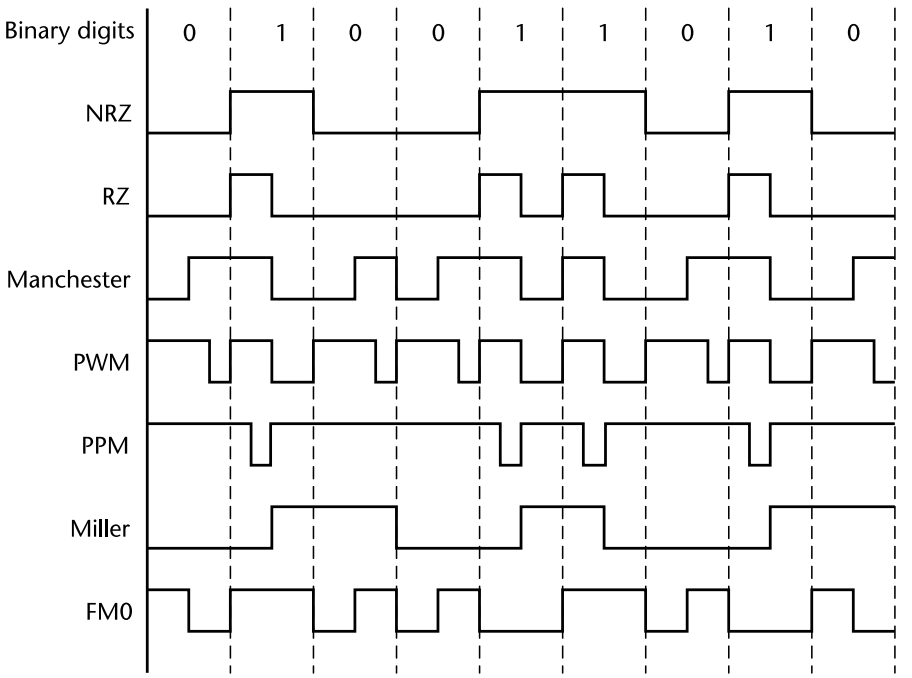


Figure 6.8 Examples of several coding schemes.

A data stream, in general, can contain long strings of ones or zeros; these would be represented in an NRZ stream by very long dc values, during which the receiving system may lose synchronization with the transmitter's clock. To combat this, the Manchester code, for example, can be used.

6.2.1.2 Return to Zero (RZ)

Return-to-zero (RZ) describes a line code used in telecommunications signals in which the signal drops (returns) to 0 between each pulse. This takes place even if a number of consecutive 0s or 1s occur in the signal. This means that a separate clock does not need to be sent alongside the signal, but suffers from using twice the bandwidth (data rate) as compared to NRZ format. The signal is self-clocking.

Although RZ contains a provision for synchronization, it still has a dc component resulting in *baseline wander* during long strings of 0 or 1 bits, just like the NRZ line code. A variant of RZ is RZ-inverted, which swaps the signal values for 1 and 0.

6.2.1.3 Biphase Mark (Manchester) Coding

Manchester coding incorporates a transition in the middle of every transmitted bit. Logic 1 is represented by a transition from low to high, and logic 0 is represented by a transition from high to low. Because the transmitted signal must change at least once for every bit transmitted, the problem of transmitting long dc values is eliminated. The Manchester code provides for efficient communication since the bit rate is equal to the bandwidth of the communication. Biphased data streams have generally a signal change in the middle of each bit, independent of the value. Therefore, the signal does not necessarily return to zero. The characteristics of the biphase method are:

- *Synchronization*: Because the transition for each bit is predictable, the receiver can synchronize on this edge. These codes are also known as self-clocking.
- *Error immunity*: To cause an error, the noise must invert both, the signal before and after the transition.
- A 1-to-0 transition represents a 0 bit.
- A 0-to-1 transition represents a 1 bit.
- The mid-bit transition is used as clock as well as data.
- The residual dc value is eliminated by having both polarities for every bit.
- The bandwidth required could be twice the bit rate (efficiency of this code can be as low as 50%).

Unlike NRZ data with a separate clock signal, the clock is not provided explicitly to the device receiving Manchester encoded data; instead, it is given, encoded, in the transmitted data so the clock must be recovered from the data. Traditionally, this is accomplished via a Phase-locked Loop (PLL) that takes in the received data stream and outputs the transmitter's clock.

6.2.1.4 FM0 Coding

EPCglobal Class 1 Gen 2 provides multiple options for tag coding and the simplest approach is FM0 coding. In FM0 coding (biphase space), a transition has to occur at the end of each bit period, but for a 0 bit, an addition transition in the middle is required. The duty cycle of a 00 or 11 sequence, measured at the modulator output, shall be a minimum of 45% and a maximum of 55%, with a nominal value of 50%.

FM0 encoding has memory; consequently, the choice of FM0 sequences depends on prior transmissions. FM0 signaling always ends with a *dummy data,* which is a 1 bit at the end of a transmission.

6.2.1.5 Miller Coding

Miller coding is also called a *delay modulation* and provides a transition for every bit. In this code a 1 is encoded as a transition occurring at the center of the bit cell, while consecutive 0s have a transition at the cell boundary between them. This means that a pattern such as 10101 has no transitions at the cell boundaries.

- There is a transition in the middle of a bit period, if it is a 1 bit.
- There is a transition at the start of the bit period if the bit 0 is followed by a 0 bit.
- For a 0 followed by a 1 or a 1 followed by a 0, no transition occurs at the symbol interval.

This code is very efficient in terms of the desired bandwidth (half of the desired bandwidth of Manchester coding). Miller coding is self-clocking and has a relatively low LF content but is not dc-free, however.

Miller squared coding (so called because it was the result of a modification of Miller coding by a second, quite separate Miller) has one additional rule. This states that the final transition of an even number of 1s occurring between two 0s is omitted (i.e., 01110 occupies five cells and has three transitions, while 011110 occupies six cells but also has three transitions), thus making the code dc-free.

6.2.2 Modulation

The data coding scheme determines how the data is represented in a continuous stream of bits. How that stream of bits is communicated between the tag and the reader is determined by the modulation scheme. Each RFID standard employs one modulation scheme for the forward link (reader-to-tag) and another for the reverse link (tag-to-reader).

The modulation schemes reflect the different roles of reader and tag. The reader must send enough RF power to keep the tag powered. A passive tag does not transmit its own signals but modulates by changing the phase or amplitude of the reader's transmitted signal that is being backscattered (in UHF system) from its antenna.

For convenience, RF communications typically modulate a high-frequency carrier signal to transmit the baseband code. RFID systems usually employ modulation techniques and coding schemes that are simple to produce. The three classes

of digital modulation are ASK, FSK, and PSK. The choice of modulation is based on power consumption, reliability requirements, and bandwidth requirements. All three forms of modulation may be used in the return signal, although ASK is most common in load modulation at 13.56 MHz, and PSK is most common in backscatter modulation.

The choice is essentially determined by performance requirements and cost. For example, ISO 18000 Type C (also known as EPC Gen2, Class 1) calls for double sideband-ASK (DSB-ASK), single sideband-ASK (SSB-ASK), and phase-reversal-ASK (PR-ASK). ASK digital modulations are spectrally inefficient, requiring substantial RF bandwidth for a given data rate. Bandwidth efficiencies of 0.20 bit per hertz of RF bandwidth are not uncommon for DSB-ASK.

It is possible to improve bandwidth efficiency using SSB-ASK. This is particularly important in European countries where bandwidth restrictions may preclude DSB-ASK. The power efficiency of DSB-ASK and SSB-ASK is dependent on the modulation index. With a modulation index of 1, or on-off keying (OOK) of the carrier, the lowest carrier-to-noise (C/N) required to achieve a given bit error rate (BER) is obtained for DSB-ASK and SSB-ASK. Unfortunately, this also provides the least amount of RF power transport on the downlink to supply the tag with energy. Ideally, the off time of the carrier should be minimized so that the tag does not run out of power. The C/N requirements should also be minimized to maximize the ID read range. For many modulations, these are conflicting requirements.

A modulation that can minimize the C/N requirement in a narrowband, while maximizing the power transport to the tag, is PR-ASK. Similar to a PSK signal, PR-ASK changes phase 180° each time a symbol is sent. PR-ASK also creates an amplitude modulation depth of 100% or a modulation index of 1, as the phase vector of the old symbol and the new symbol cross and briefly sum to a zero magnitude. This provides an easily detected clock signal as the amplitude briefly goes to zero, but minimizes the time the carrier power is off, so power transport to the passive tag is optimized. PR-ASK has carrier to noise and bandwidth requirements that more closely match PSK than DSB-ASK, making it attractive for narrowband and longer-range applications.

DSB-ASK is the least bandwidth efficient modulation, but the easiest to produce by OOK of the carrier signal. ASK modulation specifications often have a modulation depth as well as rise and fall time requirements. The rise and fall time is typically related to the bandwidth filtering while the modulation depth is set by the attenuation difference between the keying states.

Passive RFID tags operate in a way that may seem unusual because there is only one transmitter; the passive tag is not a transmitter or transponder in the purest definition of the term, yet bidirectional communication is taking place. The RF field generated by the reader (the energy transmitter) has three purposes:

1. Induce enough power into the tag coil to energize the tag. Passive tags have no battery or other power source, so they must derive all power for operation from the reader field. The 125-kHz and 13.56-MHz tag designs must operate over a wide dynamic range of carrier input, from the very near field (in the range of 200 V_{PP}) to the maximum read distance (in the range of 5 V_{PP}).

2. Provide a synchronized clock source to the tag. Many RFID tags divide the carrier frequency down to generate an on-board clock for state machines, counters, and so forth, and to derive the data transmission bit rate for data returned to the reader. Some tags, however, employ onboard oscillators for clock generation.
3. Act as a carrier for return data from the tag. Backscatter modulation requires the reader to peak-detect the tag's modulation of the reader's own carrier.

Although all the data is transferred to the host by amplitude-modulating the carrier (backscatter modulation), the actual modulation of 1s and 0s is accomplished with three additional modulation methods:

- *Direct modulation*: The amplitude modulation of the backscatter approach is the only modulation used. A high in the envelope is a 1 and a low is a 0. Direct modulation can provide a high data rate but low noise immunity.
- *FSK*: This form of modulation uses two different frequencies for data transfer; the most common FSK mode is $F_c/8$ and $F_c/10$. A 0 is transmitted as an amplitude-modulated clock cycle with period corresponding to the carrier frequency divided by 8, and a 1 is transmitted as an amplitude-modulated clock cycle period corresponding to the carrier frequency divided by 10.
- The amplitude modulation of the carrier thus switches from $F_c/8$ to $F_c/10$ corresponding to 0s and 1s in the bit stream, and the reader has only to count cycles between the peak-detected clock edges to decode the data. FSK allows for a simple reader design and provides very strong noise immunity, but suffers from a lower data rate than some other forms of data modulation.
- *PSK:* This method of data modulation is similar to FSK, except only one frequency is used, and the shift between 1s and 0s is accomplished by shifting the phase of the backscatter clock by 180°. Two common types of PSK are: (1) change phase at any 0 or (2) change phase at any data change (0 to 1 or 1 to 0). PSK provides fairly good noise immunity, a moderately simple reader design, and a faster data rate than FSK.

Regardless of what method of carrier modulation is implemented, any voltage modulation sequence controlled through the tag's memory or circuitry will result in the transmission of a bit pattern that mimics the modulation sequence. Therefore, essentially any binary information stored on the tag can be wirelessly transmitted back to the receiver.

As already mentioned, a problem unique to RFID systems is the vast difference in power between the signal outgoing from the reader and that returning to the reader as reflected from the tag. In some situations, this difference may be in the range of 80 to 90 dB, and the return signal may be impossible to detect. To avoid this problem, the return signal is sometimes modulated onto a subcarrier, which is then modulated on to the carrier. For example, in the ISO 15693 standard for RFID, a subcarrier of 13.56 MHz/32 = 423.75 kHz is used.

As defined by ISO/IEC 18000-3 (13.56 MHz), the modulation used is typically ASK (either 10% or 100%) for the forward link (reader to tag) and load

modulation for the reverse link with a rate defined as a division of the carrier. The load modulation produces subcarriers, which utilize binary phase-shift-keying (BPSK) modulation. Load modulation appears in the frequency domain as sidebands offset by the subcarrier frequency from the transmission frequency. Figure 6.9 illustrates this approach.

Future development of RFID devices will include more complex modulation schemes, for example technologies such as software defined radio (SDR) are implemented in dedicated short-range communications (DSRC) and other applications.

6.2.3 Data Encryption

RFID readers in public places can read the RFID data and connect to networks that provide real-time data about the owners and thus infringe on privacy. Encryption of data will help to solve this problem to certain extent. Encryption of the RFID tag contents will ensure that unauthorized readers are not able to access the data or, even if they can get the encrypted content, they do not have the access to encryption key. There are many encryption algorithms available that can be suitably used to encrypt the data on RFID.

EPCglobal has ratified its Gen2 global standard that uses frequency and power in a way that complies with the major regional regulatory environments. In addition to improvements in security of the data on the tag, the standard includes the ability to lock the identification fields in the tag, so that they cannot be spoofed or changed without a password. It also includes a strong kill mechanism, so retailers and others have the option of automatically erasing all data from the tag as it passes through a reader.

However, the standard does not allow for encryption, because one of the user requirements for the standard was that the tags be inexpensive. However, security issues will continue to be addressed in the hardware and policy working groups [10] and, in the meantime, implemented as proprietary solutions (AES, for example) by some of the equipment suppliers.

Current implementations of secure RFID rely on digital cryptographic primitives in the form of hashes and block ciphers. The presence of these blocks is motivated by privacy requirements, but they increase the overall processing latency,

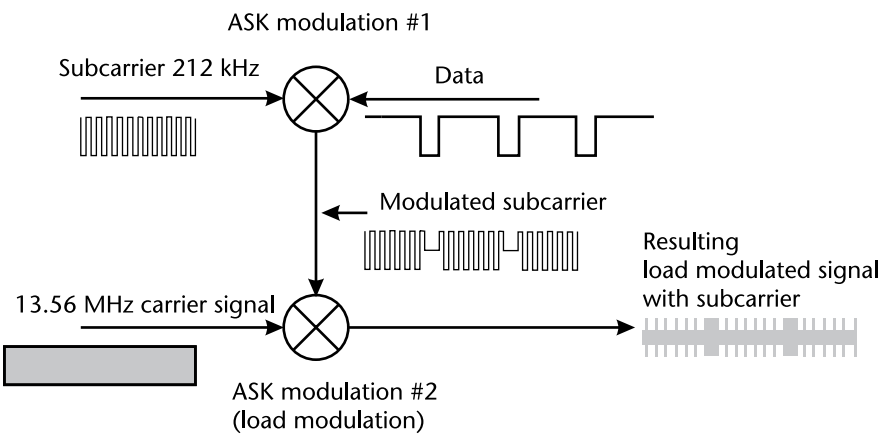


Figure 6.9 Load modulation diagram.

the power consumption, and the silicon area budget of the RFID tag. In addition, existing passive RFID systems rely on simple coding and modulation schemes using narrowband frequencies, which can be easily eavesdropped on or jammed.

6.2.3.1 Data Encryption Standard

The Data Encryption Standard (DES) algorithm is the most widely used encryption algorithm in the world. For many years, and among many people, secret code-making and DES have been synonymous. DES works on bits or binary numbers: the 0s and 1s common to digital computers. Each group of 4 bits makes up a hexadecimal, or base 16, number. Binary 0001 is equal to the hexadecimal number 1; binary 1000 is equal to the hexadecimal number 8; 1001 is equal to the hexadecimal number 9; 1010 is equal to the hexadecimal number A; and 1111 is equal to the hexadecimal number F.

DES works by encrypting groups of 64 message bits, which is the same as 16 hexadecimal numbers. To do the encryption, DES uses *keys*, which are also apparently 16 hexadecimal numbers long or apparently 64 bits long. However, every eighth key bit is ignored in the DES algorithm, so that the effective key size is 56 bits. In any case, 64 bits (16 hexadecimal digits) is the round number upon which DES is organized.

In cryptography, *triple-DES* is a block cipher formed from the DES cipher by using it three times. Given a plaintext message, the first key is used to DES-encrypt the message. The second key is used to DES-decrypt the encrypted message, and since the second key is not the right key, this decryption just scrambles the data further. The twice-scrambled message is then encrypted again, with the third key to yield the final ciphertext. In general, Triple-DES with three different keys has a key length of 168 bits: three 56-bit DES keys (with parity bits triple-DES has the total storage length of 192 bits), but due to the meet-in-the-middle attack, the effective security it provides is only 112 bits.

Triple-DES is slowly disappearing from use, largely replaced by its natural successor, the Advanced Encryption Standard (AES). One large-scale exception is within the electronic payments industry, which still uses these methods extensively and continues to develop and promulgate standards based upon it. This guarantees that triple-DES will remain an active cryptographic standard well into the future. By design, DES and therefore triple-DES, suffer from slow performance in software; on modern processors, AES tends to be around six times faster.

Triple-DES is still in use in connection with some hardware implementations, but even there AES outperforms it. Finally, AES offers markedly higher security margins; a larger block size, potentially longer keys, and hopefully freedom from cryptanalytic attacks.

6.2.3.2 Advanced Encryption Standard

AES was the result of a worldwide call for submissions of encryption algorithms issued by the U.S. National Institute of Standards and Technology (NIST) in 1997 and completed in 2000. The winning algorithm, Rijndael, was developed by two Belgian cryptologists, Vincent Rijmen and Joan Daemen. AES provides strong encryption and was selected by NIST as a Federal Information Processing Standard in

2001, and in 2003 the U.S. National Security Agency (NSA) announced that AES was secure enough to protect classified information up to the top-secret level, which is the highest security level, and defined as information which would cause exceptionally grave damage to national security if disclosed to the public [11].

The AES algorithm uses one of three cipher key strengths: a 128-, 192-, or 256-bit encryption key (password). Each encryption key size causes the algorithm to behave slightly differently, so the increasing key sizes not only offer a larger number of bits with which the data can be scrambled, but also increase the complexity of the cipher algorithm. AES is fast in both software and hardware, is relatively easy to implement, and requires little memory, and, as a new encryption standard, it is currently being deployed on a large scale.

6.3 Testing and Conformance

6.3.1 Test Equipment

The RFID engineers today face a variety of design and test challenges to bring a product to market. First, the product must meet local frequency regulations to emit energy into the spectrum; next, the interrogator and tag interaction must reliably work together. To accomplish this, both the interrogator and tag must comply with the appropriate industry standard. Finally, to be competitive, the RFID system's performance must be optimized to appeal to a particular market segment. This could mean maximizing the number of transactions per second, operating in a dense reader environment, or stretching the reader's ability to communicate over longer distances.

RFID systems, particularly those with backscattering passive tags, present some unique challenges for test and diagnostics. Timing measurements are of particular concern, as system readers can be required to read the ID data from many tags very quickly without error. Most RFID systems use transient Time Division Duplexing (TDD) schemes, where the interrogator and tags take turns communicating on the same channel. To read many ID tags within a very short period of time with a serial TDD multiplexing scheme, the standards call for very precise timing. Timing measurements on the data interchange thus present a unique RFID challenge. The transient RFID signals often contain spectrally inefficient modulations using special PCM symbol encoding and decoding. Troubleshooting the homodyne interrogators or tags that receive these unusual signals requires special signal analyzer capabilities. Traditionally, swept tuned spectrum analyzers, vector signal analyzers, and oscilloscopes have been used for wireless data link development.

The spectrum analyzer has historically been the tool of choice to characterize the RF spectral output of a transmitter to ensure compliance with regulatory emission restrictions. The traditional swept tuned spectrum analyzer was developed primarily for the analysis of continuous signals, not the intermittent RF transients associated with modern RFID products. This can lead to a variety of measurement issues, particularly the accurate capture and characterization of transient RF signals.

Similarly, the vector signal analyzer possesses little ability to capture transient RF signals, also being initially developed for CW signals. Though most vector

signal analyzers have extensive demodulation ability for popular spectrally efficient modulations, current offerings have virtually nothing to support the spectrally inefficient RFID modulations and their special PCM decoding requirements.

The oscilloscope has long been a valuable tool for analysis of baseband signals. In recent years some oscilloscopes have extended their sampling speed to very high microwave frequencies. They are, however, still suboptimal tools for UHF or higher frequency measurements on RFID systems. Relative to the modern real-time spectrum analyzer, the fast oscilloscope has substantially less measurement dynamic range and lacks modulation and decoding capability.

The real-time spectrum analyzer (RTSA) solves the limitations of the traditional measurement tools to provide a substantially more efficient test and diagnostic experience for the RFID engineer [12]. Pulsed tag reads and writes require an RF analyzer optimized for transient signals. The RTSA has the digital processing speed necessary to transform the input signal from time domain samples into the frequency domain with a real-time FFT prior to capturing a recording of data. This enables the RTSA to compare spectral amplitudes to a frequency mask set by the user in real time. The RTSA can then trigger a capture on a spectral event of interest for subsequent detailed off-line analysis.

Many RFID and near-field communications (NFC) devices use proprietary communications schemes that are optimized for specific market applications, so the test equipment should offer a variety of flexible modulation measurements that enable testing of the proprietary systems with manually configured measurements. The instrument should allow a user to define the modulation type, decoding format, and data rate.

Once the basic specifications are met, it is important to optimize some of the RFID product's features to gain a competitive advantage in a particular market segment. One such example is optimizing the number of tag reads possible in a given amount of time, resulting in the overall system capacity increase, and thus making it more appealing to high volume applications. An important element in maximizing capacity is minimizing the turnaround time for each tag reply; available RF power, path fading, and altered symbol rates can lengthen the time it takes for the tag to reply to the interrogator's query. The slower the reply, the longer it will take to read large number of tags.

6.3.2 Frequency and Bandwidth-Related Measurement

As mentioned earlier, bending tags or placing them in close proximity to conductive objects and other tags can detune the tag antenna, preventing them from going into resonance and thus either becoming inoperative or significantly reducing the operating range. Consequently, frequency deviation measurements are critical to ensuring compliance with various RFID and transmitter standards. For example, frequency accuracy for a 13.56-MHz RFID interrogator is typically specified at ±7 kHz.

Similar to frequency deviation measurements, occupied bandwidth measurements ensure compliance with standards designed to prevent interference with other signals. RFID readers and tags are intentional transmitters that fall under

regional regulations, such as FCC 47 part 15 in the United States, EN 300 330 in Europe, and ARIB STDT60/T-82 in Japan. As RFID heads towards global acceptance, the most stringent of these regulations will apply.

Also under close scrutiny is the effect of human exposure to RFID electromagnetic fields as spelled out in documents such as IEEE C.95-1 and EN50364:2010 ["Limitations of Human Exposure to Electromagnetic Fields from Devices Operating in the Frequency Range 0 Hz to 10 GHz, Used in Electronic Article Surveillance (EAS), Radio Frequency Identification (RFID), and Similar Applications"].

Various regulations will define limits using different units of measurement. The power flux density, S [mW/cm^2], electric field strength, E [V/m], and magnetic field strength, H [A/m], are interchangeable (see Chapter 5 for more details) according to the following equation:

$$S = \frac{E^2}{3700} = 37.7H^2 \tag{6.6}$$

6.3.3 Polling and Timing Measurements

When the RFID reader/interrogator searches for a tag, it is referred to as *polling*. Associated with polling is a number of timing measurements called out in various RFID standards. One key timing measurement is *turnaround time*, for both the transmit-to-receive and receive-to-transmit modes. Other timing measurements are dwell time or interrogator transmit power on ramp, decay time or interrogator transmit power down ramp, and pulse pause timing.

6.3.4 Collision Management

ISO/IEC 18000-3 calls for reading 500 tags within 390 ms. It also calls for reading 50 words of data within 930 ms from static tags, and 944 ms from active tags. If only a PC is used to time stamp the interactions and test for compliance, there is no way to know why the interaction takes so long, at which point a is collision occurring, or where there is a particular tag that is being problematic.

However, by using test equipment and monitoring the over-the-air interface during polling, it is possible to troubleshoot when a collision occurs and determine the cause (interference, faulty tag, hopping pattern error, and so on).

6.3.5 Multivendor Interoperability and Testing

If RFID is to successfully penetrate into large open systems, RFID interoperability is a necessity. Not only must tags from any vendor be able to communicate with readers from any vendor, but a given tagged object must be able to be identified by readers of any user in a wide variety of application conditions.

Today, RFID systems are primarily composed of systems that may not always interoperate due to the mix of RF propagation technologies and information protocols. The formation of the EPCglobal, a joint venture formed by UCC and EAN,

is expected to drive global retailer and manufacturing adoption of RFID technologies, especially in the supply chain management applications.

Key to market proliferation of UHF Gen 2-compliant products is the EPCglobal program to certify the hardware that implements the standard. This includes the testing of tag chips, readers, and printer/encoders with embedded reader modules. Products that pass the tests conducted by MET Laboratories (an independent third-party lab contracted by EPCglobal) earn the EPCglobal certification marks, a seal of approval, indicating the products' adherence to the stringent requirements of the standard. EPCglobal has defined three phases of certification:

- Compliance;
- Interoperability;
- Performance.

Compliance testing verifies that products comply with the UHF Gen 2 standard, and products bearing the certification mark are your assurance that they have been rigorously tested against EPCglobal standards. Some of the relevant EMC and safety standards are the following.

In the United States:

- U.S. EMC—FCC Rule Part 15 or 90;
- Safety—UL 60950 for Tag Interrogators and NRTL Certification.

In Canada:

- Canada EMC—RSS-210;
- Safety—CSA 60950 and SCC certification body.

In Europe:

- Europe EMC testing in accordance to ETSI EN 301 489-1 and ETSI EN 301 489-3;
- Radio testing in accordance to ETSI 300-220;
- Safety testing in accordance to EN 60950;
- Declaration of Conformity for CE marking requirements.

FCC recently classified passive RFID chips as *unintentional radiators*. This implies that passive RFID chips are exempt from the same clearance tests as other technical devices with electromagnetic fields.

Interoperability testing builds on the compliance certification and verifies the ability of different compliance-certified Gen 2 components to work together. Despite the obvious advantages of assuring interoperability among products, not all RFID vendors have been able to achieve this level of certification.

Clearly, certified UHF Gen 2 interoperability is a major milestone in the development of RFID systems. As important as that is, though, performance is still what

matters most to RFID deployments. RFID hardware must have a high degree of receptivity, meaning both tags and readers are not only extremely sensitive to each other's signals, they are also able to reject the interference from other RF sources operating in the area. EPCglobal, recognizing the critical importance of receptivity to system performance, created a working group to address these and other issues.

In the process of defining minimum requirements, they will address not only the performance of tags applied to various classes of products, such RF-friendly materials as paper, plastic, wood, and so forth, as well as more problematic materials such as liquids and metals. They will also the key aspects of tag performance, like sensitivity, interference rejection, orientation, Electrostatic Discharge (ESD), and others parameters.

The Hardware Action Group (HAG) handles the specifications for *performance testing* and thus completing the UHF Gen 2 standard. The group is specifically chartered to address conformance and interoperability work for the UHF Gen2 conformance document and any future UHF work, conformance and interoperability work to for HF, and conformance and interoperability work for any other air interface specifications that may be developed in the HAG.

Although all products submitted to interoperability testing must first be certified for compliance to the Gen 2 standard, it is not uncommon for some manufacturers to have misinterpreted certain elements of the specification, preventing their tags, for example, operating with other Gen 2 devices. More insidiously, certain tags and readers may be interoperable with each other, but not with all other Gen 2 devices. As such, the scope of interoperability tests should be designed to exercise, as much as possible, the full functionality of the Gen 2 spec (including operation at timing limits) with a prime objective of assuring true multivendor compatibility.

Interoperability problems between the various industries are on the horizon because cross-industry requirements usually play a minor role within a given industry. However, to fully exploit the technology's potential, cross-industry standards have to ensure interoperability. Therefore, cross-industry consultation is necessary to prevent the emergence of different standards from impeding a use of RFID across industries.

We have mentioned a few times already that RFID system calculations are just a first approximation of the real-world environment. RFID systems do not always necessarily work as well as theoretically described or as advertised. This can be due to interference in the environment, unforeseen reflections, or simply an installation that does not account for the peculiarities of the technology. Theoretical formulas and analysis, even computer simulations, may not be always sufficient to represent customer environments and predict the RFID system behavior.

Independent test labs simulating the real customer environment are the answer; services offered usually include the identification, evaluation, and integration of prototypes, support for middleware and applications, and development of hardware including antennae and tags. These labs will allow companies to test RFID systems in real customer environments and iron out any potential problems. In addition, to ensure that different scenarios can be tested at the center, some of these labs have built prototypes for a number of industries, including pharmaceuticals, retail, logistics, manufacturing, electronics, government, and transportation.

6.4 Review Questions and Problems

1. RFID tag may be used in situations where tagged objects like pallets or boxes travel on a conveyor belt at speeds up to either 10 feet/second (3.048 m/s). The tag spends less time in the read field of RFID reader, meaning that a high read rate capability is required. In such cases, the RFID system must be carefully planned to ensure reliable tag identification. The other potential problem could be the Doppler effect.

   Calculate the Doppler shift [using (6.7), where f is transmitting frequency, Δf is the change in the frequency of the reflected signal, and c is speed of light] for this case, at 915 MHz, and check if the operation could be affected.

$$\Delta f = \frac{f \cdot v}{c} \tag{6.7}$$

   (*Answer:* The Doppler shift is only around 9 Hz and will not affect the correct operation of the system.)
2. Describe one application of active and passive RFID technology and discuss and compare the following aspects: range, multitag operation capabilities, data storage, sensor capabilities (temperature, humidity, shock, temper detection, security), business process impact, country-specific and global standards, and coexistence with other technologies.
3. One example of the battery-assisted (semipassive) UHF RFID system one that uses 50% duty cycle Manchester encoding for the forward link; this is possible because the RF signal does not have to power the tag. The tag is designed not for power gathering, but for optimized signal detection. This allows much smaller signals to be picked up and amplified on the chip. FSK backscatter signaling is using different frequencies to signal a high or low, making it easier to discern signal from noise because the reader must only find a frequency, not a signal edge.

   Discuss the design summary and decide where and when you would use this not-all-that-inexpensive system.
4. To provide information on otherwise invisible tag detection process, some propose the use of a so-called *watchdog tag* to provide required transparency. Simply speaking, the watchdog tag is a sophisticated version of an ordinary tag, as it features an additional battery, a small screen, and potentially even a long-range communication channel. The watchdog tag's main task is to decode the commands transmitted by a reader and make them available on the screen of the device for inspection by the user or to log all data transfers and provide consumers with detailed summaries whenever needed.

   Although the watchdog tag could be carried by the user as a separate device, its functionality could also be integrated into a mobile phone, allowing it to leverage the existing display, battery, memory capacity, and long-range communication features of the phone.

How much complexity would this additional feature add to reader-tag protocols? Would the read speed and the number of tags read be affected? What are your thoughts on usefulness of the watchdog tags?

5. A reasonable definition of privacy is necessary to consider various policy choices. However, privacy is a notoriously difficult concept to define. The United Nations in Universal Declaration of Human Rights, Article 12, codifies that "no one shall be subjected to arbitrary interference with his privacy" as a basic human right. It has even been suggested that "all human rights are aspects of privacy" [13]. In 1890, Supreme Court Justice Louis Brandeis famously articulated privacy as the "right to be left alone." Ruth Gavison of the *Yale Law Journal* defined three core aspects to privacy: secrecy, anonymity, and solitude.

   Considering these issues, Simson Garfinkel has written an "RFID Bill of Rights" based on the U.S. Department of Health and Education's Code of Fair Information Practices. Garfinkel's bill of rights reads as follows:

   > Users of RFID systems and purchasers of products containing RFID tags have:
   > a. The right to know if a product contains an RFID tag;
   > b. The right to have embedded RFID tags removed, deactivated or destroyed when a product is purchased;
   > c. The right to first class RFID alternatives: consumers should not lose other rights (e.g., the right to return a product or to travel on a particular road) if they decide to opt-out of RFID or exercise a RFID tag's "kill" feature;
   > d. The right to know what information is stored inside their RFID tags and what information is associated with those tags in associated databases. If this information is incorrect, there must be a way to correct or amend it.
   > e. The right to know when, where and why an RFID tag is being read."

   Which one of these rights will be most difficult to implement and why? List these rights from the most important towards the least important ones. Do you agree with all of them? Add a few other requirements/rights that are important for you and the society in which you are living.

6. In operating outside the allocated spectrum, in many jurisdictions it may be possible to obtain special permission from the regulators of the radio spectrum to use equipment that operates outside the prevailing legislation. For example, the testing and development of new forms of RFID equipment or the temporary use of noncompliant RFID equipment may be allowable on a case-by-case basis.

   A large number of factors will typically be taken into account under such circumstances, for example, the location of the equipment to be deployed, the potential for interference with legitimate users of the spectrum, and so forth. This approach has been used by some companies in Europe wishing to start their RFID trials with equipment conforming to North American legislation. In these cases, the anticipated use of the noncompli-

ant equipment is for a limited period and operation of the equipment has been modified to minimize the chance of it causing interference.

Assume that you are working as an RFID system designer for a large company that would like to test and potentially implement a newly designed RFID system in a frequency band that was not originally allocated to RFID. What kind of approvals and licenses would you require in the area in which you are living? Where would you start your regulatory battle to implement nonstandardized system? What would happen if you decide next year to open warehouses in another country and want to use the same RFID system over there? Is international harmonization a good idea? Could international harmonization slow down or expedite your company's project? Discuss the answers to these questions in detail.

7. Item management includes both the identification of an item and its location. Whereas RFID provides a means of radio identification, RTLS provides a means of radio location. RTLS is being applied in a number of industries, ranging from health care to manufacturing. RTLS has an important role to play as far as real-time data is concerned and where assets are required to be located in transit.

   According to an analysis, it is expected that the revenue generated by RTLS will increase steadily. The shortcomings in the existing systems have initiated the need for RTLS. The ability to generate real-time data is the major driver behind RTLS.

   a. What type of RFID system would you use for the application shown in Figure 6.10? What other components of the network are required to make the system work? Discuss the project.
   b. Condition monitoring is another valuable function that can be added to supply chain or asset management applications. Temperature,

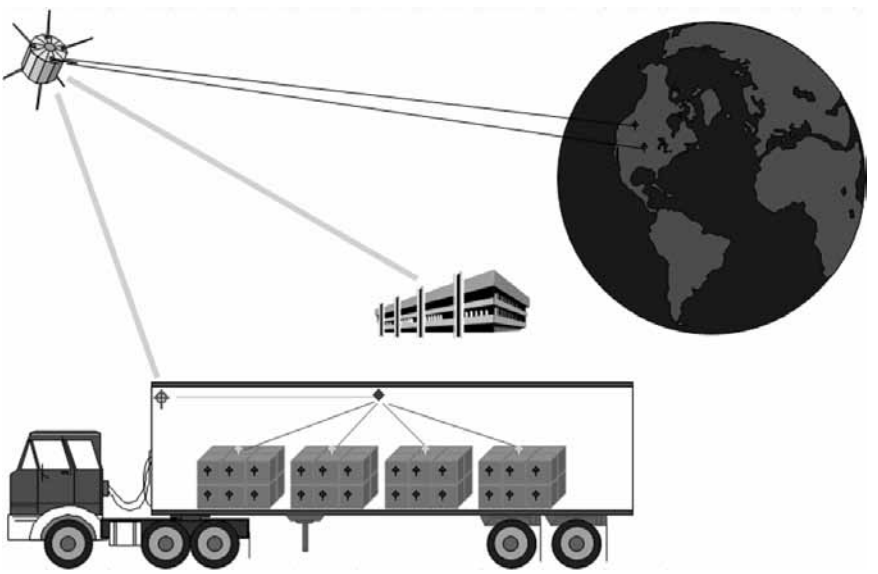


Figure 6.10 Satellite/RFID system for tracking of the transported goods.

humidity, access, and use of assets can be monitored. Discuss adding condition monitoring to the previous project of the location monitoring.

8. One priority for RFID use is hardware certification since RFID interoperability is a condition for successful open applications. Tags from any vendor must communicate with readers from any vendor, and tagged objects must be readable in a variety of application conditions. Discuss the RFID certification process in the country in which you are living.
9. The EU's Directive on Waste Electrical and Electronic Equipment (WEEE 2002/96/EC Directive) does not explicitly rule out the possibility that RFID chips will be seen as waste electrical and electronic equipment. With RFID technology becoming increasingly more widely adopted, do you think that RFID chips should be treated as dangerous material? Does the size of RFID chips play role in making that decision? Should active tags be discarded differently than passive tags?
10. It has been said that the real-time identification and location facilitate tracking, effective tracking improves communication, and better communication is necessary to streamline and coordinate workflow which is a goal of every organization, including a health care organization. Application of RFID will enable the better analysis and improvement of its operations both in real time and long term. Where do you see the application of RFID in health care being useful and where, in your personal opinion, should society be apprehensive and careful about it?

References

[1] Chawathe, S. S., et al., "Managing RFID Data (Extended Abstract)," University of Maryland, 2005.

[2] Nikitin, P. V., and K. V. S. Rao, "Performance Limitations of Passive UHF RFID Systems," Paper, Intermec Technologies Corporation, 2006.

[3] Avoine, G., and P. Oechslin, "RFID Traceability: A Multilayer Problem," Lausanne, Switzerland: Ecole Polytechnique Federale de Lausanne, 2006.

[4] Piramuthu, S., "Anti-Collision Algorithm for RFID Tags," *Mobile and Pervasive Computing*, CoMPC-2008, 2008.

[5] Weis, S., et al., "Security and Privacy Aspects of Low-Cost Radio Frequency Identification Systems," Laboratory for Computer Science, Auto-ID Center, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[6] Jain, S., and S. R. Das, "Collision Avoidance in a Dense RFID Network," Computer Science Department, Stony Brook University, New York, 2006.

[7] Kim, S., et al., "Reader Collision Avoidance Mechanism in Ubiquitous Sensor and RFID Networks," Korea University, Sungbuk-ku, Seoul, Korea: Department of Electronics Engineering, Korea University, 2006.

[8] Engels, D. W., et al., "Colorwave: An Anticollision Algorithm for the Reader Collision Problem," *IEEE International Conference on Communications*, Vol. 2, 2002, pp. 1206–1210.

[9] Stallings, W., *Data and Computer Communications*, 6th ed., Upper Saddle River, NJ: Prentice-Hall, 2000.

[10] Juels, A., "RFID Security and Privacy: A Research Survey," RSA Laboratories, September 28, 2005.

[11] Federal Information Processing Standards, Publication 197, *Announcing the Advanced Encryption Standard (AES),* November 26, 2001.

[12] Tektronix, "RFID and NFC Measurements with the Real-Time Spectrum Analyzer," Application Note, 2005.

[13] Volio, F., *The International Bill of Rights: The Covenant on Civil and Political Rights, Chapter Legal Personality, Privacy, and the Family,* Columbia University Press, 1981.

CHAPTER 7

RFID Technology for Medical Applications

Many organizations, including health organizations, are adopting RFID technologies as part of their information supply chains for the myriad benefits that come through the use of such devices. Today RFID is not just for supply chain applications anymore, and significant investment has gone into making the technology simple, practical, and cost-efficient, especially in embedded applications. A conflicting set of challenges that manufacturers continually face is how to increase safety and revenue while decreasing costs.

In health care challenges include making sure that health care providers have the right products at the right place at the right time and avoiding overstocking, spoilage, and shrinkage. Another problem is how to protect authentic disposables against counterfeits and the resulting threats to diagnosis/treatment reliability and eroded recurring revenues. In addition, they have to continually improve the quality of patient care through better process reliability and proper device configuration to drive brand equity, increase revenue, and minimize legal liability.

Embedded RFIDs[1] now enable medical and health care solutions through supporting applications such as inventory management, disposables authentication and configuration, and patient management.

This chapter will focus on a completely different area of RFID applications: their integration with sensor networks and utilization for the medical implants. *In-body implants* and/or *on-body sensors* are already allowing hearing for the deaf, sight for the blind, and mobility for the disabled. Implants can stimulate muscles or nerves in response to movement detected by sensors elsewhere on the body, allowing a paralyzed patient to walk again. Similarly, a radio-controlled valve for the urinary tract is in development that will be operated on demand to restore bladder control.

Section 2.5 provided a useful introduction into the WBANs and applications of wireless technologies in the health industry, with a brief overview of the most important standards. There are a number of other medical applications of implantable medical devices and sensors, so here we will take a brief look at their principles of operation. Technical and ethical (especially in the case of neural implants) challenges will be addressed as well.

1. Embedded RFID refers to enclosing RFIDs into a device and using a reader, which is in many cases completely hidden from the end user.

7.1 Integrating RFIDs and Sensor Networks

Only a small part of all RFIDs are directly used in (simpler) medical applications, some of which were described earlier in Chapter 2. Examples are tracking patients, tracking medical supplies, and so on. At the same time, many of the same technologies and same principles are used both in commercial RFIDs and some more sophisticated medical applications, such as wireless body implants. By combining RFID principles of operation and sensor and wireless sensor networks, the foundations for more sophisticated WBANs have been created (see Figure 7.1). Sensor networks were covered in more detail in Section 3.4.

Both near-field and far-field operations are used in commercial RFIDs and also in their medical counterparts. In addition, they share some of the same requirements and constraint: small size, long-lasting power supply, RF propagation challenges, and so forth.

Of course, wireless principles used in medical applications bring a whole new set of challenges, such as requirements for increased reliability, biocompatibility, no harm to patient, ethical dilemmas, and so forth, that have to be discussed and resolved. This chapter will focus on some of the issues specific to the medical applications of wireless principles and RFIDs.

It is important to keep in mind that this technology is still very young and will require many years of development until its widespread application in the health/medical world.

7.1.1 Basics of Biomedical Signals

Signals received from a biological or medical source are called *biomedical signals* (or biosignals for short), and their source can be at the molecular level, cell level, or systemic or organ level.

Examples include the Electrocardiogram (ECG), or electrical activity from the heart; speech signals; the Electroencephalogram (EEG)[2], or electrical activity from the brain; evoked potentials (EPs) (i.e., auditory, visual, somatosensory), or electrical responses of the brain to specific peripheral stimulation; the *electroneurogram*,

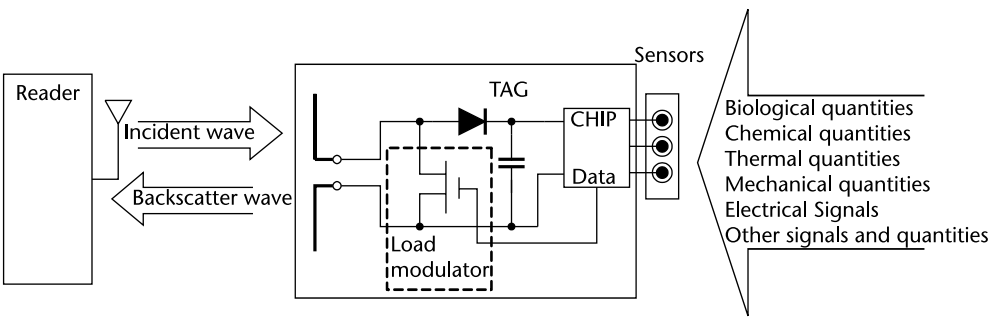


Figure 7.1 RFID tag with sensors.

2. Discovered by Hans Berger (1873–1941) in 1924, electroencephalography (EEG) is currently one of the most commonly utilized methods for the detection and measure of brain activity. Berger first published his brainwave results in 1929 as *Uber das Elektrenkephalogramm des Menschen*, but the English translation did not appear until 1969.

or field potentials from local regions in the brain; action potential signals from individual neurons or heart cells; the Electromyogram (EMG), or electrical activity from the muscle; the *electroretinogram* from the eye; and so forth.

Typically, biomedical signals are primarily acquired for monitoring specific pathological/physiological states for purposes of diagnosis and evaluating therapy. In some cases of basic research, they could also be used for the decoding and eventual modeling of specific biological systems.

The monitored biological signal is usually a combination of signal and noise. Noise can be any signal that is asynchronous and uncorrelated with the biological signal of interest; for example, from instrumentation (sensors, amplifiers, filters), from Electromagnetic Interference (EMI), or in general. A Gaussian white noise assumption is valid in many biological cases.

The biomedical signal sources can be broadly classified into *continuous processes* and *discrete-time processes*. Each of these types of signals could be *deterministic* (or predictable), *stochastic* (or random), *fractal*, or *chaotic*. The continuous processes are typically encountered in one of the following situations:

- *Deterministic signals in noise:* Examples of this type are ECG or single-fiber EMG signals in noise. The measured signal $x(t)$ can be represented as follows:

$$x(t) = s(t) + n(t) \tag{7.1}$$

  where $s(t)$ is the actual deterministic signal and $n(t)$ is the additive noise.
- According to the information about biomedical signal processing found in [1], deterministic signals could be synchronized to another stimulus signal or perturbation in noise. Examples of this type include all the different evoked responses (auditory, somatosensory,[3] visual) and event-related potentials recorded in response to controlled stimuli administered to the body (or any biological system in general). These signals usually reveal functional characteristics of specific pathways in the body.
- *Stochastic or random signals:* Examples of this type include EEGs, EMGs, and field potentials from the brain. Random signals lack the morphology of the signals found in the preceding two categories. Depending on the underlying physiology, the stochastic biosignals could be stationary where statistics of the signal do not change with time, or nonstationary (fluctuations in the signal statistics due to physiological perturbations such as drug infusion or pathology or recovery).
- *Fractal signals:* Signals and patterns in general are *self-replicating*, which means that they look similar at different levels of magnification. They are therefore scale-invariant. There is evidence to suggest that heart rate variability is fractal in nature. The branching of the airway into bronchioles seems to have a self-replicating nature that is characteristic of a fractal.
- *Chaotic signals:* They are neither periodic nor stochastic, which makes them very difficult to describe or predict beyond a short time into the future. The

3. Somatosensory response pertains to sensations received in the skin and deep tissues.

difficulty in prediction is due to their extreme sensitivity to initial conditions, characteristic of these nonlinear systems. While fractal theory details the spatial characteristics of the nonlinear systems, chaos theory describes the temporal evolution of the system parameters or the dynamical variables.

- *Multichannel signals:* They include signals of any of the listed types but that were acquired using multichannel recording technology. The goals of signal analysis are usually to identify correlation among different channels, to achieve feature extraction under noisy conditions, and to identify underlying physiology.

7.1.2 Sensor Networks in Medicine

Telemedicine is the generic term describing application of telecommunications to provide medical information and services by collecting and processing data from human body sensors and wirelessly transmitting them in real time to a central monitoring system. The purpose of telemedicine is to provide medical expertise to places that, due to their physical distance, do not have access to this kind of competence.

The biopotential wireless sensor node targets the simultaneous monitoring of vital body signs. A sensor is a device that is used to detect a change in physical conditions and chemical compounds; sensors generally have a mechanism attached or built into them that make the change in the physical or chemical condition readable to the human eye.

A biosensor can be generally defined as a device that consists of a biological recognition system, often called a *bioreceptor*, and a transducer, converting that recognition and response into a quantifiable term that can be interpreted [2]. There are different kinds of biosensors that are classified on the basis of their interaction with the biological system: noninvasive biosensors, indwelling biosensors, and invasive biosensors.

Noninvasive biosensors (also called *wearable sensors*) can be placed on the surface of the body. *Indwelling biosensors* can be placed in body cavities or under the skin without causing many changes in the body. *Invasive biosensors* are surgically placed in the human body. These sensors have a long battery life and are encapsulated in materials which are not harmful to the individual. All of these sensors vary in physical configuration, connection with body organs, packaging, user interface, and power supply in accordance with their function.

Implantable medical devices (IMDs) are being used for the treatment of various diseases. IMDs use wireless biomedical sensors as an interface between electronic instrumentation and the human body. These sensors take information from the biological system and transfer it to an external system for monitoring, thus integrating wireless communication, sensing, and computing.

A sensor network is designed to detect events or phenomena, collect and process data, and transmit sensed information to interested users. Generally speaking, basic features of sensor networks are:

- Dense deployment and cooperative effort of sensor nodes;
- Short-range communication and multihop routing;

- Self-organizing capabilities;
- Frequently changing topology due to fading and node failures;
- Limitations in transmit power, memory, and computing power.

A WBAN intended for medical applications could be seen as a wireless sensor network since most medical applications will rely on sensors collecting data about (e.g., the heart and the brain). As such, the sensor nodes must be kept simple in order to fulfill requirements on energy-efficiency and long battery lifetime. The wireless sensors' networks require efficient hardware, software, signal-processing algorithms, and network protocols for operation.

New challenges will follow the migration of biosensor technology from in vitro (procedure performed not on a living organism but in a controlled environment) to in vivo (experimentation using a living organism). In a BAN with limited bandwidth and power constraints, the conventional method of data acquisition and analog-to-digital data conversion with signal processing taking place after transmission is no longer optimal, so the local processing will typically take place at the sensor front end before transmission.

The challenge is to develop low-power communication with low-cost on-node processing and self-organizing connectivity/protocols; another critical challenge is the need for extended temporal operation of the sensing node despite a limited power supply and/or battery life. To conserve energy, the sensor node should be kept as long as possible in power-down or sleep mode [3].

Power efficiency in wireless sensor networks (WSNs) is generally accomplished in three ways: by low-duty-cycle operation (i.e., the active time period of the sensor node), local in-network processing to reduce data volume and hence transmission time, and multihop networking. In addition, the node communication protocols should be kept simple not requiring a lot of computation, and also more advanced data/signal processing should be avoided in the sensor node. In particular, the architecture of the radio, including the use of low-power circuitry, must be properly selected, meaning that the low power consumption for transmission over low-bandwidth channels.

The new practical solutions include an application of energy scavenging in the powering of BAN devices. Body-powered applications, however, remain a great challenge because of the low specific power levels at low frequencies; therefore, substantial progress will be required in reducing power requirements before such solutions become feasible, particularly for wireless data transmission.

If every sensor can transmit directly (i.e., in a single hop) to the hub, then the network can have the star topology. However, a single sensor cannot always communicate directly to the hub because of the energy and power constrained nature of the sensors and because of the multipath fading and shadowing effects.

Multihop routing is a traditional wireless communication approach to this problem, passing the message to the node that is in the optimal position to transmit to the hub device. Multihop networking reduces the requirement for long-range transmission since signal path loss is an inverse exponent with range or distance. Each node in the sensor network can act as a repeater, thereby reducing the link range coverage required and, in turn, the transmission power [4].

7.1.3 Medical Implants

Medical implants are not a novelty. They are used in nearly every organ of the human body, and more than 1,800 types of medical devices are currently in use. Some of them have been used for decades. These vary from heart valves, pacemakers, and cochlear implants to drug infusion devices and neurostimulating devices for pain relief or to combat certain disorders such as Parkinson's disease.

What is different and new today is that increasingly more of those implants contain some kind of built-in intelligence and a capability to communicate and exchange information with the environment.

Generally, IMDs can be divided into two major groups:

- Sensory aids (retinal implants, cochlear implants, and so forth);
- Assessment and treatment devices (implantable biomedical sensors and implantable medical treatment devices).

Monitoring body functions is an essential tool in medical diagnosis, for example, repeated measurement of blood pressure at short intervals is mandatory for some diseases. Continuous monitoring can be achieved by on-body or in-body medical devices. Medical devices are defined as any product that is used to treat patients or as a diagnostic tool; they may be external or implanted and reusable or disposable.

In United States, the FDA regulates medical devices according to specific definitions, classifications, requirements, codes, and standards. The FDA's authority and framework for medical device regulation are specified in the Federal Food, Drug, and Cosmetic Act of 1938 (FDCA). For purposes of medical device regulation, several acts of Congress amending the FDCA are especially significant: the Medical Device Amendments of 1976, the Safe Medical Devices Act of 1990, the Food and Drug Administration Modernization Act of 1997, and the Medical Device and User Fee and Modernization Act of 2002 [5].

Two examples of implants, implantable camera pill and cardiac pacer, are shown in Figure 7.2.

The Wireless Capsule Endoscopy (WCE) is based on a large, vitamin pill–size capsule that captures the images of the digestive tract, while it is transported passively by peristalsis (Figure 7.3).

The device consists of an image sensor, an illumination module, an RF transmitter, and a battery. The patient swallows the capsule, and the camera takes and transmits about two images per second as it travels through the gastrointestinal (GI) tract. A camera sends images to a recorder that the patient wears on a belt [6]. These images are stored on the recording device and later downloaded onto a computer for viewing by the physician.

Wireless inductive coupling has been used to communicate between the pacemaker and a wearable reader outside the patient. The new technology using Bluetooth or MICS can provide a simple interface with the patient, where the information can be accessed by mobile phone or interface Internet to connect the patient.

Microsystems include microelectromechanical systems (MEMS), microfluidics, and microarrays and have diverse medical applications, for example, biosensors and detectors to detect trace quantities of bacteria, airborne pathogens, biological

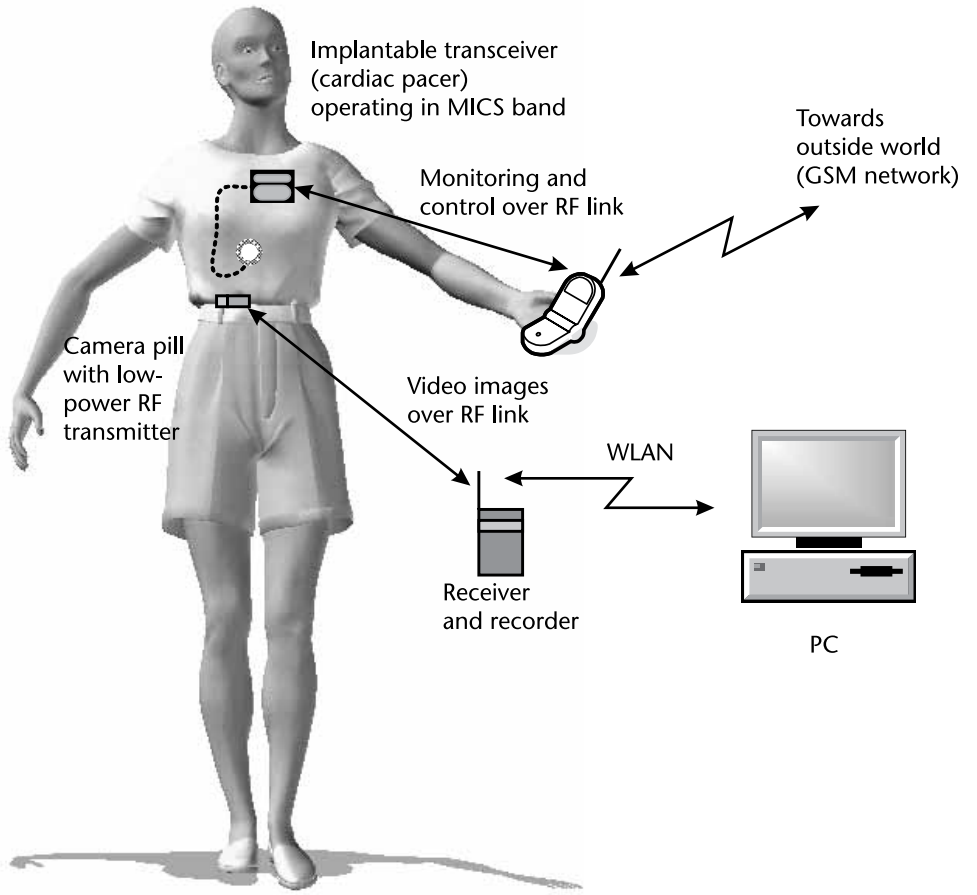


Figure 7.2 IMDs.

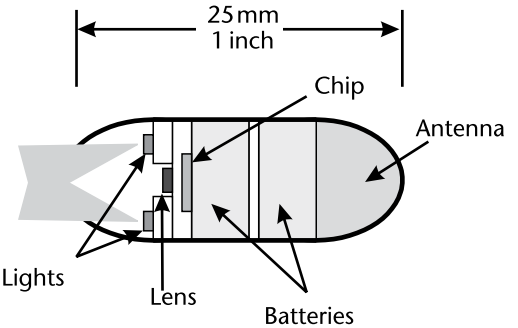


Figure 7.3 Capsule endoscopy camera. (*After:* Mayo Foundation for Medical Education and Research [104]).

hazards, and disease signatures; microfluidic applications for DNA testing and implantable fluid injection systems; and MEMS devices that contain miniature moving parts for heart pacemakers and surgical devices [7].

7.1.3.1 Sensory Aids for Hearing and Visual Loss

Loss of hearing is very common and approximately one person in a 1,000 is born deaf worldwide. Almost an equal number of people born with hearing will develop deafness during their lifetime.

Patients with severe hearing loss typically have absent or malfunctioning sensory cells in the cochlea. In a normal ear, sound energy is converted to mechanical energy by the middle ear, which is then converted to mechanical fluid motion in the cochlea. Within the cochlea, the inner and outer ear sensory cells are sensitive transducers that convert mechanical fluid motion into electrical impulses in the auditory nerve.

Cochlear implants[4] are the substitute for the function of the middle ear, cochlear mechanical motion, and sensory cells. The implants transform sound energy into electrical energy that will initiate impulses in the auditory nerve.

Today one of the highly active areas of research is the development of *visual prosthesis*. Visual prosthesis necessitates the development of sensors, very large-scale integration circuitry, power delivery, telemetry, and advance packing and tissue interfaces. Some degenerative diseases of the retina, such as retinitis pigmentosa[5] or age-related macular degeneration, could severely decrease night vision and can progress to diminishing peripheral vision and blindness.

In cases where the neural wiring from the eye to the brain is still intact but the eyes lack photoreceptor activity, photoreceptor loss could be compensated by bridging or bypassing the destroyed photoreceptors and artificially stimulating the adjacent intact cells. These artificially generated impulses will reach the brain and produce visual perception, thereby restoring some (elementary) vision.

Partial restoration of visual function (e.g., facial recognition or navigation through a building) involves the development of implantable retinal prosthesis devices with sophisticated image processing and neural interfaces [8].

One approach being developed by various groups including a project at Argonne National Laboratory is an artificial retina implanted in the back of the patient's own retina; the artificial retina uses a miniature video camera attached to eyeglasses in order to capture visual signals. The signals are processed by a microcomputer worn on the belt and transmitted to an array of electrodes placed in the eye, stimulating optical nerves, which then carry a signal to the brain.

A different approach uses a subretinal implant designed to replace photoreceptors in the retina [9]. The visual system is activated when the membrane potential of overlying neurons is altered by current generated by the implant in response to light stimulation. The implant uses a microelectrode array powered by 3,500 microscopic solar cells.

4. Cochlear implants cannot restore hearing to normal, but it can give the sensation of sounds; they work well for adults and children who have lost their hearing after acquiring spoken language and for young children who were born deaf.

5. Retinitis pigmentosa (RP) is a group of genetic eye conditions that leads to incurable blindness due to a progressive loss of photoreceptors. In the progression of symptoms for RP, night blindness generally precedes tunnel vision (reduction of the peripheral visual field) by years or even decades.

7.1.3.2 Implantable Assessment and Treatment Devices

Nanotechnology[6] is the analysis, understanding, and manipulation of matter at structure sizes from 0.1 to 100 nm, offering sensing technologies that provide more accurate and timely medical information for diagnosing disease and miniature devices that can administer treatment automatically if required. Microsensors and nanosensors can make use of a wide range of technologies that detect a targeted chemical or physical property.

One can also look at biomedical sensors from the standpoint of how they are applied to the patient or research subject. General approaches to attaching biomedical sensors are as follows:

- Noninvasive (noncontacting and skin surface or contacting);
- Invasive (indwelling or minimally invasive and implanted).

Clearly, if a measurement can be made equally well by a sensor that does not contact the subject being measured or by one that must be surgically implanted, the former is by far the most desirable. However, a sensor that is used to provide information to help control a device surgically placed in the body to replace or assist a failing organ should be implanted, since this is the best way to communicate with the internal device.

Some types of implantable sensors use MEMS devices and accelerometers for monitoring and treatment of paralyzed limbs. With Functional Electrical Stimulation (FES) therapy, a patient uses an external device to wirelessly communicate with several implanted and on-body devices. By stimulating several muscles and nerves in sequence, a paralyzed patient can stand and even walk a few steps.

Implantable sensors can also work with a series of medical devices that administer treatment automatically if required. Tiny implantable fluid injection systems can dispense drugs electrically on demand making use of microfluidic systems, miniature pumps, and reservoirs. Initial applications may include chemotherapy that directly targets tumors in the colon and are programmed to dispense precise amounts of medication at convenient times, such as after a patient has fallen asleep. Lupus, diabetes, and HIV/AIDS applications are also being investigated.

For example, the awarded patent in 2006 is the first out of 10 pending patents by Integrated Sensing Systems, Inc. (ISSYS) for its MEMS technology for safe, chronic, fast, detailed, real-time, continuous, biopressure measurements. The particular targets of ISSYS products are cardiovascular disease, especially congestive heart failure, and hydrocephalus (high brain pressure) disease. The U.S. Patent Office has granted a patent entitled "Implantable Sensing Device for Physiologic Parameter Measurement" (U.S. Patent No. 6,968,743), which covers the design,

6. Triggered by recent discoveries in the area of nanomaterials, nanoelectronics has shown the potential to introduce a new approach in electronic devices and system design. Since many nanoscale materials and devices exhibit their most interesting properties at RF, nanotechnology has entered the microwave engineering research arena. In the future, nanotechnology-based innovation may be utilized in the different areas of RF system miniaturization and diversification, energy-efficient RFID devices, and even in the extended sensing and cognitive functionalities.

fabrication, and manufacturing of miniature, wireless, batteryless, implantable sensors [10].

While sensors themselves may be neutral, the service supplied and the data gathered may give rise to ethical concerns. For example, the same type of sensors used to monitor a forest fire could also be employed on the opposite side of a wall to observe a person's activities based on body heat. The U.S. Supreme Court has barred law enforcement officials from using this form of technology in *Kyllo vs. United States* in 2001 [11].

7.2 Operational Challenges of Implanted Devices

Some of the challenges that WBANs are facing in health applications are the development of better sensors, improving wireless networks, correctly using the information, demonstrating that this technology reduces cost and improves patient care, and making sure that the privacy of the end user is guaranteed and that technology is used in the highly ethical manner.

The American Association for the Advancement of Science (AAAS) claims that scientists and ethicists should weigh the impacts of human enhancements, such as, among others, steroid use by athletes, implanted devices for treating depression, and the possibility of genetic manipulations to prolong human life [12].

Some of the challenges faced when deploying wireless-based solutions include engineering and quality issues, social issues, and patients' well-being issues. An implementation-related challenge is interoperability among various devices, as well as devices in different countries.

For patients, an important issue is how the quality of their life will be affected by using these new applications, so designing applications that can be useful while unobtrusive is another challenge for wireless networks-based application/solution developers.

Some of the main challenges related to the use of wireless devices definitively include security, privacy, and the learning curve for new technologies. Ensuring patients' information security, combined with the privacy of user data over wireless channels, can be a major challenge.

Wireless-based medical devices can be very limited in terms of power availability and processing strength. With the new technologies surrounding us in our daily lives, new users can find it challenging to use these more sophisticated devices to the fullest. Thus, it can be an issue and a challenge for engineers to create some of the best solutions without forcing the users to make unnecessary effort just to learn how to use them.

7.2.1 Biomedical Materials Inside of the Human Body

IMDs that are greater than 1 mm in diameter may affect the functions of surrounding tissue. Smaller implantable devices with nonintrusive or minimally-intrusive

systems will likely contain nanoscale materials and smaller systems approaching the nanoscale[7].

Many biological systems are designed to deal with foreign materials by trying to eliminate them. The rejection reaction that is often discussed with regard to implanted materials or transplanted tissues is an example of this. Thus, in considering biomedical sensors, the main concern is the rejection phenomenon and the way it will affect the performance of the sensor. If the rejection phenomenon changes the local biology around the sensor, this can result in the sensor measuring phenomena associated with the reaction that it has produced as opposed to phenomena characteristic of the biologic system being studied [13].

On the other hand, biologic systems can also affect sensor performance, especially indwelling and implanted sensors. The body generates a hostile corrosive environment to most materials that are commonly available since body fluids are largely composed of salt water. Thus, the sensor package must not only protect the sensor from the corrosive environment of the body, but it must allow that portion of the sensor that performs the actual measurement to communicate with the biologic system (human body in this case).

Despite recent development in biomedical field, the current state-of-the-art prosthetic platforms still lack reliable and convenient packaging methods to integrate high-density signal-driving chips, wireless telemetry circuitries, and noise-canceling amplifiers. A useful discussion on the present challenges of retinal implants and a description of the potentially new packaging technology is presented in [14].

Due to the limited access to implanted devices, sensors and electronics have to be highly reliable so that there is no need to repair or replace them very often. It is also important that these sensors are highly stable, because in most applications it is not possible to calibrate the sensor in vivo. Thus, sensors must maintain their calibration once they are implanted, and for applications such as organ replacement, this can represent a potentially long time, the remainder of the patient's life [15].

Implants should function within the human body with no side effects introduced by the implant's presence. Any implant containing moving parts (heart-valve or hip implant, for example) will almost certainly release wear particles, and unless the implant is made of a truly inert material, some corrosion will eventually occur. The body has only a limited tolerance for corrosion products and wear particles released by an implant; a major problem with orthopedic implants is the release of wear particles from the moving surfaces.

The implant with the highest load capacity is the hip implant (Figure 7.4), giving users 20 years of service without any maintenance. The longevity of the hip implant will depend, of course, on the age and the lifestyle of the patient.

Under the Safe Medical Devices Act (SMDA) issued in 1990, amended by the FDA Modernization Act (FDAMA) in 1997, and revised in 2008, manufacturers are required to track certain medical devices (mostly implantable ones) in the marketplace, so that the device can be promptly identified and removed in the event of a product recall [16].

7. While nanodevices are increasingly becoming possible to make in the laboratory, larger-scale microdevices are effective solutions in most cases. As a result, nanodevices are more likely to be in future applications and will not be discussed here.

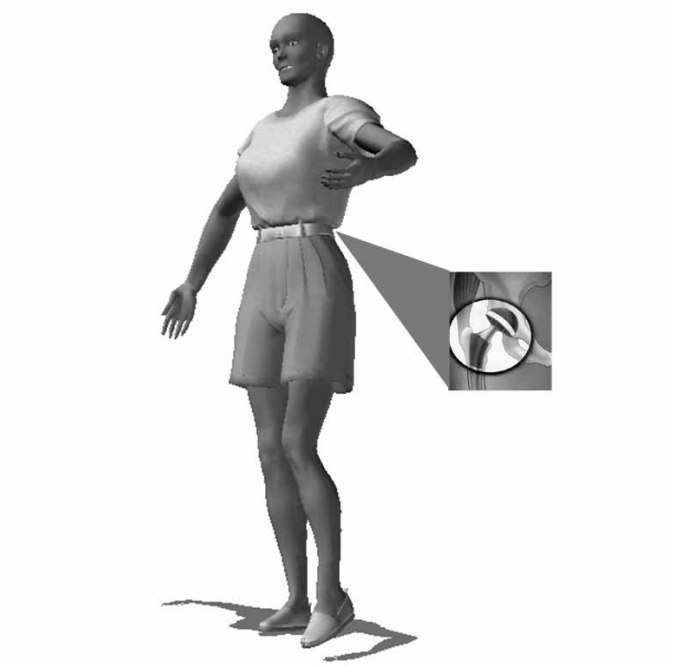


Figure 7.4 The hip implant.

Each time a biomedical material is inserted in the body, there is a risk that bacteria will be introduced into the body on the surface of the biomedical material. Even fine metal particles and oxides or hydroxides of metals such as chromium and titanium have been observed to cause inflammation, swelling, and possibly even cancer in some patients. The enzymes and proteins present in tissue fluids are observed to amplify the corrosive nature of the tissue fluids.

If there is inflammation around the implant, then the implant environment becomes even more corrosive because of the secretion of strong oxidants by the human cells that initiate the inflammation. It is not sufficient for a material to be merely nontoxic; the material should not initiate an immune or inflammatory response from the human body nor should it function as a protected niche for bacterial infections.

Medical implants are being used in every organ of the human body, and, ideally, they should have biomechanical properties comparable to those of real tissues and without any adverse effects. Whenever a biomedical material is inserted in the body, adsorbed[8] proteins will almost inevitably cover this material. In some cases, this protein adsorption may lead to blood clotting, and a lump of clotted blood could block an artery, causing a thrombosis [17].

Detailed studies of the long-term effects of medical implants must be undertaken to accurately determine the performance and safety of the implants. For example, from a clinical point of view, it is important to realize that, at least for now, the implanted microelectrodes into the brain of patients will only provide a

8. Adsorption is the adhesion of atoms, ions, biomolecules or molecules of gas, liquid, or dissolved solids to a surface.

year or so of operation. The electrodes cause scarring and would be implanted in eloquent regions of the cortex; repetitive procedures could have significant detrimental effects to the patient's long-term functional and cognitive status. Invasive brain electrodes, therefore, need a prolonged lifespan to warrant the risks of an intracranial procedure.

To date, current single-unit microelectrodes have long-term biocompatibility issues leading to limited lifespans; however, there are several groups developing new biomaterials as well as slow-release drug delivery systems that could decrease encapsulation. Dexamethasone, a potent synthetic member of the glucocorticoid class of steroid hormones that acts as an anti-inflammatory and immunosuppressant, on the microelectrode might reduce the initial injury response; its potency is about 20 to 30 times that of hydrocortisone and four to five times of prednisone. Dexamethasone is commonly used in transvenous screw-in cardiac pacing leads to minimize the inflammatory response of the myocardium [18].

The ability to predict the long-term in vivo performance of medical implants is of vital interest. The extrapolation of in vitro data to the in vivo environment remains, in most cases, unproven. Among the major challenges are the limited ability to simulate the complexities of the biological environment, the current lack of reliable computer modeling of in vivo performance characteristics of implants, and difficulties in evaluating the synergistic contributions of materials, design features, and therapeutic drug regimens.

There has been some concern about the possibility of tumor formation by the wide range of materials used in body implants. Although many implant materials are carcinogenic in rats, there are few well-documented cases of tumors in humans directly related to implants. It may be premature to make final judgment since the latency time for tumor formation in humans maybe longer than 20 to 30 years. In this case, we have to wait longer for a final assessment.

However, the number of implants being placed in the body and the lack of strong direct evidence for carcinogenicity tend to support the conjecture that carcinogenesis is species-specific and that no tumors (or an insignificant number) will be formed in humans by the implants.

7.2.2 Radio Propagation Inside the Human Body

While the characteristics of different wireless networking technologies partially overlap, some were originally designed with different particular applications in mind. Table 7.1 details the common North American and European frequency bands that are suitable for certain medical applications. The absence of global approach is preventing technology of spreading and impeding the use of these devices.

Good examples of global standardization challenges are the remote controlled infusion pumps, operating in the band 902–920 MHz, which is allowed to operate in North America and Australia/New Zealand, but cannot be used in Europe because the band is extensively used for cellular telephony.

As of the time of this writing, there were no standards for sensor networks or WBANs, and one reason for this could be the wide range of application scenarios for those networks. A standard needs to have support for as many applications as possible, but the result is a more complex and time-consuming document.

Table 7.1 Frequencies for Medical Applications

Frequency (MHz)	*Region*	*Band*	*Regulation*
402–405	North America, Europe	MICS (license-exempt)	FCC, 47 CFR 95.601-673, Subpart E; EN 301839
433.05–434.79	Europe	General telemetry band (license-exempt)	EN 300220
868–870	Europe	General telemetry band (license-exempt)	EN 300 220
602-614	North America	WMTS	CFR, Part 95
902–928	North America, South America, Australia	915-MHz ISM band (license-exempt)	FCC 15.247
2,400–2,483.50	Europe	2.4-GHz ISM band (license-exempt)	ETS 300 328
2,400–2,500	North America	2.4-GHz ISM band (license-exempt)	FCC 15.247
5,650–5,925	North America	5.8-GHz ISM band (license-exempt)	FCC 15.247

A study group of the IEEE was launched in November 2007 to work on the WBAN[9] standardization; the IEEE 802.15 Task Group 6 (BAN) is presently developing a communication standard optimized for low-power devices and operation in regards to the human body (but not limited to humans) to serve a variety of applications including medical, consumer electronics/personal entertainment, and other [19].

Table 7.2 summarizes features of different technologies considered for WBANs and is based on [20].

According to [21], the communications between an access point to the backbone network are dependent on the application. From the hospital to the backbone, IEEE 802.11 WLAN can be employed. Besides WLAN, a traditional PCS/cellular network can also be used in the home case. IEEE 802.11p Vehicular Ad hoc Networks (VANET) or IEEE 802.16e Worldwide Interoperability for Microwave Access (WIMAX) is proposed for the mobile ambulance case. For the ambulance helicopter, IEEE 802.16e WIMAX is a good candidate.

A good summary of the specifications, output power, frequency bands, and international use of several applicable RF wireless technologies can be found in the ISO technical report [22].

7.2.2.1 Implanted Wireless Systems

When upgrading medical devices with any kind of wireless connectivity, it is important to remember that the communication link quality and data integrity are vital if they are used for diagnostics or therapeutics. Even wearable sensor networks present problems in terms of managing transmitted power (and therefore distance and channel conditions) and power consumption that can dramatically change with

9. Wireless body area networks (WBAN) has emerged as a key technology to provide real-time health monitoring of a patient and diagnose many life threatening diseases. WBAN operates in close vicinity to, on, or inside a human body and supports a variety of medical and nonmedical applications.

Table 7.2 Comparison of Technologies for WBANs

	UWB IEEE 802.15.6	*ZigBee IEEE 802.15.4*	*Bluetooth IEEE 15.1*	*WLAN IEEE 802.11b/g*	*MICS*	*WMTS*
Frequency (MHz)	1–10 GHz	2.4 GHz, 868, 915	2.4 GHz	2.4 GHz	402–405	608–614, 1,395–1,400, 1,427–1,432
Bandwidth (MHz)	>500	5	1	20	3	6
Data rate (kbps)	850	250 (2.4 GHz)	1 Mbps	>11 Mbps	>250	>250
Tx Power (EIRP)	–41 dBm	0 dBm	4 dBm, 20 dBm	24 dBm	–16 dBm (25 μW)	The maximum field strength at 3m*
Multiple Access	ALOHA	CSMA/CA	FHSS	OFDMA, CSMA/CA	CSMA/CA, Polling	CSMA/CA, Polling
Range (feet)	6	30	30, 300	300	30	>300

*The maximum field strength at 3m: 608–614 MHz (200 mV/m), 1,395–1,400 MHz, and 1,427–1,429.5 MHz (740 mV/m).
Source: [20].

the posture of the body and movements. The topic was covered by Quwaider in his article [23]. Much more challenging are medical applications that utilize wireless communication with implanted electronic circuits.

As with all measured quantities, the user must be confident that the data retrieved is accurate, requiring the usual techniques of instrument design and calibration. However, an additional constraint for wireless devices is that the data must be secure. Since many devices operate in the license-exempt ISM frequency bands, there is a severe risk of interference that could be particularly dangerous in the context of a medical device.

Power consumption is another major challenge, especially for medical implants in particular, leading to highly optimized solutions for ultralow-power consumption from the implanted battery. From the engineering perspective, power consumption, range, and data rate are opposing design requirements for wireless implants and have to be optimized within the radio regulatory constraints.

In the simple wireless system block diagram, shown in Figure 7.5, the analog data from a sensor measurement is first converted from its continuous-time analog signal into a discrete-time digital signal. A baseband processor then adds error-detection information and formats the data for transmission, and after that the baseband data is sent to the transceiver, encoded onto an RF signal and broadcasted by the transmitter, detected by the receiver, and finally decoded back into digital data.

The human body is a difficult medium for radio propagation; it is partially conductive and consists of materials of different dielectric constants and characteristic impedance. This means that at the interface of two body materials, such as muscle and fat, the difference can cause a wave to be partly reflected rather than transmitted. Signal penetration into body tissue is also important. A reduction of signal to 36% of the peak is an often-used reference, and it can be seen that propagation at the certain frequency is about four times better in the fat than in the muscle tissue.

The cerebrospinal fluid is also highly conductive, and transmitting the RF signal within the head is similar to transmitting a radio wave through an electrically

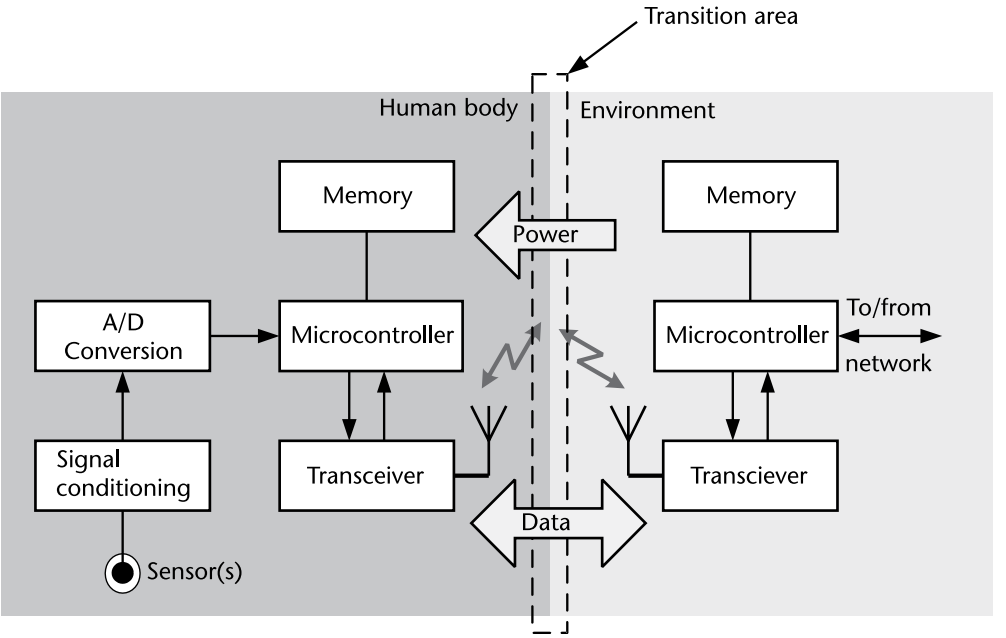


Figure 7.5 A simplified implanted wireless systems.

shielded room. Such transmission is possible only when the RF signal is strong and its frequency is relatively low, which requires more energy consumption [24].

A potential RF performance problem is the placement of the implanted device. A surgeon will place an implant where it is required and will be clinically most effective, with little concern for RF propagation. Therefore, the antenna must operate effectively from various depths and through layers of fat, muscle, and skin with unpredictable thickness. In addition, all these parameters may change with time.

The power limit applies to the signal level outside of the body (environment), which allows for implant power levels to be increased compensating for body losses. Once the implant is in place, the RF link can be used to adjust the signal level to the highest allowable level. The impedance radiated by the antenna will also vary as the patient moves or ages and the location of the implanted device shifts. An automatic tuning circuit and firmware routine that operates each time the transceiver is powered up can compensate for impedance change.

At low frequencies, electromagnetic energy has significant penetration capabilities, and the body can be used to support communications channels. For example, at 10 MHz the penetration depth is about 200 mm (8 inches) for muscle and over 1m (3 feet) for fat. At 2.45 GHz the depths are 25 mm (1.0 inch) and 120 mm (5 inches), respectively [25]. Radio propagation through the human body and on the body-environment boundary (transition area) has been a topic of intense study and discussions for some time.

7.2.2.2 Wearable and Implanted Antennas

Transmitting data, images, and videos from inside the body taken by a radio system of a size of a pill, with capabilities of tracking the system, delivering drugs to

specified organs and entering the body in noninvasive way, are all being developed today. A key element of an RF-linked implant is the in-body antenna, which must meet stringent biocompatibility and size-limit requirements.

Antenna design is a mature scientific and engineering discipline, but all the literature on this topic has one thing in common, that is, they mainly describe antennas placed in a nonconducting surrounding with a relative permittivity of 1, or close to 1. In other words, they describe antennas placed in vacuum or air. The only structure that is typically found close to the antenna is a radome, which is made of low loss materials with low permittivity. When the antenna is placed inside a human body, a completely different situation is created: the antenna is surrounded by a lossy material with high permittivity. This is a completely new area of research since there are only two instances in classical antenna applications in which similar conditions occur: buried antennas and submarine antennas.

When the medical implant is placed with an antenna inside a patient, the antenna will be affected by the immediate surroundings, and, as a result, the antenna behaves differently if placed in an arm, deep in the abdomen, or just beneath the skin in the chest. In addition to this, there will be a dependency on the surrounding tissue type, for example, variations in the subcutaneous fat layer. This layer varies in thickness between patients and varies also over time when a patient gains or loses weight; therefore, the far field from the antenna is affected by the patient's size, body shape, and position. Movements of the patient change the immediate surroundings of the implanted antenna.

The MICS frequency band, 402–405 MHz, corresponds to a wavelength of approximately 30 inches in the air and approximately 4 inches inside the body. The body surface is in the near field of the implanted antenna and any change of the permittivity or conductivity of materials placed in the near field of an antenna changes its radiation characteristics. Thus, a change in posture changes the far-field pattern and affects the radio channel between the medical implant and the external base station [26].

Unlike free-air performance, the human body is an unpredictable and hostile environment for a wireless signal. The human body is a conductive medium representing a lowpass filter for the electromagnetic waves, and engineers are then often forced to use low frequencies or, equivalently, long wavelengths. If the space for the antenna is limited, the antenna will be small compared to the wavelength; the drawback is that small antennas in high-loss materials consume more power, due to the losses in the near zone of the antenna. Hence, the design of the antenna and the choice of frequency are delicate problems, where two power loss mechanisms with counteracting frequency dependences are involved [27].

Conducting materials, such as metals, act similar to perfect reflectors for UHF radiation. Materials such as glass, concrete, and cardboard are effectively RF transparent for waves that are incident upon them with an angle of incidence of 90°, but they become less transparent as the angle of incidence becomes more oblique.

In modeling antenna behavior in the human body, there are two different cases: the antenna in an infinite body of lossy matter and the antenna in a finite body of lossy matter that is placed in air. The last one follows from the first, if the finite body with the internal antenna is treated as one large antenna. At every boundary between two materials, electromagnetic waves incident upon that boundary will

be both transmitted from one material to the other and reflected back into the material in which they are traveling.

7.2.2.3 Modeling Human Body

Some materials, such as water, act as both good reflectors of electromagnetic waves and good attenuators, or absorbers, of electromagnetic energy. The partial reflection of a wave results in the energy of the wave being separated to traverse multiple paths. The result is that a partial reflection attenuates the partially transmitted wave by the amount of energy reflected at the boundary.

Detailed radio propagation characterization within the human body through means of various simulation techniques and measurements was described and statistical models of the radio channel were derived and shown in [28]. The study demonstrated the importance of optimum frequency choice taking into account attenuation factors, tissue conductivity, and also antenna size and orientation, which not only can affect radiation inside the body but also can determine the optimum distance of which good performance can be achieved in the surrounding environment.

Animal organs are sometimes used to characterize and simulate human tissues; the animal organs used are sheep liver, heart, and lungs, with the heart possessing electric properties close to the stomach dielectric constant and conductivity. In the literature [29] there are recipes for tissue-simulating liquids for muscle, brain, lung, and bone tissue as well as descriptions of polyacrylamide solutions, which simulate fat tissue.

Until recently, the majority of work that has been done to obtain a detailed understanding of radio transmission near the human body has focused on the head and neck. Early work used relatively simple body models that assumed that human tissue was homogenous, but slowly the research moved to more sophisticated models [30]. More recently, there have been a number of studies looking at the abdominal region using detailed models for both female bodies [31] and male bodies [32].

Results also show that the higher frequency is attenuated more when propagating through the body, and thus the lower frequency is better for implants placed deeper inside the body. Despite the fact that absorption of electromagnetic radiation increases with frequency, it is found that, up to a point, increasing the frequency can improve the far-field signal strength from an ingested or implanted source. The reason for this is that the size of the capsule device demands that an electrically small antenna be used, and that typically the antenna should be smaller than $\lambda/2$ (the preferred size for the simplest radiating device). As a consequence, when the frequency is increased, the wavelength decreases towards the antenna dimensions, increasing the antenna's efficiency.

Table 7.3 lists the conductivities, the dielectric constants, and the penetration depths of muscle, fat, bone, and skin tissue at 405 and 650 MHz, based on [33].

The electrical properties of the body tissues are frequency-dependent and should be identified for the frequency of interest. The permittivity of a material is a function of numerous factors including its constituent materials. For biological materials, water is one of the major constituents. It is important to realize when making biological measurements that the relative percentage of water in the body varies with such things as gender, age, physiologic state, and tissue type [34].

Table 7.3 Electrical Properties of the Human Body

Tissue Type	*Frequency f (MHz)*	*Conductivity σ (S/m)*	*Relative Permittivity* ε*	*Penetration Depth δ (mm)*
Muscle	405	0.797590	57.0880	52.464
	650	0.864120	55.7620	46.874
Dry skin	405	0.690200	46.6720	55.064
	650	0.782760	43.1330	45.846
Fat	405	0.041199	5.5777	308.280
	650	0.045587	5.5068	275.040
Bone cortical	405	0.091780	13.1360	212.080
	650	0.115300	12.7240	165.500

*The older term, still is use, is the relative dielectric constant.

The formulas for the radiation resistance and reactance assume a medium with a complex propagation constant $\gamma = \alpha + j\beta$, where α and β depend on the electrical parameters σ and ε. In addition to the high losses that the transmitted RF wave will face, the composition of the biological tissues inside the human body is very complex and varies from one person to another. Thus, numerical models, no matter how complex, will only be estimates since an accurate solution to the problem of wave propagation in the human body is still being developed [35].

It has been found that there is a competing effect between the increased efficiency of the antenna and the increasing absorption of the body tissue, so an optimum trade-off frequency is found to be in the region of 650 MHz.

7.2.3 Power Requirements for Implanted Devices

In medical applications, where the diagnostic device is to be implanted inside the human body, only limited power is available for the electronic circuits and sensor. In such cases, power can only be found via a battery implanted together with the sensing system or through passive telemetry. For an implant device with a battery life of 7 to 10 years, energy must be carefully conserved, since power consumption determines the life of the implant as well as the battery size.

Energy is mostly consumed during data transmission, so the smaller the transmission range and the slower the data can be sent, the lower the power consumption. A common approach to saving power is to keep the transmitter circuits powered off when not required and keeping radio circuits in a receive mode when not transmitting, but an even more effective technique is to also power-down the receiver circuits and periodically wake the receiver up to check for an asynchronous transmit request. The receive section is duty-cycled on and off in less than 100 μs, which makes it possible to achieve very low average power consumption while monitoring the MICS channel for transmitted messages. When on, the receiver checks to see if a signal is present and if a strong enough signal is located, the entire receive section can be quickly powered up to obtain data. The technique heavily depends on a rapid-start oscillator that can wake up the receiver and (if needed) the transmitter in an extremely short time [36].

Many receivers are using two independent receive channels (diversity) to boost the reception range and improve the reliability of MICS transmissions. More sophisticated transceivers also contain baseband Clock and Data Recovery (CDR) circuits that postprocess the demodulated incoming data stream to produce both a sampled data bit stream and a clock signal. That process helps improve transmission reliability by synchronizing the data processing clock with the incoming data and the result is less power wasted with retransmissions.

There have been a number of research projects over the last few years dedicated to these alternative (and sometimes unusual) sources of energy with the potential to power RFID and WBAN devices and networks, described in [37, 38]. For example, University of Maryland's A. James Clark School of Engineering and College of Agriculture and Natural Resources is harnessing and exploiting the "self-renewing" and "self-assembling" properties of Tobacco Mosaic Virus (TMV) to build a new generation of small, powerful, and highly efficient batteries and fuel cells.

In medical devices, for example, a patient's normal daily activities could power an implantable pump that delivers insulin to a diabetic. The use of piezoelectric materials to harvest power has already become popular since they have the ability to transform mechanical strain energy into electrical charge; for example, they can also be embedded in shoes to recover "walking energy." A new and more efficient piezoelectric energy harvester without rectifying diodes principle is described in [39].

In *passive telemetry*, energy may be harvested from a remote electromagnetic field transmitted outside the body. The same field may also be used to receive control data and transmit sensor data. Passive RFID tags obtain impinging energy during reader interrogation periods, and this energy is used to power tag IC. In the near field, tag to reader communication is achieved by modulating the impedance (load modulation) of the tag as seen by the reader [40].

The other solution is the so-called *far-field energy harvesting* that utilizes the energy from the interrogation signal's far-field signal to power the tag. In the far field, tag-to-reader communication is achieved by modulating the RCS of the tag antenna (i.e., backscatter modulation).

Electromagnetic power can also be delivered through skin [41], allowing implanted devices to be powered indefinitely. However, such a scheme requires an external transmitting device to deliver EM energy continuously, or to recharge a battery. A novel power supply for implantable biosensors has been described by Goto et al. [42]. In this power supply, Near Infrared (NIR) light transmission recharges a lithium secondary battery wirelessly through the skin. The Sun is a good source of NIR light, and its use requires no other external device to deliver energy to the recharging system. A photovoltaic cell array and the rectifier using Schottky diodes implanted beneath the skin can receive NIR light through the skin, and charge the battery that is directly powering an implanted biosensor(s).

A new approach to reduction of power consumption and miniaturization is to create a small network of sensors within the body; all the sensors are wirelessly communicating with the central unit, which, in turn, communicates with the external signal processing unit. In [43] authors describe an Intra-brain Communication (IBCOM), a wireless signal transmission method that uses the brain itself as a conductive medium to transmit the data and commands between neural implants and data processing systems outside the brain. The concept was validated through

a series of experiments on rat brain. More details regarding powering RFID tags were given in Section 5.5.

7.3 Development of Medical Devices

7.3.1 Technology Transfer

Technology transfer is the communication of information from research and development to the users or vice versa. Technology transfer also occurs when an innovation is transferred to another application as an idea, prototype, or useful product. There are many examples of technology transfer such as applications that began as military or space exploration (Teflon comes to mind) and later became commercial products. Medical technologies usually develop as a result of breakthroughs in other disciplines but also medical technologies can also transfer to other nonmedical and/or commercial applications.

The development of computed tomography (CT)[10] scans in 1970s is an example of a medical technology that now has a variety of other uses. The imaging techniques essential in many of the sensing technologies carry over into areas of fine art, drawing, and photography, for example. Technology that is not successful in medicine may prove successful elsewhere and vice versa. Figure 7.6 graphically depicts this process.

As described in [44], cognitive neuroscience combines experimental strategies of cognitive psychology with various techniques to find correlation between brain functions and mental activities. Sometimes technology can be used for purposes for which it was never intended. For example, the development of functional neuroimaging (fMRI)[11] has brought many ethical issues. While some of these ethical issues are typical bioethical issues, including safety (e.g., for scans involving radiation or high magnetic field strengths) and researchers' obligations when incidental findings of abnormal brain structure or function are observed in research scans, others arise from the ability to correlate brain activity with psychological and mental states.

One of the most widely discussed new applications of functional neuroimaging is based on correlations between brain activity and intentional deception (lying) [45]. All lie detection methods, like all human cognitive behavior, have their roots in neuroscience. The most common and commonly used lie detector, the *polygraph*, does not measure directly activity in the subject's brain. From its invention around 1920, the polygraph has measured physiological indications that are associated with the mental state of anxiety: blood pressure, heart rate, breathing rate, and galvanic skin response (sweating).

10. CT is a medical imaging method employing tomography created by computer processing. Digital geometry processing is used to generate a 3-D image of the inside of an object from a large series of 2-D X-ray images taken around a single axis of rotation.
11. Functional MRI (fMRI) uses MRI technology to measure blood flow in the live brain, which corresponds with activity in the brain [blood oxygen level dependent (BOLD) signals]. fMRI dominates the brain mapping field due to its relatively low invasiveness, absence of radiation exposure, and relatively wide availability.

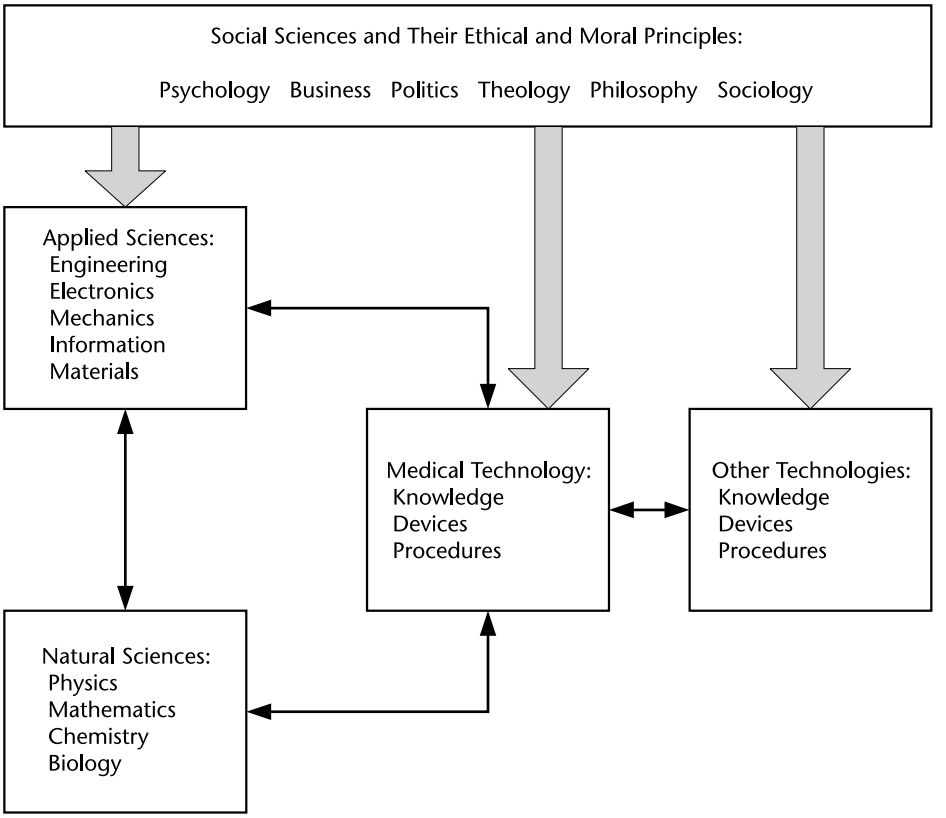


Figure 7.6 Technology transfer process.

The term *neuroscience-based lie detection* describes newer methods of lie detection that try to detect deception based on information about activity in a subject's brain. In spite of the lack of convincing proof of efficacy, some companies in the United States, due to the lack of regulations, are able to offer fMRI-based lie detection services.

7.3.2 Medical Product Development

A *hazard* is a potential source of harm, and in order to minimize use-related hazards, to assure that intended users are able to use medical devices safely and effectively throughout the product life cycle, and to facilitate review of new device submissions and design control documentation, a thorough understanding of the medical device is required [46].

For the safe and effective use of the medical device, the user, the device, and the environment have to be closely examined[12]. ISO 14155 standard defines procedures for conducting clinical investigations of medical devices that will protect human subjects, ensure proper scientific conduct in the clinical investigation, and

12. Regulations for conducting medical device clinical trials around the world can vary widely. A medical device clinical trial can cost between $5 and $10 million in the United States or Western Europe and more in Japan. The cost of the same trial conducted in Eastern Europe will be considerably lower, and due to the relaxed or even nonexistent regulations in India, China, or Korea, it may be 10 times cheaper.

assist sponsors, monitors, investigators, ethics committees, regulatory authorities, and bodies involved in assessing medical device conformity.

ISO 14155 helps assure that data generated anywhere in the world meets minimum standards of quality for study design, planning, conduct, documentation and human subject protections [47] and it consists of two parts: *ISO 14155-1:2003 Clinical Investigation of Medical Devices for Human Subjects—Part 1: General Requirements* and *Part 2: Clinical Investigation Plans*.

When used as intended, the medical device should work perfectly well; the question is what happens when the device is in the different environment and/or used in an incorrect or different way than for what it was designed. Many of these and similar issues can be addressed through a risk analysis during the design stage, but in most cases not all potential situations could be predicted and/or accounted for [48].

In addition to the functional assessment, the ethical evaluation of implantable devices is required to assess, at the minimum, the following areas of concern: issues of safety and informed consent, issues of manufacturing and scientific responsibility, anxieties about the psychological impacts of enhancing human nature, worries about possible usage in children, and issues of privacy and autonomy. As is the case in evaluation of any future technology, the reliable prediction of all the effects is not possible; nevertheless, the potential for harm must be considered through risk analysis.

Medical devices (MDs) are subject to the general controls of the Federal Food Drug & Cosmetic (FD&C) Act which are contained in the final procedural regulations in Title 21 Code of Federal Regulations Part 800-1200 (21 CFR Parts 800—1299). These controls represent the baseline requirements that apply to all medical devices necessary for marketing, proper labeling, and monitoring its performance once the device is on the market [49].

As described on the FDA's Web site [50], the three-step process for obtaining marketing clearance from the Center for Devices and Radiological Health (CDRH), are:

- *Step 1* is making sure that the product is indeed a medical device. It could also be an electronic radiation emitting product with additional requirements.
- *Step 2* is a determination of how the FDA may classify the device: into which one of the three classes the device may fall.
- *Step 3* is a development and/or collection of data and/or information necessary to submit a marketing application and to obtain FDA clearance to market.

In the EU, the first step in launching of an MD is to obtain the CE marking. The MD, active implantable medical devices (AIMD), and in vitro diagnostic medical devices (IVDMD) market is based on a European regulatory framework governed by three directives, 90/385/CEE for AIMD, 98/79/CEE for IVDMD, and 93/42/CEE for the others, the so-called new approach. These specify that MDs and IVDMDs can be marketed only if their manufacturers have previously appended the CE marking. This requirement does not apply to devices intended for clinical

investigation, to custom-made medical devices, or to in vitro diagnostic medical devices for evaluating performance [51].

The manufacturer must prove that his or her device conforms to the requirements of the directive in question, before appending the CE marking to the device. The CE marking symbolizes compliance of the device with the essential requirements of the directives.

The three stages of the clinical development of a new MD are the preclinical phase, the clinical phase, and a launch of a new product. The preclinical phase of the MD development consists of technological revisions, in vitro tests, and possible animal experiments. The clinical phase consists of feasibility studies (patient selection, surgical technique, clinical efficiency, complications, and risks) and studies demonstrating clinical benefits. MDs are divided into classes as a function of their level of risk, and the principal texts of the European regulations that apply to MDs are available on the European Commission Web site [52].

The FDA has also officially recognized device-specific and general standards published by standards bodies such as the Association for the Advancement of Medical Instrumentation (AAMI) and the International Electrotechnical Commission (IEC). The FDA general and specific guidance and standards recognized by the FDA are listed on the FDA's Web site.

To support decision making, the Medical Technology Assessment (MTA)[13] evaluates medical technology based on medical efficiency and other aspects. The Constructive Technology Assessment (CTA) in the early stage aims to provide decision-makers involved in technological development and health care with a tool to help steer technological development [53].

Many of the decisions here are cost-related. Three groups of stakeholders are likely to lobby on behalf of new technologies even if their benefit is less than their cost: product manufacturers who want to make the new technology, physicians and hospitals that want to use it, and patients who want or need to receive it. All three are relatively small groups whose members have relatively high individual stakes in the issue; on the other side are the people who will ultimately pay for the very expensive technology, but are never included in the decision-making process: taxpayers and insured people [54].

Many emerging technologies, in addition to the technical challenges and cost-related issues, present ethical challenges, and wireless body implants are not an exception. There are two aspects of ethical and moral issues important for discussion about WBIs and their implementation. The first one is related to the problem of security and privacy invasion, and it is commonly discussed in society today. The second aspect is related to the human enhancement; in spite of the assumption of the safe environment without the possibility of breach of security and privacy of an individual, some people might ask, and rightly so, if society should go that far and really do that.

13. The technology assessment literature generally defines *medical technology* as including not only machinery, devices, and drugs, but also medical practices and procedures.

7.3.3 Laws and Regulations Regarding Wireless Body Implants

To answer consumers' concerns about the privacy and use of their personal information, many applicable consumer protections are already written into law. For example, retailers are already restricted in the sale or distribution of consumer information, and secure computer systems and data encryption schemes are already used for the electronic transfer of private data. There are already federal guidelines in place that address many of these concerns including the Privacy Act (1974), the Electronic Communications Privacy Act (1986), the Telecommunications Act (1996), the Health Insurance Portability and Accountability Act (1996), and the Financial Modernization Act (Gramm-Leach-Bliley Act–2000), among others.

Most international ethics committees were created in the 1990s; UNESCO's International Bioethics Committee (IBC) was created in 1993, and the Steering Committee on Bioethics of the Council of Europe dates from 1992. Most national and international ethics committees have been, in fact, until now committees on bioethics.

The Electronic Privacy Information Center (EPIC) is a public interest research center in Washington, D.C., established in 1994 to focus public attention on emerging civil liberties issues and to protect privacy, freedom of expression and constitutional values in the information age [55].

It seems at the first glance that body implants are not ethically problematic in the context of cardiac pacemakers. However, although particular information and communication technologies (ICT) may be used to repair deficient bodily capabilities, others are ethically more problematic (especially neural implants), particularly if such devices are accessible via digital (and/or wireless) networks. Due to their network capability, ICT could be misused in several ways for all kinds of social surveillance or manipulation.

In December 1997 the European Commission set up the European Group on Ethics and New Technologies (EGE), which is the first international committee with a broader scope, succeeding the Group of Advisers on the Ethical Implications of Biotechnology (GAEIB), which existed from 1991 to 1997. During its first mandate (1998–2000) the EGE provided opinions on subjects as diverse as human tissue banking, human embryo research, personal health data in the information society, doping in sports, and human stem cell research. The group also wrote the *Report on the Charter on Fundamental Rights* related to technological innovation.

On April 24, 2001 the European Commission appointed 12 members for the period 2001 to 2004 and amended the EGE remit in order to strengthen the role of the group. The EGE is an independent, multidisciplinary and pluralist advisory group, composed of 12 members. Its role is to advise the European Commission on how ethical values should be taken into consideration in the regulation of scientific and technological developments.

On March 16, 2005, EGE adopted Opinion No. 20 on the ethical aspects of body implants in the humans and presented it to the European Commission [57]. The idea of placing electronic devices inside our bodies in order not just to repair but even to enhance human capabilities gives rise to science fiction visions with threat and/or benefit characteristics. Not surprisingly, the respect for human dignity has been the fundamental basis of EGE discussions of where the limits should be drawn for different applications of body implants.

Body implants can be used for both medical and nonmedical purposes. Both types of implants clearly require informed consent. This information should not only concern possible benefits and health risks but also risks that such implants could be used to locate people and/or obtain access to information stored in these devices without the permission of the individuals in whom the devices are implanted. The use of body implants in order to obtain remote control over the will of people should be strictly prohibited.

Nevertheless, the ethical notion of the inviolability of the human body should not be understood as a barrier against the advancement of science and technology but as a barrier against its possible misuse. A broad social and political debate is needed to define what kind of applications should be accepted and legally approved, particularly concerning surveillance and enhancement.

The EGE insists that surveillance applications of body implants may only be permitted if there is an urgent and justified necessity and that there are no less intrusive methods. Currently, nonmedical body implants in the human body are not explicitly covered by existing legislation, particularly in terms of privacy and data protection.

In the EGE's view, implantable devices for medical purposes should be regulated in the same way as drugs when the medical goal is the same, particularly as such implants are only partly covered by Council Directive 90/385/EEC on the approximation of the laws of the Member States relating to active implantable medical devices. The EGE recommends that the European Commission should launch legislative initiatives in these areas of body implant applications [57].

The first infoethics goal, derived from the Universal Declaration of Human Rights [UN General Assembly resolution 217 A (III) of December 10, 1948], establishes the fundamental priority of putting technology in the service of human rights. Derived from that goal are three others that aim to promote the public domain, diversity of content, and access to information and the means of communication, with these three based on the premise that all people should be able to share the benefits of body implants [59].

On July 17, 2007, the American Medical Association (AMA) officially established a code of ethics designed to protect patients receiving RFID implants [60]; the code of ethics is not a law but a recommendation defining the essentials of honorable physician behavior. The recommendations focus on safeguarding a patient's privacy and health, and are the result of an evaluation by the AMA's Council on Ethical and Judicial Affairs (CEJA) regarding the medical and ethical implications of RFID chips in humans.

The AMA's report identified three specific recommendations: (1) the requirement for informed-consent process, (2) ensuring patients' privacy by storing confidential information only on secure RFID devices, and (3) requirement for physicians to support research into the safety of RFID devices implanted in human beings and examine the role of doctors regarding the nonmedical uses of the technology. The document unfortunately does not go into any details or specifics of the real-word implementation.

U.S. President George W. Bush created the President's Council on Bioethics on the basis of the Executive Order 13237 from November 28, 2001. The role of the Council is to advise the President on ethical issues related to advances in biomedi-

cal science and technology. They had not published anything on the topic of ethics of implanted medical devices [58].

7.4 Wireless Neural Implants

Neuroscience starts with the study of the brain, spinal cord, and nerves and continues on to incorporate the study of genetics, pharmacology, biochemistry, and other powerful influences on nervous system development and function. This relatively new field of science (started in about 1970) and includes research into neurological and psychiatric disorders and studies of aging, stress, sleep, memory, and movement. Cognitive neuroscience is usually a more controversial branch of neuroscience that focuses on the mind, learning, and human behavior.

The human body reacts to a number of stimuli, both internally and externally. The mechanism to achieve the response is controlled via the nervous system. Impulses travel from the tips of the fingers along nerves to the brain. The signals that travel along the nervous system result from electrical impulses and neurotransmitters that communicate with another body tissue, for example, muscle. For convenience, the nervous system is divided into two sections, which communicate with each other in order to achieve an overall steady state for the body: central and peripheral nervous systems.

The *central nervous system* consists of the brain and the spinal cord and can be thought of as a central processing component of the overall nervous system. The *peripheral nervous system* consists of nerve cells and their fibers that emerge from the brain and spinal cord and communicate with the rest of the body. There are two types of nerve cells within the peripheral system: the afferent or sensory nerves, which carry nerve impulses from the sensory receptors in the body to the central nervous system, and the efferent or motor nerve cells, which convey information away from the central nervous system to the effectors (muscles and body organs.)

The following brief nervous system description and pictures are adopted from [61–63].

7.4.1 The Brain and the Spinal Cord

The center of the nervous system is the brain. It has four major subdivisions: the brain stem, the cerebellum, the cerebrum, and the diencephalon. The location in the brain of these various divisions is depicted in Figure 7.7. Each is concerned with a specific function of the human body.

The *cerebrum* (Latin for brain) is the largest part of the brain. It is composed largely of white matter with a thin outer layer of gray matter, the *cerebral cortex* (surface). It is within the cortex that the higher brain functions of memory, reasoning, and abstract thought occur. The cerebrum is divided into two hemispheres by a deep groove, the longitudinal fissure. Each hemisphere is further divided into lobes with specialized functions.

The *cerebellum* is concerned with coordination for skeletal muscle movement. The cerebellum is under the cerebrum and dorsal to the pons and medulla. Like the cerebrum, it is divided into two hemispheres. It helps to control voluntary muscle movements and to maintain posture, coordination, and balance.

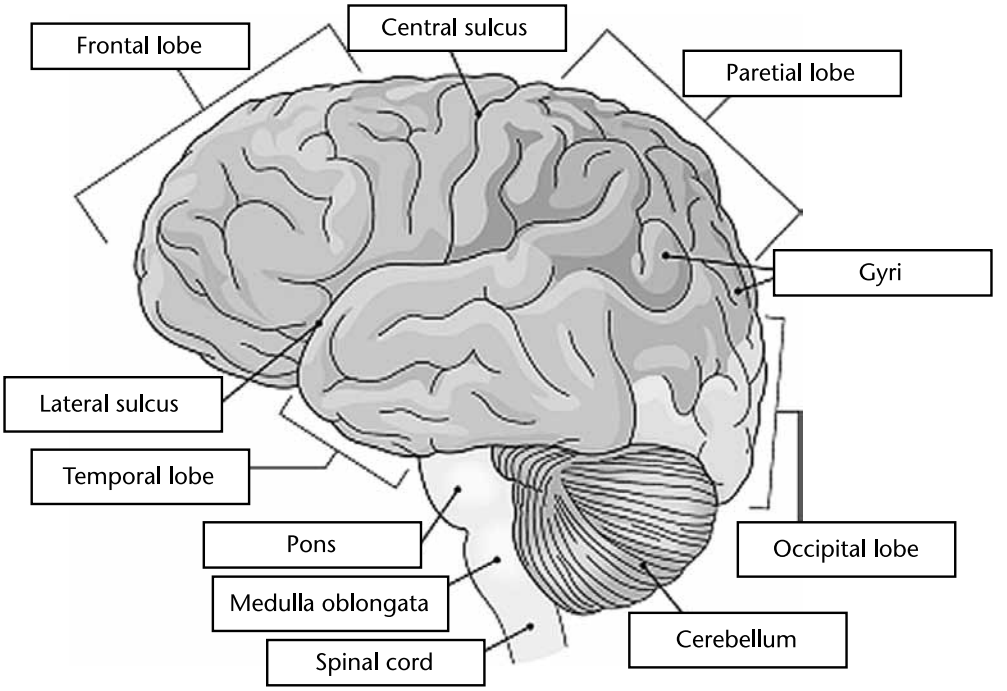


Figure 7.7 The human brain.

The *brain stem* sends messages between the spinal cord and the brain and helps control the heart rate, respiratory rate, swallowing, and blood pressure and is involved with hearing, taste, and other senses. The brainstem consists of the midbrain, the pons, and the medulla oblongata.

The *midbrain* contains reflex centers for improved vision and hearing. The pons forms a bulge on the anterior surface of the brain stem. It contains fibers that connect different regions of the brain. The medulla connects the brain with the spinal cord. All impulses passing to and from the brain travel through this region.

The *diencephalon* connects the midbrain with the cerebral hemispheres. Within its area it has the control of all sensory information, except smell, and relays this information to the cerebrum. Other areas within the diencephalon control the autonomic nervous system regulate body heat, water balance, sleep/wake patterns, food intake, and behavioral responses associated with emotions.

The cerebrum and the diencephalon together constitute the *forebrain*. The frontal lobe is responsible for voluntary movement and planning and is thought to be the most significant lobe for personality and intelligence.

The human brain is mostly water (about 75% in the adult) and has a consistency similar to that of set jelly. The brain is protected by the skull. Within the brain are four ventricles (cavities) in which cerebrospinal fluid (CSF) is produced. This fluid circulates around the brain and spinal cord, acting as a protective cushion for these tissues. Covering the brain and the spinal cord are three protective layers, together called the *meninges*. The outermost and toughest of the three is the *dura mater*. The middle layer is the *arachnoid*. The thin, vascular inner layer, attached directly to the tissue of the brain and spinal cord, is the *pia mater*.

The capillaries within the brain have walls that are highly impermeable and therefore prevent toxic substances causing damage to the brain. Without this protection the delicate neurons could easily be damaged.

The brain is connected to the spinal cord via the brain stem. The spinal cord extends from the skull to the lumbar region of the human back. Twelve pairs of cranial nerves, identified by Roman numerals and also by name, connect with the brain.

Presented in Figure 7.8 is the distribution of the nerves from the spinal cord. Similar to the brain, the spinal cord is surrounded by cerebrospinal fluid. The cord and the cerebrospinal fluid are contained within a ringed sheath called the duramatter. All these structures are contained within the vertebral column.

The vertebral column is made up of individual vertebrae that are separated from each other by annular *intervertebral discs*. These discs have a consistency similar to rubber and act as shock absorbers for the vertebral column. Each vertebra has a canal from which the spinal nerve can leave the spinal column and become a peripheral nerve.

Figure 7.9 illustrates the function of a peripheral nerve. It transmits sensory information to the spinal cord, from which information can either be transmitted to the higher nervous system, the brain, for interpretation and action, or can be acted on directly within the spinal cord and the information sent back down the ventral

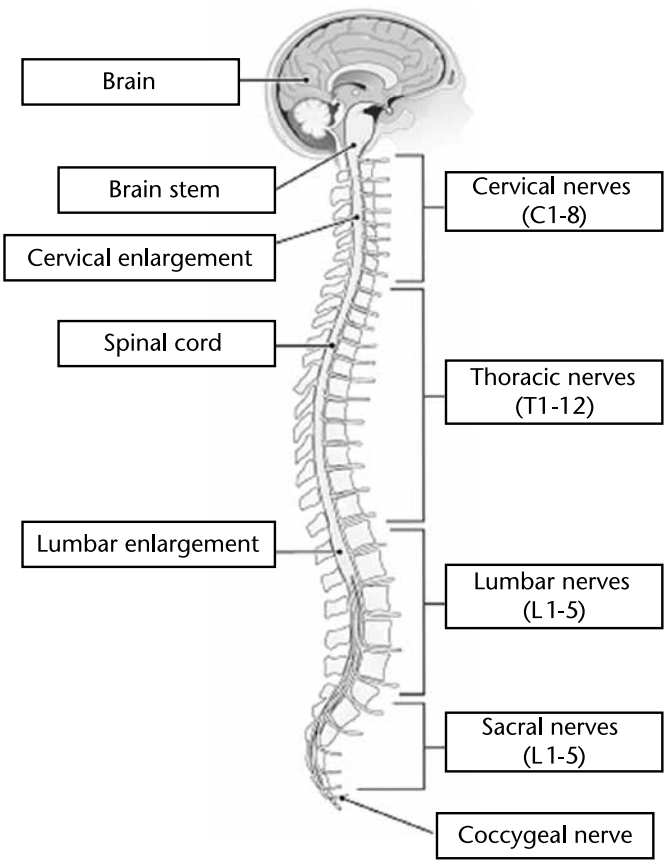


Figure 7.8 The human spinal cord.

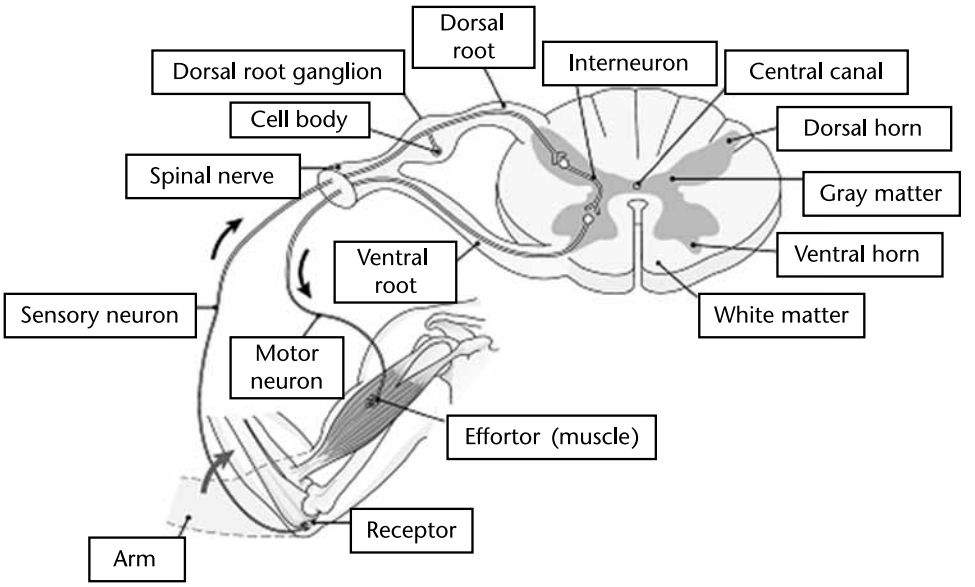


Figure 7.9 The human peripheral nerve.

route to initiate the response. If the spinal cord is injured, the resulting disability is related to the level of the injury.

Injuries of the spinal cord near the brain result in larger loss of function compared to injuries lower down the cord. *Paraplegia* is the loss of motor and sensory functions in the legs. This results if the cord is injured in the thoracic or upper lumbar region. Quadriplegia (or tetraplegia) involves paralysis of all four limbs and occurs from injury at the cervical region. Injuries above the C4 level may require a ventilator or electrical implant for the person to breathe. This is because the diaphragm is controlled by spinal nerves exiting at the upper level of the neck.

Hemiplegia results in the paralysis of the upper and lower limbs on one side of the body, and occurs due to the rupture of an artery within the brain.

Due to the architecture of the connections between the right- and left-hand sides of the brain, damage to the right-hand side of the brain would result in hemiplegia in the opposite side.

7.4.2 The Neurons and the Neurostimulation

The nervous system contains nerve cells, or *neurons;* they are specialized cells that enable the transmission of impulses from one part of the body to another via the central nervous system. Brains of the most advanced insects (honey bees) have about 1 million neurons, snails have about 20,000, and primitive worms (nematodes) have about 300, while 100 billion or so are required for human levels of intelligence.

From the engineering prospective, the human brain can be looked at as a highly distributed, parallel, and hierarchal biological computer, and its neural cells as nature's own transistors [64]. Neurons have two properties: excitability, or the ability to respond to stimuli, and conductivity, the ability to conduct a signal. Neurons

transmit information via electrical pulses [65]; more information can also be found in [66].

Similar to all other body cells, transmission depends upon the difference in potential across the membrane of the cell wall. The change in potential is mediated by transmembrane flow of ions. Electromagnetic fields of different frequencies could affect this system by induced current or direct field effect on molecular interactions. The therapeutic effect of electromagnetic fields has been studied on a variety of neurological diseases, and in these studies extremely low-frequency (ELF) magnetic or electric fields have been used. The application of electromagnetic energy for the treatment of neurological and psychiatric diseases is still in its infancy.

Cognitive science is advancing the understanding of brain functions using imaging techniques, multi-electrode sensing, and neural prosthesis. Multichannel recording of neural activity from the central nervous system (CNS) and peripheral nervous system (PNS) has long been pursued by physiologists as a means of understanding the operation of individual neurons, of deciphering the organization and signal processing techniques of biological neural networks, and of controlling a variety of prosthetic devices.

In September 2010, the National Institutes of Health (NIH) awarded grants totaling $40 million to map the human brain's connections in high resolution [67]: the Human Connectome Project[14].

Brain implants, commonly referred to as *neural implants*, are devices that connect directly to the brain, usually placed on the surface of the brain, or attached to the brain's cortex. Different techniques have been used to produce recording probes for interfacing with the nervous system, using percutaneous (skin-penetrating) connectors for power and data transfer. A common purpose of modern brain implants and the focus of intense research are to establish a biomedical prosthesis bypassing areas in the brain, which became dysfunctional after a stroke or other head injuries.

Neural prostheses interface nerves for therapy and rehabilitation; spinal cord stimulators treat incontinence, cochlea implants restore hearing, vagal nerve stimulators suppress epileptic seizures and depression, and so forth [68, 69].

Neurostimulation is a process by which nerves partially losing their function as a result of disease or trauma are stimulated using artificial electrical pulses for regeneration. Electrical signals used for this purpose must be consistent with the natural activity of human neurophysiology. Some brain implants involve creating interfaces between neural systems and computer chips, which are part of a wider research field called brain-computer interfaces. Brain-computer interface devices detect and translate neural activity into command sequences for computers and prostheses.

Electrical devices called deep brain stimulators (DBS) are essentially a pacemaker for the brain. Since the FDA approved such brain pacemakers and the electrically based treatment they deliver for a disorder called an essential tremor in

14. The Human Connectome Project should provide better understanding of brain connectivity and promises improved diagnosis and treatment of brain disorders. This project will lead to advances in understanding what makes us uniquely human and will help future studies of abnormal brain circuits in many neurological and psychiatric disorders.

1997, for Parkinson's disease in 2002, and for a degenerative brain disease called dystonia[15] in 2003, over 75,000 people have had them installed [70].

Implanted electrical stimulators were first used in 1967. They were primarily developed for the management of chronic pain. In the case of persistent and extensive pain, especially neurogenic pain that does not generally respond to drugs, transcutaneous[16] electrical nerve stimulation (TENS) is not adequate due to the need for multiple electrode placement and increased skin impedance. In order more effectively to cover the painful area, direct stimulation of the spinal cord is necessary via an implantable electrode system. A discussion and a summary on neural implants and direct brain control of prosthetic systems are given in [71].

The brain stimulators are not the cure for the disease, but they can give patients a better quality of life; the beneficial effect has lasted for almost a decade so far in Parkinson's patients, and it is expected the dystonia effect will also be long lasting [72].

In the last few years, scientists have developed a way to stimulate neurons using light rather than electricity. Researchers are developing a prototype neural implant that uses light to alter the behavior of neurons in the brain. The device is based on the emerging science of optogenetic neuromodulation, in which specific brain cells are genetically engineered to respond to light [73]. This effort is still some time away from the implementation on humans.

Used for therapy in cases like these, implantable brain chips are not very controversial and represent desirable procedures. The issues that arise with such therapeutic uses of implantable brain chips primarily involve questions of equity and the costs of implementing this technology.

Although deep brain stimulation is increasingly becoming routine for patients with Parkinson's disease, there may be some behavioral side effects (possibility of apathy, hallucinations, compulsive gambling, hypersexuality, cognitive dysfunction, and depression.) However, these may be temporary and related to the correct placement and calibration of the stimulator and so are potentially reversible.

A disorder such as obesity can be treated by applying an electrical signal to an autonomic nerve (e.g., a vagus nerve[17]). The signal has a duty cycle, including an on time (30 to 180 seconds) during which the signal is applied by to the nerve followed by an off time during which the signal is not applied to the nerve. When applying an electrical signal to a nerve, the signal is commonly a series of pulses applied over a period of time. For example, to treat obesity, a bipolar signal is applied to both the anterior and posterior vagus nerves via electrodes placed on the nerves and connected to a pulse generator.

As shown in U.S. patent application Publication No. US 2005/0038484 A1 published February 17, 2005, the signal may be any signal in excess of a 200-Hz blocking signal reported in [74]. A 5-kHz signal is currently preferred in most cases. The current of the signal is selected to block the nerve without injury to the nerve; amplitudes may range from about 1–6 mA. These signals are applied with a

15. Dystonia is a neurological movement disorder in which sustained muscle contractions cause twisting and repetitive movements or abnormal postures. The disorder may be inherited or caused by other factors such as birth-related or other physical trauma, infection, poisoning (e.g., lead poisoning), or reaction to drugs.
16. Transcutaneous or transdermal means without the skin penetration.
17. The vagus nerve is also called the pneumogastric nerve since it innervates both the lungs and the stomach.

duty cycle, for example, applying a signal for 5 minutes (referred to herein as an on time) followed by 10 minutes of no signal (referred to herein as an off time). The pattern is repeated during the day (for example, while the patient is awake) and repeated for an indefinite number of days (e.g., daily for 6 months, 12 months, or more).

7.4.3 Brain-Computer Interface (BCI)

With the today's evolution of neuroscience, engineering, and computing technology, the era of clinical neuroprosthetics is becoming a practical reality for people with severe motor impairment.

Amyotrophic lateral sclerosis (ALS),[18] brain-stem stroke, and severe brain or spinal cord injury can damage the neural pathways that control muscles or damage the muscles themselves. Individuals most severely affected may lose all voluntary muscle control, including eye movements and respiration, and may be completely locked in to their bodies, unable to communicate in any way [75].

Brain-computer interfaces (BCI)[19] are devices that capture brain signals involved in a subject's intention to act, with the potential to restore communication and movement to those who are immobilized, thus providing an alternative method of communication and control for the most severely affected individuals. Other commonly used terms include: motor neuroprosthetics, direct brain interface (DBI), brain-machine interface (BMI), and neurorobotics. Andrew B. Schwartz in [76] describes the beginnings of BCI in the late 1960s and early 1970s:

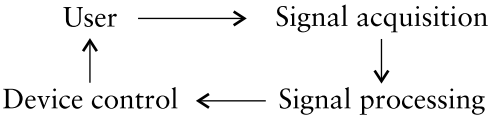
1. *1970s:* Research developed algorithms to reconstruct movements from motor cortex neurons, which control movement.
2. *1980s:* Johns Hopkins researchers found a mathematical relationship between electrical responses of single motor-cortex neurons in rhesus macaque monkeys and the direction that monkeys moved their arms (based on a cosine function).
3. *1990s:* Several groups were able to capture complex brain motor center signals using recordings from neurons to control external devices.
4. *2000s:* Permanent implants became possible due to the advances in electronics (miniaturization) and development of new materials.

BCIs have been researched mainly as aids to patients with severe neuromuscular impairments but could also be used in other applications. Current devices record electrical activity from the scalp, on the surface of the brain, and within the cerebral cortex, and one of the goals of BCI research is to better understand

18. Amyotrophic lateral sclerosis (ALS), also referred to as Lou Gehrig's disease, is a form of motor neuron disease. ALS is a progressive, fatal, neurodegenerative disease caused by the degeneration of motor neurons, the nerve cells in the central nervous system that control voluntary muscle movement. The condition is often called Lou Gehrig's disease in North America, after the famous New York Yankees baseball player who was diagnosed with the disease in 1939.
19. The main difference between BCI and DBS described earlier is that DBS delivers fixed stimulation to activate or inactivate certain pathways in specific brain regions, whereas in BCI the neural prostheses record specific neural activity and respond with appropriate feedback.

the neural coding of information. These signals are being translated to command signals driving prosthetic limbs and computer displays.

A functional BCI system is a closed-loop, real-time system:



The most important component in a BCI is a signal processing decoding algorithm that converts the raw electrophysiological signal into an output that is suitable to control the external device.

A prolonged learning phase may be required to train the subject to encode the desired action into observable changes in his measured neural activity. Therefore, a feedback mechanism and adaptation are important part of the process. As a new output channel, the user must have feedback to improve the performance of how they alter their electrophysiological signals. Continuous alteration of the neuronal output[20] matched against feedback from the overt actions (same for learning to walk, complex movements, and so forth), so the subject's output can thus be tuned to optimize their performance toward the intended goal. The better the subject and computer are able to adapt, the shorter will be the training required for control [77].

EEG arrays allow interface between mind and machine but do not require direct implantation of a device. One of the very first successful attempts was on October 14, 2003, when the Associated Press announced that monkeys with brain implants could consciously move a robot arm with their thoughts. This was achieved by researchers at Duke University, who were hoping to allow paralyzed people to perform similar tasks.

Beyond offering a beneficial therapeutic technique for the patient with a damaged brain, BCI could potentially lead to an augmentation in normal brain performance, significantly extending the brain's power to interact with machines. This prospect is both exciting and frightening, and it certainly demands serious ethical analysis [78].

In 2008, monkeys in the Schwartz Laboratory (named after Andrew Schwarz, a professor of neurobiology) were able to move a robotic arm to feed themselves marshmallows and chunks of fruit while their own arms were restrained. The probes (wide as a human hair) are inserted into neuronal pathways in the monkey's motor cortex, a brain region where voluntary movement originates as electrical impulses. The neurons' collective activity is then evaluated using software programmed with a mathematic algorithm and then sent to the arm, which carries out the actions the monkey intended to perform with its own limb. Movements are fluid and natural, and evidence shows that the monkeys come to regard the robotic device as part of their own bodies.

It is important to realize that the success of patient communication using BCI depends on the existence of an external and established telecommunications and/

20. Neuronal implants are artificial devices that are implanted into the human body and have contacts to nerves or neural tissues. These devices interact with the body by electrical stimulation. Neuronal implants are also termed *neuroprosthetics*.

or wireless infrastructure (i.e., without phone lines, computers, the Internet, e-mail, and simple electrical power), a neural implant would be useless. What makes a prosthetic, monitoring, enhancement application, and/or microchips and biosensors useful is not simply their implantation into the human body, but the fact that they integrate the human body into an external information and telecommunications environment, sort of AmI.

7.4.3.1 Electroencephalography (EEG)

EEG is the safest way of recording brain activity because the electrodes are placed on the scalp (noninvasive procedure). Unfortunately, the human scalp is 20–30 mm (about 1 inch) away from the surface of the cortex (Figure 7.10). The potential from an individual dipole falls off at one over the square of distance, a 300-μV action potential, recorded 0.1 mm away from a neuron, would be reduced to an amplitude of 25 pV when recorded 20 mm away.

Although spikes of individual neurons generate extracellular potentials with frequency components up to 5–10 kHz, the EEG signal on the surface of the brain does not contain significant frequency content above 70 Hz, skull serving as a lowpass filter for the underlying cortical activity. For instance, the large distance between the recording electrode and the underlying cortex allows capacitive effects of the tissue to shunt high frequency currents more locally.

μ (8–12 Hz)[21] and β (18–25 Hz) frequencies are the two dominant bands used in EEG BCI. High-amplitude α waves are associated with a restful, meditative

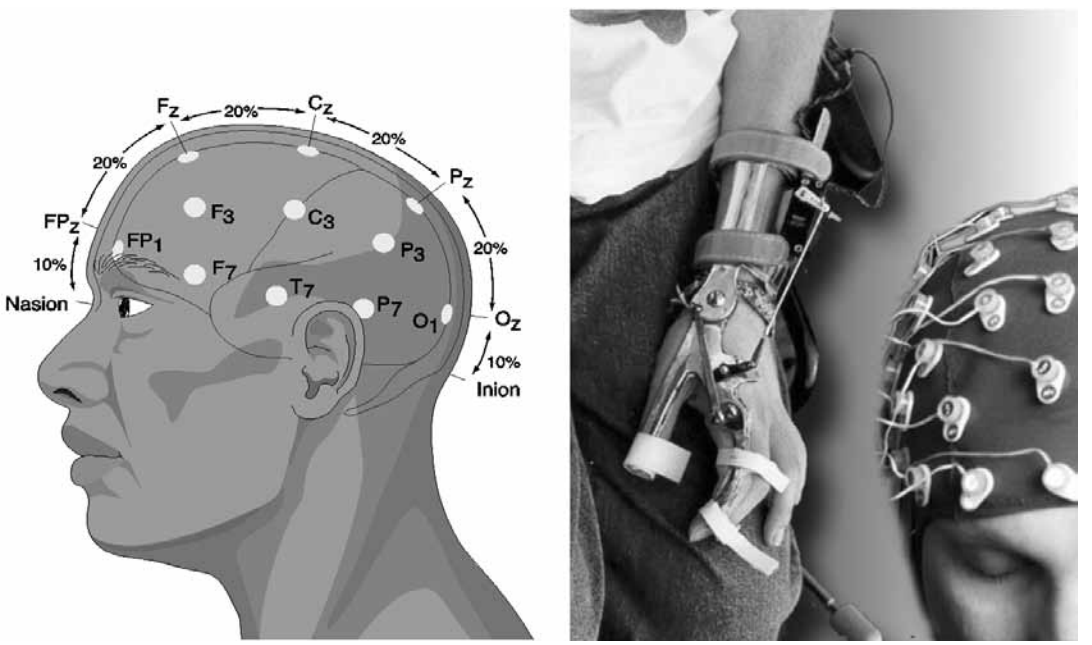


Figure 7.10 EEG electrode placement.

21. Alpha (α) waves are electromagnetic oscillations in the range of 8–12 Hz arising from synchronous and coherent (in phase or constructive) electrical activity of thalamic pacemaker cells in humans. They are also called Berger's wave in memory of the founder of EEG. An alpha-like variant called mu (μ) can be found over the motor cortex (central scalp) that is reduced with movement, or the intention to move.

mind, while low-amplitude β waves with multiple and varying frequencies are often associated with active, busy, or anxious thinking and active concentration.

7.4.3.2 Single-Unit BCI

From the engineering point of view, the optimal method of recording this electrical information would be to place a series of small electrodes directly into the dipole sheet to intercept signals from individual neurons (i.e., single-unit BCI designs). Single unit BCI system consists of a 10×10 array of microelectrodes; the array is attached by cable that transmits signals to a specialized head-mounted, titanium-based connector (Figure 7.11). This high-density microelectrode array is designed to function as a direct cortical interface device and can be implanted in human cortical tissue[22] without acute clinical complications [79].

The ability of a microelectrode to record single-unit action potentials depends on many factors, such as electrode impedance, tip size and shape, whether the target cell has an open or closed extracellular field, and the size and orientation of the target cell bodies in the cerebrum (>100 mm or about 4 inches) and generate large electrical fields, making them an ideal source for extracellular recording. The sig-

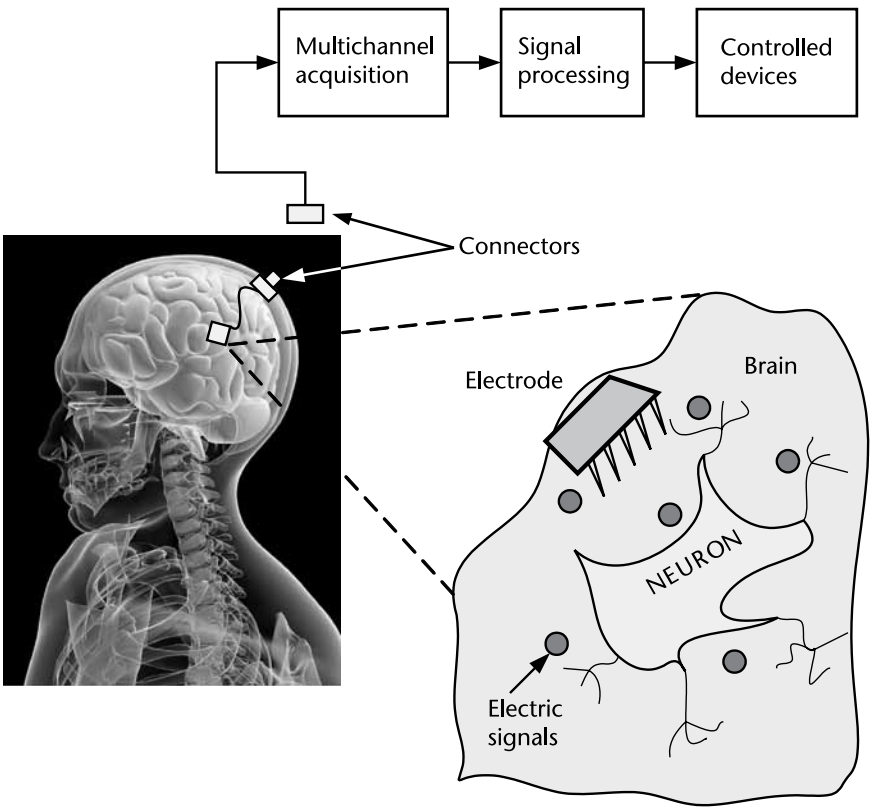


Figure 7.11 Detection and conversion of neural signals.

22. The majority of BCI research in North America involves invasive technologies, while the majority of BCI science in Europe involves noninvasive technologies (due to constraints and intimidations imposed by animal rights organizations).

nals from penetrating microelectrodes used in single-unit recordings are typically bandpass filtered between 300 and 5,000 Hz.

The connector enables cabled access to external electronics which combine low-noise preamplifiers for each channel with their signal multiplexing, perform analog-to-digital conversion, and execute spike analysis and other signal processing tasks. Multichannel neural recording systems potentially produce large quantities of continuously streaming data that must be transmitted while the power dissipation of implanted devices must be very low in order to prevent excessive tissue heating that can kill nearby cells [80].

Neurons rarely fire faster than 100 spikes per second (though rapid bursts of several spikes are possible), with firing rates around 10 Hz somewhat typical in cerebral cortex. Signals from penetrating microelectrodes used in single-unit recordings are usually bandpass filtered between 300 and 5,000 Hz; however, the same electrodes can be used to record lower-frequency (<250 Hz), so called local field potentials (LFPs). This band has advantages because the lower-frequency components seem to be much less affected by geometry and the tissue-electrode interface is so critical for single-unit recordings.

The energy of the robust LFP signals in the primate premotor and motor cortex[23] has been shown to correlate with specific arm movement reach parameters such as direction, distance, and speed, and thus may be useful in neuroprosthetics (also called neuronal implants) applications. In some experiments using electrode arrays, scar tissue forms around microelectrode tips. This scar tissue tends to attenuate spike signals from nearby neurons, but LFP signals seem to be less affected. In many applications, it is desirable to separate LFP and spike signals and analyze them separately.

Based on [81], since the potential from the neurons outside their membrane drops rapidly with distance, a rule of thumb is that any given exposed microelectrode tip should be located within about 30–50 μm from the neuron's cell body within the background of conductive brain tissue to acquire a usefully measurable signal.

Microelectrode arrays composed of bundles of wires have been now largely replaced by monolithic arrays for work in primates (monkeys) and recently in first human trials. In research experiments, electrodes only needed to provide stable neural recordings for enough time to complete the experiment and/or prove the hypothesis. In [82] Ludwig discussed a need to design electrodes for a long-term implantation and realization of the clinically useful neuroprosthetics device.

7.4.3.3 Electrocorticography (ECoG)

ECoG offers a different approach, using a plastic sheet filled with electrodes (Figure 7.12). ECoG is a measure of the electrical activity of the brain taken from beneath the skull (subdural or epidural). The advantage of ECoG-based BCI systems is that recording electrodes are approximated on the cortical surface, yielding a much finer spatial resolution as well as the ability to record higher-frequency (10–200 Hz) content in the signal.

23. The premotor cortex is an area of motor cortex lying within the frontal lobe of the brain.

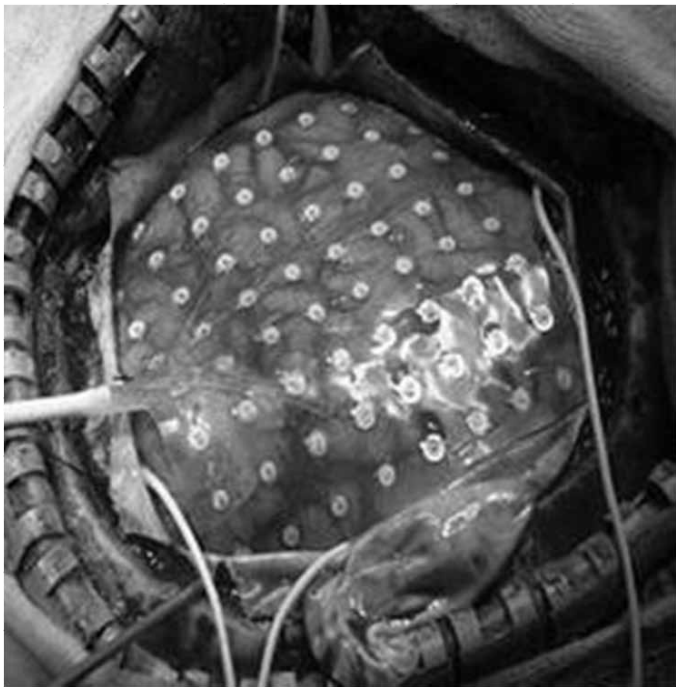


Figure 7.12 ECoG.

ECoG has a much better SNR and higher resolution than traditional EEG recorded with scalp electrodes, and it is less invasive compared to the single-neuron recording technique, where arrays of microelectrodes are inserted into the cortex. The sheet rests on the surface of the brain, recording signals from many neurons at once. The scar tissue does not form around the ECoG grid because it is implanted on the surface of the brain.

Research had shown that the ECoG approach can reveal useful insights into what a patient wants to do by analyzing signals from groups of neurons, rather than single neurons. Examples include a desire to move a hand or to speak [83]. BCIs based on EEG have focused exclusively on μ and β frequencies because γ frequencies are inconspicuous at the scalp. In contrast, γ frequencies as well as μ and β frequencies are prominent in ECoG during movements. The ECoG signal is much more robust compared to the EEG signal and it has five times the magnitude, finer resolution, and higher frequencies.

While ECoG systems are invasive, because they are on the brain surface, they result in stronger signals than penetrating electrodes, but they have not been studied extensively until recently, due to the limited access to subjects. According to Washington University in St. Louis, although the ECoG implants are currently left in place only temporarily, their researchers hope that one day they could be implanted for long-term usage [84].

ECoG electrodes suitable for chronic implants are just now being developed, and they are expected to be tested in nonhuman primates in the near future.

7.4.4 Wireless Neural Implants: Principle of Operation

Early working implants in humans now exist, designed to restore damaged hearing, sight, and movement. The commonality throughout the research is the *cortical plasticity* (also referred to as *neuroplasticity*[24]) of the brain, which often adapts to BCIs.

Currently, most brain implants contain implanted multi-electrode arrays using bundles of fine wires that tether the array to a skull-mounted connector; all electronics for amplification and recording is external to the body. The transcutaneous connector provides a path for infection, external noise and interfering signals easily couple to the wires conveying weak neural signals from high-impedance electrodes, and the connector and external electronics are typically large and bulky compared to the miniature electrode arrays.

These issues present some major barriers to the development of practical neuroprosthetics devices. To eliminate these problems, signals from/to the implanted electrodes should be transmitted out of the body and to the extracutaneous receiver wirelessly, increasing safety and convenience for the patient. The wireless link has to fulfill multiple functions: power transmission and data transmission to and/or from the body. Passive telemetry (similar to load modulation in passive RFID systems) in most cases does not provide sufficient bandwidth, so the most commonly used are inductively coupled RF links or active systems using a transmitter and a receiver operating in MICS or some other frequency band.

One available therapy offering possible hope to a subset of chronic pain[25] patients is Spinal-cord Stimulation (SCS), described in [85]. This therapy consists of electrical impulses triggering selected nerve fibers along the spinal cord. The stimulation of these nerve fibers inhibits pain messages from being transmitted to the brain.

A small ASIC can be programmed to deliver various stimulation patterns (random stimuli, ramping stimuli, combined stimuli, and control of nerve fatigue) to the patient. Current pulses rather than voltage stimuli are delivered so as to be independent of the adjacent tissue impedance (i.e., load). The microprocessor can execute commands that it wirelessly receives from the external (host) computer and create and administer the desired therapy program.

Power for the implant as well as data could be transmitted from the host through an inductive RF link. AM-modulated data is Manchester-coded so as to include also clock synchronization (reference) information for the implant and are serially transmitted. The internal circuitry is powered by the received RF signal picked up by an internal coil and rectified by the AC/DC-voltage converter while the AM demodulator is used for extracting the AM envelope.

24. Neuroplasticity is the capability of brain to act and react in many different and changing circumstances. Because of its plasticity, the brain can rebuild damage from trauma and disease. Healthy brain cells near an injured area of the brain can take on the functions of the damaged part of the brain. Sometimes a brain that has suffered trauma can figure out a new approach by reorganizing preexisting neuronal networks.

25. Usually, the purpose of a pain signal is to act as a warning that protects the body from potential harm. When the cause of the pain has been remedied and no additional injury or healing is occurring, at this point, pain no longer serves the purpose of warning, so it itself becomes the disease that needs treatment. Chronic pain can be very persistent and disabling and may not respond to drugs and other standard therapies.

7.4.5 Fully Implantable Wireless Neural Implants

Fully implantable, wireless neural implants still represent a serious biomedical engineering challenge. Once implanted, wireless neural implants communicate with the outside world via the signals transmitted only transcutaneously [i.e., without any skin-penetrating (percutaneous) wires or feed-through connectors]. These implants are body-embedded and brain-interfaced microsystems where all neural sensor arrays and all the active microelectronic circuits are sealed within the human body.

A team of researchers at the Stanford University created a fully implantable neural recording system that was used on freely behaving animals [86]. The communications frequency was around 4 GHz, with no interference issues from satellites[26]. In addition, frequency was high enough to enable a design of small and efficient antennas while providing a high bandwidth with increased throughput.

Figure 7.13 depicts the design discussed in [87]; the outside unit provides power and clocking through inductively coupled RF unit while the data from the implant is optically transmitted through the skin. The data from the implant could also be transmitted via either an inductively coupled link or even an MICS band.

Once the digitized neural signals from the brain of a health subject are extracted, decoding algorithms and filters correlate the rates of spike activity recorded across the microelectrode array to the observed motion such as using a joystick or

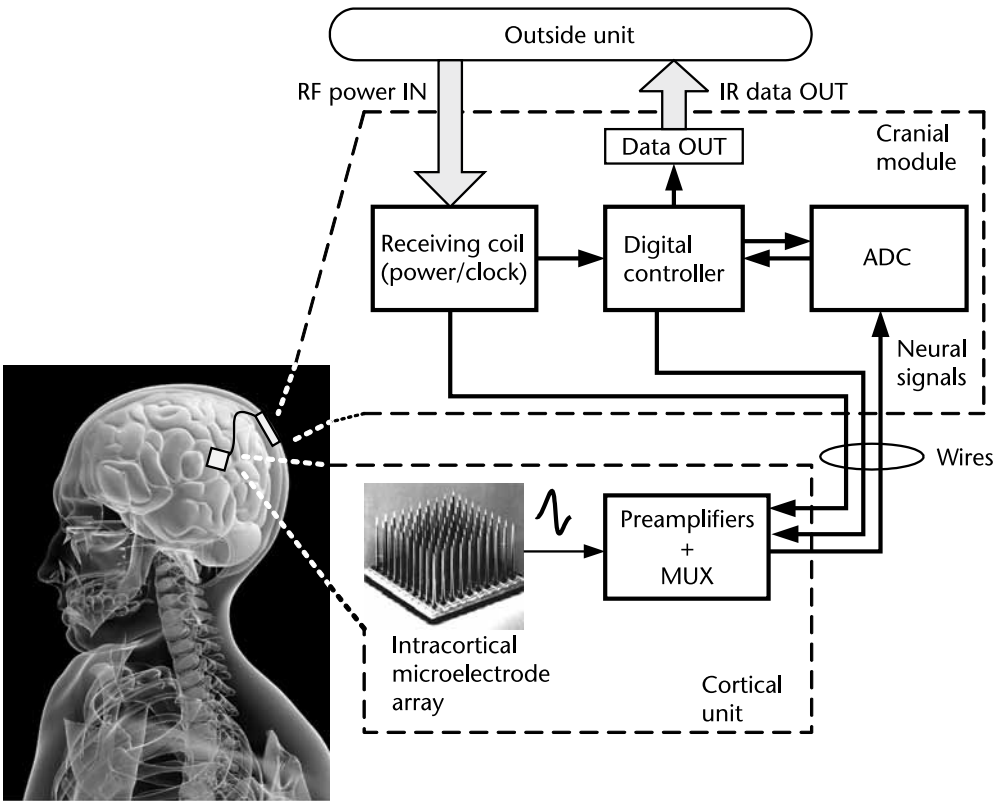


Figure 7.13 Fully implantable neural implants.

26. Geostationary communications satellites have a potential for causing interference into terrestrial radio systems since present INTELSAT satellites as well as a number of U.S. domestic satellites transmit a downlink frequency in the 4-GHz common carrier band.

mouse to move a cursor on computer screen. This information will later be used in order for the brain signals to control desired motion of the paralyzed patients, for example.

There are primarily two methods for retrieving data from an implant: passive impedance reflection and active transmission [88].

Passive transmission is also referred to as *load modulation*, where changes in the loading of the implanted secondary coil are reflected back as a change in the impedance of the primary coil (outside the body), as described in more detail in Chapter 1. External decoding circuitry can sense the loading changes to detect the transmitted data.

In *active transmission*, the implant circuitry drives an implanted antenna to actively transmit the signals to the external receiver, where the carrier is modulated in amplitude, frequency, or phase. Active transmission consumes more power, but achieves higher data rates and greater range than passive transmission.

7.5 Patient's Risks

There are numbers of potential risks associated with the neural implants, some more obvious than others, so risk-benefits analysis is sometimes required to assess whether suggested procedure should be undertaken.

This analysis is somewhat subjective since there are no universal performance criteria for the implanted neural or any other type of medical devices; different patients will have different levels of tolerance for the idea of undergoing an unproven and, in many cases, experimental procedure. In addition, in most cases, it may be difficult to quantify the risks of these proposed procedures.

7.5.1 Surgical Risks

Implanting foreign objects into various parts of the body all carry certain risks, and brain implants are no exception [89]. Chronic subdural hematoma, seizure, infections, subcortical hemorrhages, pulmonary embolisms, potential for a long-term damage to brain tissue, and even perioperative deaths[27] are just a few examples of things that could go wrong. Another important consideration here is a need of neural implants for maintenance, replacement, and upgrades.

Neurons can survive in an environment with a temperature of 30°–40°C; an increased temperature for an extended period of time can be harmful and could eventually cause the death of brain tissue. Neural implants, in order to avoid excessive heating of the brain tissue, should produce a power density below 62 mW/cm^2 [90].

Clinical research is different from the medical treatment, and physicians as well as patients must understand the difference. Clinical research, by definition, investigates a clinical intervention/procedure involving humans in order to obtain scientific knowledge that may or may not benefit the subject. The new field of *neuroethics,* in conjunction with institutional review boards, ensures that human trials are closely monitored for ethical and legal compliance.

27. The perioperative period is the time period describing the duration of a patient's surgical procedure; this commonly includes ward admission, anesthesia, surgery, and recovery.

Today the use of unwilling research subjects is ethically (and legally) unacceptable. What is required is an informed consent, i.e., voluntary participation by a subject after a clear dialogue regarding the nature of the intervention, the risks and benefits, and any other alternatives.

If the risks that could harm the patient are suspected, physicians must quite often balance beneficence and nonmaleficence[28]. Treatment of any kind should not cause any contraindications such as unwanted side effects and/or additional complications. One of the common statements, which should be investigated further, is that implanted RF devices could cause a growth of a tumor and/or cancer in a human body. It should be emphasized that at the time of this writing, there were no reliable scientific studies and/or conclusions concerning the long-term health impact of implants in the human body.

7.5.2 Security and Privacy Risks

Forty years ago, in the preinformation society, local social norms such as simple norms of decency and strong physical (walls) and temporal (limitation of human memory) boundaries were sufficient in maintaining the free and democratic characteristics of society. Today technology creates new privacy issues, for example, known and even accepted surveillance, collection of nonintimate information, and collection of information in public. As a result, society has to deepen its understanding of traditional concerns regarding privacy in order to respond to these new situations [91].

People value a sense of freedom and privacy[29]. While there are variations to its formal definition, privacy can be defined as freedom from the intrusion of others in one's private (personal) life or affairs. Moreover, many view privacy as a right and a legal and absolute standard that is one of our inherent civil liberties; privacy as a legal right is actually a recent creation, starting in the mid-1960s [92].

Information privacy is the interest that individuals have in controlling, or at least significantly influencing, the handling of data about themselves. *Communicational privacy* is the enjoyment of a certain level of intimacy when one communicates with others, even in the public space, as well as a guarantee of some confidentiality of the content of one's communications with others. Privacy can be achieved at home, protected from undesired and unconsented intrusions by others, but home does not provide protection against ubiquitous and pervasive computing, which has the potential to interfere with people's spatial privacy.

Legislations have been passed to protect the information privacy of individuals; one such act is the Privacy Act, Public Law 93-579 (1974), which requires the U.S. government to safeguard personal data processed by federal agency computer

28. Nonmaleficence is a physicians' duty to do no harm. This includes avoiding even the risk of harm so anyone knowingly or unknowingly subjecting a patient or colleague to unnecessary risk has violated the principle of nonmaleficence.

29. The right to privacy is explicit in the European human rights framework, but not in the United States; the right to privacy has no explicit written constitutional basis except in the context of government intrusions, through the constitutional protection against unreasonable searches and seizures of the Fourth Amendment to the U.S. Constitution.

systems and provide ways for individuals to find out what information is being recorded on them and the means to correct inaccuracies.

Today the concept of privacy has been fused with data protection, which interprets privacy in terms of management of personal information. It seems that because of technical advancements it is difficult for important elements, security, and privacy to completely coexist in certain environments.

Medical devices such as oximeters, defibrillators, pacemakers, patient monitors, and even neural implants are commonly equipped with wireless capabilities; interoperability between different systems has been studied for some time now [93]. The safety of the patient from the technical perspective is defined with the security of the communication channel between the neural implant and the outside unit or another implant/prosthesis in or on the patient's body.

Controlling electromechanical systems (prosthetic limbs), deep brain stimulation, and cognitive function augmentation/enhancement via neural signals are new and developing field of biomedical engineering. *Neurosecurity* is a relatively new term describing the protection of the confidentiality (learning private information), integrity (change the settings and modify software), and availability (disabling or causing the malfunction) of neural implants from malicious hackers, attackers, or adversaries [94].

Today's wireless technology used for neural implants requires the transmitter and the receiver to be in a very close proximity, from distance of a few millimeters (1/8 inch) to a few meters (10 feet). In addition, the processing capabilities of the electronic circuitry are limited and so is the potential damage to the patient's safety and privacy. Eventually, the wireless communication channel between the neural implant and the outside unit will allow for the adjustment of the prosthesis, deep brain stimulation, or enhancement of the cognitive functions.

The Internet started as a great idea of connecting people around the world, and while engineers were involved in making things work better and faster, increasing the bandwidth (i.e., transmission capacity and adding new features), others were involved in trying to maliciously attack and penetrate the networks. These attacks could be sometimes triggered by the financial gains, while in some cases a pure savagery or inappropriate application of the one's ingenuity may be the reason.

Similar situations will happen with wireless implants in the future. Although the concerns regarding the wireless neural implants today is probably not a top priority, over the next 10 to 20 years they could become critical. Malicious attacks could change the therapy, disable implants, and cause brain damage by flooding neurons with meaningless random signals, or record (and steal) sensitive and personal information of the patient. The *dignity principle* prohibits the transformation of the human body into an object manipulated and/or controlled remotely.

Aside from the malicious attacks, the users themselves may want to increase or modify the settings of their neural implants as well. The wireless neural implants will have to be designed not only to the highest standards of safety and effectiveness, but also designed ethically and to be capable of operating in different and sometimes hostile environments.

7.5.3 Ethical Issues

The next logical step in this development of the implantable brain chip is *direct neural interfacing*. Today a considerable research is being devoted to neural (brain) implants or a direct brain-machine interface; there could be serious implications and complications of directly communicating with computers or connecting to the Internet, or even a direct brain-to-brain interaction [95].

The implantable chips could generate, among other possibilities, an increased range of senses, enabling, for example, seeing infrared light, ultraviolet light, and chemical spectra; enhancing memory; enabling invisible communication with others when making decisions; and facilitating access to information where and when it is needed. These enhancements will produce major improvements in quality of life or in job performance, but a number of technical, ethical, and social concerns should be considered before proceeding with implantable chips.

According to the Danish Board of Technology [96], the brain is the place where experiences are stored and memories are retrieved, the place where ideas arise and thoughts are born, where decisions are made, and pleasure, pain, grief, and joy are felt. Hence, any increase in our understanding of the brain is bound to trigger important moral, ethical, legal, and socioeconomic questions. This knowledge results in the new methods of treatment, but it also raises some big questions such as the definition of normal versus abnormal, healthy versus ill, and even issues as basic as the definition of human life.

One of the most obvious and basic problems involving the safety and evaluation of the costs and benefits of these implants requires a consideration of the surgical and long-term risks. In addition, the issue of whether there should be a higher standard for safety when technologies are used for enhancement rather than therapy needs public debate. Due to the enormous potential for societal impact, the informed consent of recipients may not be sufficient for allowing implementation.

In addition to the functional assessment and medical usefulness, ethical evaluation of implantable devices is required to assess, at the minimum, the following areas of concern: issues of safety and informed consent, issues of manufacturing and scientific responsibility, anxieties about the psychological impacts of enhancing human nature, worries about possible usage in children, and issues of privacy and autonomy.

Ethics and ethical decision-making are a highly abstract and intangible area in human behavior; ethical decisions that are strongly supported by one or more of the ethical principles without any contradiction or challenge from others may be regarded as very strong and well founded. However, decision-makers will encounter circumstances in which it is impossible to reconcile all the applicable principles and choosing between principles may be required.

A decision or course of action does not necessarily become unethical just because it is contentious or other decision-makers would have reached different conclusions in similar circumstances. A decision-maker's obligation is to consider all the relevant circumstances with as much care as is reasonably possible and to be appropriately accountable for decisions made. As is the case in the evaluation of any future technology, the reliable prediction of all the effects is not possible; nevertheless, the potential for harm must be considered through risk analysis.

Probably one of the most useful ethical principles in everyday life is the Golden Rule: Do unto others as you would have them do unto you. The Golden Rule (also called the *ethics of reciprocity*) does not replace usual moral norms but extends them as a consistency principle and tests people's moral coherence (i.e., the spirit of fairness). The Golden Rule is universal and can be found in Christianity, Confucianism, Buddhism, Hinduism, Islam, Judaism, Taoism, and Zoroastrianism.

Regardless of where the new technologies will go, the most important factor from the patient's perspective is an *informed consent*. Patients should not be subjected to technological risk until they have clearly understood the risk and have granted their consent without being unduly constrained by economic (health insurance issue) or other external pressures. The contention is that the concept of free and informed consent as applied in the field of medicine is applicable to technology in general and ought to be a part of what guides morally grounded public policy.

Whether it is an implant, medication, or another procedure in question, the patient should be a part of a decision-making process when it comes to deciding what will be the course of action. In order to make that decision, a patient has to be informed of all the immediate and future potential benefits and harm.

7.6 Review Questions and Problems

1. What would be some of the most important technical considerations for engineers designing medical implants of any type? How about wireless implants?
2. Systems for monitoring hip implants have been under development for some time now. One system has been researched for the detection of hip prosthesis loosening and using vibration analysis. Special signal processing hardware allows the measurement of small amplitudes in the presence of noisy sensor signals. Try to conceptually describe what would a system like that look like.
3. User satisfaction usually includes *usability* (also called *user testing*) and *usefulness*. When designing any type of human-made object, the ease with which this object can be used is very important. *Usability* includes the testing in real situations and the study of the principles behind an object's perceived efficiency or elegance and is often associated with the functionalities of the product. In this context, the learning curve (i.e., how fast user can get used to the device and master its operation) is also a part of usability.

   Usefulness determines how useful the product is, so the product could have a high usability (great user satisfaction, simple to operate, easy to learn) but very low usefulness since it is a very simple device without too many advanced (and useful) features.

   User satisfaction concerns how sometimes technical achievements in medicine are celebrated in the media (transplantation medicine, prosthetics) while the problems that the individual patient is experiencing are not discussed. A very important question for patients is how the quality of their life will be affected by using these new applications.

Think of a medical device of your choice. Describe how you would test that medical device in practice. What constitutes usability? What constitutes usefulness? Which one (usability or usefulness) would you test first? Why? Which one is more important? Please elaborate your answer.

4. MICS and other wireless communication can be accomplished at the substantial distance and without the assistance of another person so that a certain level of independence could be achieved. However, there is always a potential danger that the wireless communication between the patient's brain and the outside unit could be intercepted, manipulated, and/or altered in some way, so many ethicists are concerned that the technology has the potential to be used in nonmedical ways to affect the human mind and behavior. Discuss the issue.
5. Access to medical records and health care information as well as treatment modification can be a result of the malicious attack by a third party, but it can also be *self-administered* (i.e., the users themselves may want to increase or modify the settings of their neural implants or use medications outside of the boundaries prescribed by the medical staff).

   Discuss the possibility of these self-administered treatment modifications and their long-term consequences on a person's health and well-being. Discuss how these potentially dangerous treatment modifications are impacted by the security mechanisms built into the implanted device(s) and its communications channel.
6. In some developed countries it is a common practice to give volunteers (usually terminally ill patients) experimental medications[30], which cannot be tested on human subjects otherwise. This is for patients who have nothing to lose and is their last chance to get better while providing valuable information to researchers in medicine. Of course, this is done with the full consent of the patient or patient's family. Please discuss this practice.
7. Research and explain why the metals are generally less biocompatible than ceramics or polymers. What can be done to improve this disadvantage of metals as implant materials?
8. Some people, for one reason or another, refuse to accept a surgery (or even medications) as a solution for their medical problems. They usually turn to alternative medicine for help. In Western culture, alternative medicine is any healing practice that does not fall within the realm of conventional medicine or that which has not been shown consistently to be effective.

   Alternative medicine may include herbal medicine, acupuncture, homeopathy, chiropractic medicine, hypnosis, meditation, and prayer. Although alternative medicine most likely will not help the patient to walk again, it may alleviate some of the physical and maybe even some of the psychological pain and depression.

30. Experimental medication, also called *investigational agent* or *investigational drug*, is a substance that has been tested in a laboratory and has gotten approval from the FDA to be tested on people. An experimental drug may be approved by the FDA for use in one disease or condition but could be considered investigational in other diseases or conditions.

For example, the placebo effect[31] can have clinically important effect in some cases and it is not to be neglected in many of the alternative medicine treatments.

Please discuss the importance of placebo effect in helping to improve patient's health. Can alternative medicine replace or augment common medical practice? Do you have any personal experience?

9. The next step from RFID systems is AmI, a new field described as a seamless environment of computing, advanced networking technology, and specific interfaces. AmI will have networking technology embedded in everyday objects such as furniture, clothes, vehicles, appliances, roads, and smart materials. This environment should be aware of the specific characteristics and the needs of users and be capable of responding intelligently to spoken or gestured indications of desire and possibly even result in systems that are capable of engaging in intelligent dialogue. AmI should also be unobtrusive and simple in implementation and usage [97].

   AmI is based on three key, fairly new, technologies: ubiquitous computing, ubiquitous communication, and intelligent user interfaces. By providing an intelligent environment, an innovative intelligent personal health services can be developed while improving the quality and cost control at the same time [98].

   The WSN and BAN are the necessary technology for the development of the concept of AmI where users (patients) are provided services depending on their context [99]. Intelligent interfaces can empower people with severe motion impairments that can result from nonprogressive disorders, such as cerebral palsy, or degenerative neurological diseases, such as ALS, multiple sclerosis (MS), or muscular dystrophy (MD) [100].

   Although the application of the AmI vision may lead to a dramatic lowering of costs (reduction of time to diagnosis and time to treatment, outpatient diagnosis and treatment), the risk of dehumanization and depersonalization of the patient should be carefully considered. The problem may exist in a progressive dehumanization and identification of the patient with the collection of his vital parameters. In other words, the risk is that the patient will be progressively disembodied, reduced to the sum of his or her biological and physiological functions.

   A second risk may arise from the possibility for the patient to monitor directly data detected and stored by wearable biometric devices. This capability may contribute to increase awareness of patient's body, but it may also increase the likelihood of self-diagnosis, with potential serious implications for the patient's health [101].

   Generally speaking, people are concerned about the social consequences of a world full of embedded wireless implants, tags and readers. WBANs

31. The phenomenon of an inert substance resulting in a patient's medical improvement is called the placebo effect. The word placebo, Latin for "I shall please," dates back to a Latin translation of the Bible by St. Jerome. It was first used in a medicinal context in the eighteenth century. A substance containing no medication is prescribed or given to reinforce a patient's expectation to get well; positive results are called placebo effects and negative effects are called nocebo effects.

are a part of AmI; monitoring and tracking implanted devices outside of the designed purpose are a critical issue, as they provide details about the actual person. Society may need laws to specify who can access personal data logs and for what purpose. In Europe, the Data Protection Act already limits access to computer records of this kind, and the United States should probably enact similar legislation.

What are your thoughts on the application of AmI for medical purposes? Should people be concerned with unexpected and undesired side effects in the attempt to improve human lives using the latest technology? Should we even continue with the development of the AmI? Write a short essay and provide arguments that support your answer.

10. Fabrication of a very, very small RFID device with communicating ability is already achievable, and it appears technically possible to fabricate RFID device at the biomolecular scale that interrogates its local chemical environment (sensing) or/and controls the biological machinery inside a single living cell [102]. The remote monitoring at such a very small-scale level is assumed to provide the selectivity requirement, that is, the local control of biomolecular machinery without damaging cells or even affecting the biological events in the surrounding medium. Discuss the concept of remote monitoring and control (activation/deactivation) of human biological functions wirelessly by using RFID technology inside a single living cell.
11. The assumption that practically anything, including ethical dilemmas, can be described and analyzed mathematically [103]. Discuss at least one approach to the objective mathematical decision-making process that could be used in cases wherer you have intangible data.

References

[1] Kutz, M., *Standard Handbook of Biomedical Engineering and Design*, New York: McGraw-Hill, 2002.

[2] Vo-Dinh, T., and B. Cullum, "Biosensors and Biochips: Advances in Biological and Medical Diagnostics," *Fresenius J. Anal. Chem.*, Vol. 366, 2000, pp. 540–551.

[3] Bilstrup, K., *A Preliminary Study of Wireless Body Area Networks*, Technical Report IDE0854, School of Information Science, Computer and Electrical Engineering, Halmstad University, Sweden, August 2008.

[4] Kailas, A., and M. A. Ingram, "Wireless Communications Technology in Telehealth Systems," School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 2009.

[5] Webster, J. G., *Medical Devices and Instrumentation,* 2nd ed., Vol. 2, *Capacitive Microsensors for Biomedical Applications: Drug Infusion Systems,* New York: John Wiley & Sons, 2006.

[6] http://www.mayoclinic.com/health/capsule-endoscopy/MY00139, last accessed December 2011.

[7] Gordon, N., and U. Sagman, "Nanomedicine Taxonomy," Briefing Paper, Canadian Nano-Business Alliance, February 2003.

[8] Panescu, D., "MEMS in Medicine and Biology," *IEEE Engineering in Medicine and Biology Magazine,* September/October 2006.

[9] Chow, A. Y., et al., "The Artificial Silicon Retina Microchip for the Treatment of Vision Loss from Retinis Pigmentosa," *Archophthalmol.*, Vol. 122, April 2004.

[10] www.mems-issys.com (accessed August 24, 2010).

[11] *Kyllo v. United States*, 533 U.S. 27, 2001.

[12] American Association for the Advancement of Science (AAAS), Annual Report, 2006.

[13] Neuman, M. R., "Physical Measurements," Ch. 46 in *Medical Devices and Systems*, J. D. Bronzino, (ed.), Boca Raton, FL: CRC Press, 2006.

[14] Kui-Jui Huang, R., "Flexible Neural Implants," Ph.D. Thesis, California Institute of Technology, 2010.

[15] Bronzino, J. D., (ed.), *The Biomedical Engineering Handbook,* 2nd ed., Boca Raton, FL: CRC Press, 2000.

[16] FDA, "Medical Device Tracking; Guidance for Industry and FDA Staff," August 15, 2008.

[17] Batchelor, A., and M. Chandrasekaran, (eds.), *Service Characteristics of Biomedical Materials and Implants,* Series on Biomaterials and Bioengineering, Vol. 3, Imperial College Press, London, U.K.: World Scientific Publishing, 2004.

[18] http://www.prescriptiondrug-info.com/topics/deca/ (accessed July 18, 2010).

[19] http://www.ieee802.org/15/pub/TG6.html (accessed July 17, 2010).

[20] Yuce, M. R., and C. K. Ho, "Implementation of Body Area Networks Based on MICS/WMTS Medical Bands for Healthcare Systems," The School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, Australia, 2008.

[21] Taparugssanagorn, A., et al., "A Review of Channel Modelling for Wireless Body Area Network in Wireless Medical Communications," Centre for Wireless Communications, University of Oulu, Finland, 2009.

[22] ISO Technical Report TR 21730:2007, *Health Informatics: Use of Mobile Wireless Communications and Computing Technology in Healthcare Facilities, Recommendations for Electromagnetic Compatibility (Management of Unintentional Electromagnetic Interference) with Medical Devices*, 2007.

[23] Quwaider, M., et al., "Body-Posture-Based Dynamic Link Power Control in Wearable Sensor Networks," *IEEE Communications Magazine,* July 2010.

[24] Sun, M., et al., "Data Communication Between Brain Implants and Computer," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* Vol. 11, No. 2, June 2003.

[25] Hall, P. S., and Y. Hao, "Antennas and Propagation for Body-Centric Communications," Norwood, MA: Artech House, 2006.

[26] Scanlon, W. G., J. B. Burns, and N. E. Evans, "Radiowave Propagation from a Tissue-Implanted Source at 418 MHz and 916.5 MHz," *IEEE Transactions on Biomedical Engineering,* April 2000, pp. 527–534.

[27] Karlsson, A., "Physical Limitations of Antennas in a Lossy Medium," Department of Electroscience Electromagnetic Theory, Lund Institute of Technology, Sweden, 2003.

[28] Alomainy, A., et al., "Modelling and Characterisation of Radio Propagation from Wireless Implants at Different Frequencies," Department of Electronic Engineering, Queen Mary, University of London, 2005.

[29] Hartsgrove, G., A. Kraszewski, and A. Surowiec, "Simulated Biological Materials for Electromagnetic Radiation Absorption Studies," *Bioelectromagnetics,* No. 8, 1987, pp. 29–365.

[30] Okoniewski, M., and M. A. Stuchly, "A Study of the Handset Antenna and Human Body Interaction," *IEEE Transactions on Microwave Theory and Techniques,* Vol. 44, No. 10, 1996, pp. 1855–1864.

[31] Scanlon, W. G., and N. E. Evans, "Radiowave Propagation from a Tissue-Implanted Source at 418MHz and 916.5MHz," *IEEE Transactions on Biomedical Engineering*, Vol. 47, No. 4, 2000, pp. 527–534.

[32] Chirwa, L. C., et al., "Electromagnetic Radiation from Ingested Sources in the Human Intestine Between 150 MHz and 1.2 GHz," *IEEE Transactions on Biomedical Engineering,* Vol. 50, No. 4, 2003, pp. 484–492.

[33] Gabriel, C., and S. Gabriel, "Compilation of the Dielectric Properties of Body Tissues at RF and Microwave Frequencies," Armstrong Laboratory, 1996, http://niremf.ifac.cnr.it/tissprop/htmlclie/htmlclie.htm (accessed March 16, 2009).

[34] Schroeder, M. J., et al., "An Analysis on the Role of Water Content and State on Effective Permittivity Using Mixing Formulas," *Journal of Biomechanics, Biomedical and Biophysical Engineering,* Vol. 2, No. 1, 2008.

[35] Abdelsayed, S. M., et al., "Radiation Characteristics of Loop Antennas for Biomedical Implants," McMaster University, Department of Electrical and Computer Engineering, Canada, 2006.

[36] Falcon, C., "Improving Data Transfer and Battery Life for Implanted Devices," originally published by MDDI, June 2005.

[37] Loreto Mateu Saez, M., "Energy Harvesting from Passive Human Power," Ph.D. Thesis Project, Electronic Engineering, January 2004.

[38] Holleman, J., et al., "Neural WISP: An Energy-Harvesting Wireless Neural Interface with 1-m Range," University of Washington, 2009.

[39] Kwon, D., and G. Rincon-Mora, "A 2μm BiCMOS Rectifier-Free AC-DC Piezoelectric Energy Harvester-Charger IC," *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 4, No. 6, December 2010.

[40] Jiang, B., et al., " Energy Scavenging for Inductively Coupled Passive RFID Tags," *IMTC 2005, Instrumentation and Measurement Technology Conference*, Ottawa, Canada, May 2005.

[41] Hamici, Z., R. Itti, and J. Champier, "A High-Efficiency Power and Data Transmission System for Biomedical Implantable Electronic Devices," *Meas. Sci. Technol.*, No. 7, 1996, pp. 192–201.

[42] Goto, K., et al., "An Implantable Power Supply with an Optically Rechargeable Lithium Battery," *IEEE Transactions on Biomedical Engineering*, Vol. 48, No. 7, 2001, pp. 830–833.

[43] Al-Ashmouny, K. M., et al., "IBCOM (Intra-Brain Communication) Microsystem: Wireless Transmission of Neural Signals Within the Brain," *31st Annual International Conference of the IEEE EMBS,* Minneapolis, MN, September 2–6, 2009.

[44] Cabeza, R., and A. Kingstone, *Handbook of Functional Neuroimaging of Cognition*, 2nd ed., Cambridge, MA: MIT Press, 2006.

[45] Bizzi, E., et al., *Using Imaging to Identify Deceit: Scientific and Ethical Questions*, American Academy of Arts and Sciences, 2009.

[46] Guidance for Industry and FDA Premarket and Design Control Reviewers Medical Device Use-Safety, R. Kaye and J. Crowley, "Incorporating Human Factors Engineering into Risk Management," U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Division of Device User Programs and Systems Analysis, Office of Health and Industry Programs, July 18, 2000.

[47] Richter, C. K., "Harmonizing Regulations and Standards That Guide Clinical Investigation of Medical Devices," Food and Drug Administration, Center for Devices and Radiological Health, 2006.

[48] Teixeira, M. B., and R. Bradley, *Design Controls for the Medical Device Industry*, New York: Marcel Dekker, 2003.

[49] http://www.fda.gov/cdrh/devadvice/3122.html#link_2 (accessed August 24, 2010).

[50] http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/HowtoMarketYourDevice/default.htm (accessed May 13, 2010).

[51] HAS-Haute Autorité de Santé (French National Authority of Health) Guidebook, *Medical Device Assessment in France,* 2009.

[52] http://ec.europa.eu/enterprise/sectors/medical-devices/regulatory-framework/legislation/index_en.htm (accessed September 9, 2010).

[53] Gelijns, A. C., and H. V. Dawkins, (eds.), Committee on Technological Innovation in Medicine, "Adopting a New Medical Technology," Institute of Medicine, 1994.

[54] Elhauge, E., "The Limited Regulatory Potential of Medical Technology Assessment," Harvard Law School, 1996.

[55] The Electronic Privacy Information Center, *EPIC Annual Report, 2007–2008*.

[56] Opinion No. 20, "Ethical Aspects of ICT Implants in the Human Body," Opinion of the European Group on Ethics in Science and New Technologies to the European Commission, March 16, 2005.

[57] http://europa.eu.int/comm/european_group_ethics/index_en.htm (accessed August 23, 2010).

[58] http://www.bioethics.gov, last accessed July 2010.

[59] Rundle, M., and C. Conley, "Ethical Implications of Emerging Technologies: A Survey," *Geneva Net Dialogue*, UNESCO, Paris, 2007.

[60] Report of the Council on Ethical and Judicial Affairs, CEJA Report 5-A-07, *Radio Frequency ID Devices in Humans,* presented by R. M. Sade, MD, Chair, 2007.

[61] Jennings, D., et al., *Introduction to Medical Electronics Applications*, London, U.K.: Edward Arnold, 1995.

[62] Cohen, B. J., *Medical Terminology: An Illustrated Guide*, 4th ed., Baltimore, MD: Lippicott Williams and Wilkins, 2007.

[63] Kandel, E. R., J. H. Schwartz, and T. M. Jessell, (eds.), *Principles of Neural Science*, 4th ed., New York: McGraw-Hill, 2000.

[64] Nurmikko, A. V., et al., "Listening to Brain Microcircuits for Interfacing with External World—Progress in Wireless Implantable Microelectronic Neuroengineering Devices," *Proceedings of the IEEE,* Vol. 98, No. 3, March 2010.

[65] Hobbie, R. K., and B. J. Roth, *Intermediate Physics for Medicine and Biology*, 4th ed., New York: Springer Science+Business Media, 2007.

[66] http://www.medical-look.com/human_anatomy/ (accessed July 21, 2010).

[67] http://www.nih.gov/news/health/sep2010/nimh-15.htm (accessed December 18, 2010).

[68] Stieglitz, T., "Challenges of Neuroprosthetics and Sensor Technologies for Rehabilitation," *European Symposium: Technical Aids for Rehabilitation (TAR 2007)*, Technical University of Berlin, January 25–26, 2007

[69] Stieglitz, T., et al., "Implantable Biomedical Microsystems for Neural Prostheses," *IEEE Engineering in Medicine and Biology Magazine*, September/October 2005.

[70] Leach, J. B., et al., "Bridging the Divide Between Neuroprosthetic Design, Tissue Engineering and Neurobiology," *Frontiers in Neuroengineering*, Vol. 2, No. 18, February 2010.

[71] Ryu, S. I., and K. V. Shenoy, "Human Cortical Prostheses: Lost in Translation?" *Neurosurgical Focus*, Vol. 27, July 2009.

[72] http://www.ninds.nih.gov/disorders/dystonias/detail_dystonias.htm (accessed August 24, 2010).

[73] Williams, M., "A Brain Implant That Uses Light," *Technology Review*, February 24, 2010.

[74] Solomonow, M., et al., "Control of Muscle Contractile Force Through Indirect High-Frequency Stimulation," *Am. J. of Physical Medicine*, Vol. 62, No. 2, 1983, pp. 71-82.

[75] McFarland, D. J., "Noninvasive Communication Systems," Chapter 7, *WTEC Panel Report on International Assessment of Research and Development in Brain-Computer Interfaces*, 2007.

[76] Schwartz, A. B., et al., "Brain-Controlled Interfaces: Movement Restoration with Neural Prosthetics," *Neuron*, Vol. 52, October 5, 2006, pp. 205–220.

[77] Sanguineti, V., et al., "Neuro-Engineering: from Neural Interfaces to Biological Computers," *Communications Through Virtual Technology: Identity Community and Technology in the Internet Age,* G. Riva and F. Davide (eds), ISO Press: Amsterdam, 2001.

[78] O'Shea, M., *The Brain: A Very Short Introduction*, Oxford, U.K.: Oxford University Press, 2005.

[79] House, P. A., et al., "Acute Microelectrode Array Implantation into Human Neocortex: Preliminary Technique and Histological Considerations," *Neurosurgical Focus*, Vol. 20, No. 5, May 2006, p. E4.

[80] Harrison, R. R., "The Design of Integrated Circuits to Observe Brain Activity," *Proceedings of the IEEE*, Vol. 96, No. 7, July 2008.

[81] Gold, C., et al., "On the Origin of the Extracellular Action Potential Waveform: A Modeling Study," *J. Neurophysiology*, Vol. 95, 2006, pp. 3113–3128.

[82] Kip A. Ludwig, "Neuroprostethic Devices: Inputs and Outputs," Ph.D. Dissertation, University of Michigan, 2009.

[83] Ojemann, J. G., et al, "Brain-Machine Interface: Restoring Neurological Function through Bioengineering," *Clinical Neurosurgery*, Vol. 54, 2007, pp. 134–135.

[84] WUSTL Newsroom, "Brain Implants May Help Stroke Patients Overcome Partial Paralysis," November 19, 2008.

[85] Mouine, J., K. A. Ammar, and Z. Chtourou, "A Completely Programmable and Very Flexible Implantable Pain Controller," *Proceedings of the 22nd Annual EMBS International Conference*, Chicago IL, July 2000, pp. 603-622.

[86] Miranda, H., et al., "A High-Rate Long-range Wireless Transmission System for Multichannel Neural Recording Applications," *IEEE ISCA*, 2009.

[87] Song, Y. -K., et al., "Active Microelectronic Neurosensor Arrays for Implantable Brain Communication Interfaces," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 17, No. 4, August 2009.

[88] Wise, K. D., et al., "Wireless Implantable Microsystems: High-Density Electronic Interfaces to the Nervous System," *Proceedings of the IEEE*, Vol. 92, No. 1, January 2004.

[89] ISO 14708-3, "Implants for Surgery: Active Implantable Medical Devices Part 3: Implantable Neurostimulators," November 15, 2008.

[90] Reichert, W., "Indwelling Neural Implants: Strategies for Contending with the In-Vivo Environment," in *BMI-Related Thermal Studies*, Boca Raton, FL: CRC, 2007.

[91] Austin, L., "Privacy and the Question of Technology," *Law and Philosophy*, Vol. 22, 2003, p. 164.

[92] Cohen, A., and C. H. Wellman, (eds.), *Contemporary Debates in Applied Ethics*, Cambridge, MA: Blackwell Publishing, 2005.

[93] Venkatasubramanian, K.K. et al., "Interoperable Medical Devices: Communication Security Issues," *IEEE Pulse*, September/October 2010.

[94] Denning, T. et al., "Neurosecurity: Security and Privacy for Neural Devices," *Neurosurgery Focus*, Vol. 27, July 2009.

[95] Naam, R., *More Than Human*, New York: Broadway Books, 2005.

[96] Newsletter from the Danish Board of Technology to the Danish Parliament, "Knowledge of the Brain Must Be Used with Care," No. 213, December 2005.

[97] Rouvroy, A., "Privacy, Data Protection, and the Unprecedented Challenges of Ambient Intelligence," *Studies in Ethics, Law, and Technology*, Vol. 2, No. 1, Article 3, The Berkeley Electronic Press, 2008, http://www.bepress.com/selt/vol2/iss1/art3 (accessed August 24, 2010).

[98] Chiarugi, F., et al., "Ambient Intelligence Support for Tomorrow's Health Care: Scenario Based Requirements and Architectural Specifications of the EU-DOMAIN Platform," 2006.

[99] Fernández, L., et al., "Wireless Sensor Networks in Ambient Intelligence," *Technologies for Health and Well-Being*, Instituto ITACA, Universidad Politécnica de Valencia, 2007.

[100] Nakashima, H., et al., *Handbook of Ambient Intelligence and Smart Environments*, New York: Springer, 2009.

[101] Gaggioli, A., et al., "From Cyborgs to Cyberbodies: The Evolution of the Concept of Techno-Body in Modern Medicine," *Psychology Journal,* Vol. 1, No. 2, 2003, pp. 75–86.

[102] Aubert, H., "RFID Technology for Human Implant Devices," *Comptes rendus à l'Académie des Sciences*, Special issue on nanosciences/nanotechnologies, March 1, 2011.

[103] Lehpamer, H., "Analysis of Technical and Ethical Acceptability of Wireless Body Implant Applications," Ph.D. dissertation, University of Zagreb, Zagreb, Croatia, 2011.

[104] http://www.mayoclinic.com/health/medical/IMO4443 (accessed December 2011).

CHAPTER 8

Sociocultural Implications of RFIDs and Their Applications

8.1 Market Trends and Usage

Many new and promising RFID applications are in the works. For example, medicinal products can be tagged and traced to combat drug counterfeiting, and logging tagged items into and out of your refrigerator can help you track when certain products are out of stock or whether certain products have gone beyond their expiration date.

An RFID chip can provide useful information over the whole life cycle of its tagged product. That is why RFID chips can improve customer relations through better after-sales services. There is a large field of applications when it comes to medical services and services for people with other types of needs. Even the tracking of criminals on parole from prison is imaginable.

One of the most popular new developments is a contact-less payment solution called SmartPay for small payments in the United States. It is also an efficient technique to reduce thefts from shops, in addition to the stock-keeping function mentioned above. Looking into the future, RFID and smart tags will allow the creation of an Internet of things, where objects and locations may be directly related to one another. These objects will also be capable of increasingly *intelligent interaction*.

Apart from its expected benefits, the more intensive and extensive use of RFID also raises major issues in the areas of privacy, security, technological reliability, and international compatibility. One key challenge for decision-makers is to create a common vision and a set of goals on how RFID can keep companies (and countries) more innovative and competitive in the world economy.

At the same time, citizens must have the tools and freedom of choice they need to protect their privacy and security. At present, technological challenges like the lack of global frequency standards, low reading rates, interference with other radio sources, insufficient encryption capabilities, and the cost of implementation and end-user concerns prevent wide adoption of RFID technology. Thus, the main issues to be addressed are consumer privacy, standards and interoperability, harmonization of the frequency spectrum, intellectual property rights, and future research needs.

What will be the social consequences of a world full of embedded RFID tags and readers? Will our privacy be affected as RFID technology makes it possible for our movements to be tracked and allows our personal information to be available in unprecedented detail? These and many other questions must be answered before RFID systems become commonplace. One of the major worries for privacy advocates is that RFID tags identifying individual items purchased with credit or debit cards would link buyers to the specific items in the card's or the store's databases. Marketers could then use these data to keep track of exactly what particular people bought, down to the color, size, style, and price, a lot more information than UPC bar codes reveal.

Another concern is that RFID equipment will produce automatic audit trails of commercial transactions. In a totally tagged world, it will be easier to detect when we lie about how we spent our time or what we did and where. This capability could have great consequences for the workplace, and the legal system might look to using logs kept by tag readers as courtroom evidence. We may need laws to specify who can access data logs and for what purpose. In Europe, the Data Protection Act 1998[1] already limits access to computer records of this kind, and the United States will probably enact similar legislation.

The problem does not lie with RFID technologies themselves; it is the way in which they are deployed that raise privacy concerns. Privacy and security must be built in from the outset (i.e., at the design stage); just as privacy concerns must be identified in a broad and systemic manner, so too must technological solutions be addressed systemically. A thorough privacy impact assessment is critical. Users of RFID technologies and information systems should address the privacy and security issues early in the design stages, with a particular emphasis on data minimization. This means that wherever possible, efforts should be made to minimize the chance to identify, observe, and link the RFID tags with personal information and other associated data.

Use of RFID information systems should be open and transparent and offer individuals as much opportunity as possible to participate and make informed decisions.

8.1.1 Barriers to RFID Adoption

National and international regulatory authorities are trying to bring at least some regional interoperability and, in some cases, international interoperability, but progress is slow and competition for valuable bandwidth is strong. There are currently two principal barriers to global implementations for low powered tags: the regulations and the laws of physics.

The saying "high speed, high frequency; high capability, high cost" still applies. With these constraints, RFID systems can now generally be grouped into two types, although with some exceptions:

- Low-cost, but capable passive tag systems;

1. The 1998 act replaced and consolidated earlier legislation such as the Data Protection Act 1984 and the Access to Personal Files Act 1987. Later, the Privacy and Electronic Communications (EC Directive) Regulations 2003 altered the consent requirement for most electronic marketing.

- High-cost specialized tags for operation at high speed and long distances.

RFID supporters envision a world where RFID reader devices are everywhere: in stores, cars, clothes, factories, and even in our home refrigerators. However, RFID tags will not become ubiquitous in consumer products as long as the price of the tags is viewed as prohibitively expensive by many businesses. RFID tags currently cost from 20 cents to $1 each, which still makes them impractical for identifying millions of items that cost only a few dollars.

Some experts predict that in quantities of 1 billion, RFID tags would approach 10 cents each. The holy grail of 5-cent tags, which is the stated primary goal of the Auto-ID Center, would be attained in lots of 10 billion. More recent technological developments may put a 1-cent tag within reach, which, in turn, would fuel demand for RFID comparable to that for bar codes.

Makers of RFID chips and so-called inlays, which include the chip, antenna, and substrate, have been trying for years to reduce prices for RFID tags to 5 cents. The types of materials and assembly methods used to package tags impact the final cost directly (around 30%) and to some extent the communication performance. In the supply chain, the cost of tags is one of the main considerations for mass adoption, with the 5-cent tag being the much-discussed target.

Traditionally, chip die size has always been the key focus, and IC companies have managed to get die sizes (chip area) down to around 0.3 mm^2 for UHF chips, resulting in a manufacturing cost of about 1 to 2 cents depending on the silicon process, leaving 3 cents for the rest of the cost. This is where the real challenge now seems to be.

However, these less expensive tags may lack the capabilities of their more expensive counterparts. As a result, most manufacturers have been content to cut prices at a steady rate of about 5% to 10% per year since 2000 while improving the technology. As a result, users of the tags are employing them in applications no one dreamed of a decade ago, despite their inability to reach the elusive nickel price.

At McCarran International Airport in Las Vegas, for example, operators attach bag tags with dual dipole antennas to luggage to ensure that RFID readers in the handling system can communicate with all bags, regardless of their orientation on conveyor belts. The technology integrates two antennas 90° from one another; thus, the RFID tags can communicate with the airport's RFID readers, no matter how baggage handlers toss the luggage onto conveyor belts. Such dual-antenna tags have not reached rock-bottom prices, but at roughly 20 cents each, they offer capabilities that nickel tags cannot match.

Similarly, retailers have begun using tags with specialized antennas to enable garments buried in stacks to successfully communicate with RFID readers. Again, cheaper tags are unlikely to achieve such feats.

There seems to be a lot of attention given to the cost of tags. The benefits that can be derived from implementing RFID can far outweigh the cost of the tags. An investment in time and money is required; however, to be successful and get the best return on investment (ROI), it is important to understand the technology and how it can benefit the organization. In other words, by understanding the technology, it is possible to optimize the design and reduce overall cost of implementation.

8.1.2 Globalization

Widespread deployment of RFID relies on availability of either dedicated or license-exempt bands. Current interest is in the UHF frequencies, which offer a good balance between antenna size and path loss. However, the requirements for these bands vary widely around the world, frustrating attempts to deploy systems in an era of global trade.

Frequency band is just one of the challenges; given often-conflicting global constraints, with the implied requirement to recognize tags wherever goods might flow, the challenge is to build in support for multiple data rates, modulation formats, and interference environments through a flexible and programmable air interface. From the technical prospective, in order to make RFID a truly global technology, some basic requirements have to be fulfilled:

- *Compatibility:* Suitability of products, processes, or services for use together under specific conditions to fulfill relevant requirements without causing unacceptable interactions. Interchangeability, interoperability, and noninterference are differing levels (or degrees) of compatibility.
- *Interchangeability:* The condition that exists between devices or systems that exhibit equivalent functionality, interface features, and performance to allow one to be exchanged for another, without alteration, and achieve the same operational service.
- *Interoperability:* The condition that exists between systems, from different vendors, to execute bidirectional data exchange functions, in a manner that allows them to operate effectively together. A guarantee of a certain level of compatibility has to be achieved between different implementations of the same standard. The desired level of compatibility is specific to a given standard and can be limited to basic services. Interconnection and interoperability are the main objectives of standardization.
- *Noninterference:* The condition that exists when standard-compliant components of various types or of different vendor origins coexist within the same space without serious detrimental effect on one another's performance. Components are not necessarily required to communicate with one another as part of a common infrastructure, but merely to peacefully coexist.

8.2 RFID Security and Privacy Aspects

8.2.1 Access to Information

If you are lost in an airport or parking lot, an RFID-based system that can guide you to your gate or car would be appealing. So too would be the ability to return items to shops without receipts, either for refunds or warranty servicing, and RFID-enhanced medicine cabinets that ensure that you have remembered to take your medications. The concern is the effect on individual privacy of RFID-enabled computing systems that can automatically see everyday objects: the clothing on your

person, the medical implants in your body, the prescription drugs you are carrying, the payment devices in your pocket, and perhaps even individual pieces of paper, such as banknotes and airline tickets [1] (see Figure 8.1).

To increase consumer acceptance of RFID technology, RFID advocates must promote and implement comprehensive security measures, along with consumer education, enforcement guidelines, and research and development of practical security technologies. Technical organizations, such as EPCglobal, Inc., are developing standards for the electronic product code, including its *Guidelines on EPC for Consumer Products*.

It is useful to understand the difference between *on-tag* and *off-tag access control*. As the name implies, on-tag access control mechanisms are located on the RFID tags themselves. On-tag access control is the most common type of RFID access control, with mechanisms including tag deactivation, cryptography, and tag-reader authentication.

In contrast, off-tag access control mechanisms put the access control mechanism on a device external to the RFID tag. Examples of this include the *RSA blocker tag*[2], a special RFID tag designed to prevent readers from performing unwanted scanning and tracking of people or goods, without any disruption to normal RFID operation, and external reencryption. Off-tag access control has the advantage that it can protect low-cost RFID tags (such as EPC tags), because the access control does not require any extra complexity (hence, extra cost) on the RFID tag itself.

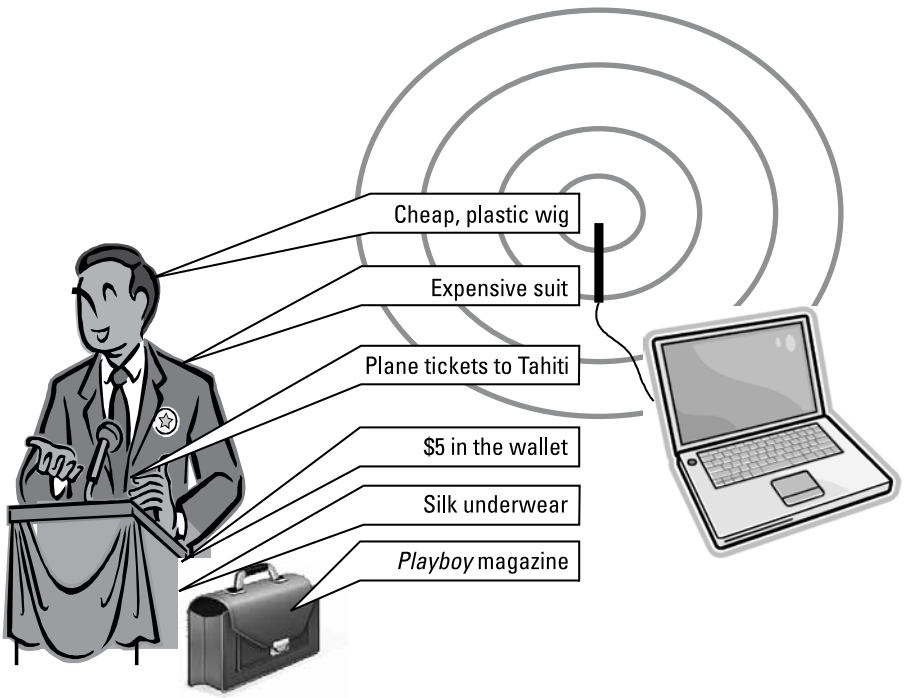


Figure 8.1 Potential RFID privacy threats.

2. RSA Laboratories is the research center of RSA, the Security Division of EMC, and the security research group within the EMC Innovation Network. The group was established in 1991 at RSA Data Security, the company founded by the inventors of the RSA public-key cryptosystem.

8.2.2 Privacy Threats and Protection

The impending ubiquity of RFID tags requires not only support mechanisms to provide adequate performance, but also measures to address privacy concerns associated with unobtrusive tags on everyday items. When people envisioned computing capabilities everywhere, embedded in the environment in such a way that they can be used without being noticed, they also acknowledged that the invisible nature of the computing devices will make it difficult to know what is controlling what, what is connected to what, and where information is flowing.

This tension between the contradicting requirements of control and privacy on the one hand and usability and performance on the other are well illustrated by the privacy concerns associated with the planned deployment of RFID technology in supermarkets and retail outlets. Two notable privacy issues complicate the adoption of RFID systems:

1. *Leaking information pertaining to personal property:* If a generic dumb RFID system is used, anyone can read, without restriction, the connection between the product and the tag and obtain information regarding the tagged contents of, say, a purse or any tagged item worn on the body in a manner about which the possessor is unaware.
2. *Tracking the consumer's spending history and patterns and physical whereabouts:* If a product ID is specific to an individual (when, say, tags are used in clothes and other personal belongings such as shoes, watches, handbags, and jewelry), tracking the person's movements over an extended period becomes an option. Not only can physical location be tracked, but an individual's personal information (stored on multiple independently managed databases) might also be accessible based on a unique ID.

These RFID privacy threats follow from the basic functionality of RFID technology that states that an ID can be read without permission, is constant and unique, and contains potentially sensitive data. A number of proposed RFID privacy protection schemes are classified based on the new functionality that they implement in RFID technology; they range from adding only memory to adding lightweight circuits. Each involves a trade-off between the cost of the tag and the value of privacy protection. Several approaches are briefly discussed next.

8.2.2.1 Kill Function

The EPCglobal standard specifies that tags must be equipped with at least one nullification function as a way to address public opposition. This function, called the *kill command*, disables the functionality of the tag after consumers purchase a product. It involves a high degree of consumer privacy protection at negligible cost; however, since the disabling process is performed manually by millions of individual consumers, human error is always a possibility. Moreover, the major problem in killing the tag is that the various RFID stakeholders would no longer be able to take advantage of the future emerging services that would rely on the millions of RFID tags likely to be dispersed throughout the consumer environment.

This simple countermeasure, a built-in option designed to kill the functionality of an RFID tag when the consumer leaves the store, has been incorporated into the

EPCglobal standard (Class 1 Generation 2 UHF Air Interface Protocol). For consumers, its purpose is easy to understand and thus easy to accept. However, killing a tag's functionality curtails the future potential use of RFID in consumer services, such as in smart refrigerators that automatically reorder food products, expiration date and product recall alarms, and personal library management.

8.2.2.2 Normal Tags and Smart Tags

Other privacy protection schemes generally reflect two main approaches: normal tags and smart tags.

The *normal tag* approach protects individual consumer privacy without having to modify the existing tag or cost the user organization more money. The normal tag approach achieves privacy protection by preventing the unauthorized reading of the output from the tag, blocking electric waves with aluminum foil, or jamming waves to interfere with a tag's ID being read by an adversary's unauthenticated reader.

Smart tags are equipped with additional components, such as rewritable memory, basic logic circuits, hash function units (turning data into a relatively small number that may serve as a digital fingerprint of the data), and common-key/public-key encryption units. When the tag incorporates rewritable memory, the reader rewrites the information in the tag to achieve privacy protection. This approach is notable for its low cost, because the tag requires only rewritable memory.

On the other hand, a lightweight circuit is incorporated into the tag, and a re-encrypted ID to the reader is calculated by the circuit. Although public key cryptosystems come close to providing good privacy protection, they are not suitable for tags because public key primitives are complex and costly. A noteworthy scheme employing this technology is the *hash-chain scheme*, in which a hash function circuit is embedded in the tag, and the tag response is calculated by the hash function. The scheme holds down the cost of the tag, because the hash function is lightweight, pseudorandom, and one way.

Here, *pseudorandom* means the output of the hash function is computationally indistinguishable from a true random value. Being *one way* means it is computationally infeasible to compute the input of the hash function from output of the hash function. The scheme addresses ID leakage and tracing problems through the pseudorandomness of the hash function, which prevents leakage and tracing. Moreover, the scheme is forward secure; that is, even after the tag's secret is exposed through tampering, the tag's past history cannot be traced due to the hash function being only one way.

The drawback to the hash-chain scheme is that the load on the server is proportional to the number of tags, though the load can be reduced through advanced computation.

8.2.3 The Blocker Tag

The RFID blocker tag takes a different approach to enhancing RFID privacy. It involves no modification to consumer tags. Rather, the blocker tag creates an RF environment that is hostile to RFID readers. The blocker tag is a specially configured, ancillary RFID tag that prevents unauthorized scanning of consumer items.

In a nutshell, the blocker tag *spams* misbehaving readers so they cannot locate the protected tags' identifiers. At the same time, it permits authorized scanners to proceed normally [2].

The blocker tag spoofs the tree-walking protocol into thinking that all tags, that is, all identifiers, are present. To do this, it simply emits both a 0 and a 1 in response to all reader queries. The result is that the reader attempts to traverse the entire identifier tree, believing that all possible tag identifiers in the world are present. The reader stalls because the tree is far too big to be fully scanned (for Class 1 EPC tags, the tree would have 2^{96} nodes).

8.2.4 Reader Signal Energy Analysis

One approach to RFID privacy does not rely on logical protocols at all. A system has been proposed, based on the premise that legitimate readers are likely to be quite close to tags (such as at a checkout counter), whereas malicious readers are likely to be far away (such as a competitor in the parking lot). In other words, the farther away a reader is, the greater the noise level in the signal that a tag receives. With some additional circuitry, therefore, an RFID tag might be able to obtain a rough estimate of the querying reader's distance and change its behavior accordingly.

A tag interacting with a distant reader might only reveal the type of product to which it is attached, a pair of trousers, for example. When interacting with a nearby reader, however, the tag might also reveal its unique identifier. A more sophisticated, multitiered approach is also possible, in which tags furnish increasing amounts of information as readers get closer.

Of course, distance alone does not provide an ideal trust metric. However, distance could be combined with traditional access control techniques, such as a challenge-response protocol between the reader and tag, to achieve a more comprehensive approach to RFID tag privacy. Indeed, the distance-measurement approach is complementary to blocker tags.

8.2.5 Protecting the Public

The Federal Trade Commission held a hearing in June 2004 to "facilitate discussion of the public policy issues surrounding the use of RFID and to encourage the development of best practices for RFID that do not compromise consumers' privacy and security." Their March 2005 released report [3] found that the privacy issues associated with RFID are linked to database security and that industry can play an important role in addressing privacy concerns raised by some RFID applications. The report emphasized the importance of industry self-regulatory programs, meaningful accountability provisions to help ensure compliance, and implementation of reasonable and appropriate measures to protect data collected by RFID systems.

Legislators in several states, recognizing privacy concerns stemming from the use of RFID, have introduced bills that seek to respond to the increasingly rapid adoption of RFID technology. While none of the proposed pieces of legislation has been passed into law, the introduction of these bills signifies that RFID-related technologies appears to be generating concerns within the legislative branches of state and federal governments.

According to some, legislation restricting RFID use at this early stage would likely stifle the technology and delay deployment in the marketplace[3]. It would be more productive to monitor the technology over the next few years, while engaging with the business and government sectors regarding their respective use of RFID and their policies on maintaining RFID privacy and security.

A number of governments around the world, trying to ensure the protection of individual privacy rights, have raised concerns that the collection, storage, transfer, and use of personal information through RFID technology could possibly violate individuals' privacy rights. For example, in the European Union, the EU Article 29 Working Party of Member State Data Protection Authority has recently expressed its concern that RFID technology may contravene the requirements of the EU Directive on Data Protection. Accordingly, the European Commission has held a number of workshops and issued inquiries concerning the privacy implications of RFID.

The Asia-Pacific Economic Cooperation (APEC) forum is considering the relationship of RFID privacy to its recent privacy guidelines. In particular, South Korea has called for the development of RFID privacy guidelines in the forum's Electronic Commerce Steering Group. In 2004, Japan also issued privacy guidelines for RFID. Finally, the Organization for Economic Cooperation and Development's (OECD) Working Party on Information Security and Privacy is currently reviewing the scope of policies and concerns with the global use of RFID on security.

The U.S. State Department is implementing its conversion program for the RFID-based electronic passports, or *e-passports*, despite warnings from security experts that these passports could be accessed or tracked by the wrong individuals. In fact, some security experts feel that the technology contained in this type of passport could be used by terrorists to construct a bomb designed to target anyone of their choosing.

There is also some concern that e-passports do not have enough security embedded to resist hackers and the advancement of technology. Nevertheless, in August 2006, the State Department began issuing e-passports[4] containing RFID chips, and at the time of this writing, the plan was that all U.S. passports were expected to include RFID chips containing personal biometric information by 2017.

8.2.6 Fair Information Practices

The Fair Information Practices (FIP), published by the OECD in 1980, are a well-established set of guidelines for consumer privacy (www.oecd.org). They have their roots in a 1973 report by the U.S. Department for Health, Education, and Welfare (HEW) and were drawn up by the OECD to better facilitate the cross-border transfer of customer information as part of trade between its member states. The eight principles can be summarized as follows:

3. The Competitive Enterprise Institute (a nonprofit public policy organization advocating nonregulatory, market-based solutions) states that as RFID technology comes into full use, various social forces would constrain it more suitably than government regulation.
4. New U.S. e-passports contain a 64-kbit RFID chip with personal information about the passport holder. U.S. Department of Homeland Security (DHS) officials claim that the passports must be held within 10 centimeters (4 inches) of a reader to have their data read.

1. *Collection limitation:* Data collectors should only collect information that is necessary, and should do so by lawful and fair means, that is, with the knowledge or consent of the data subject.
2. *Data quality:* The collected data should be kept up to date and stored only as long as it is relevant.
3. *Purpose specification:* The purpose for which data is collected should be specified (and announced) ahead of the data collection.
4. *Use limitation:* Personal data should only be used for the stated purpose, except with the data subject's consent or as required by law.
5. *Security safeguards:* Reasonable security safeguards should protect collected data from unauthorized access, use, modification, or disclosure.
6. *Openness:* It should be possible for data subjects to learn about the data collector's identity and how to get in touch with him or her.
7. *Individual participation:* Data subjects should be able to query data collectors as to whether or not their personal information has been stored and, if possible, challenge (i.e., erase, rectify, or amend) this data.
8. *Accountability:* Data collectors should be accountable for complying with these principles.

The FIP forms the basis for many of today's privacy laws, such as the EU Directive 95/46/EC (April 1995), which provides the framework for the national privacy laws of all EU-member states. For example, Article 6 of the Directive requires data collectors to collect only as much information as necessary (also called the *proportionality principle* or the *principle of data minimization*), while Article 7 requires them to obtain the unambiguous consent of the data subject before collection.

It is undisputed that the act of reading out one or more RFID tags can constitute a data collection, meaning that existing privacy laws also apply to the communication between tags and their readers. This has also been recently pointed out by the International Community of Data Protection and Privacy Commissioners; at the outset, this would mean that RFID readers would need to be openly announced with the help of public signs and placards explaining the purpose and extent of the data collection, as well as the identity of the data collector.

More comprehensive unofficial text of the EU Directive 95/46/EC can be found at the Center for Democracy and Technology's Web site[5].

8.3 Health Risks from RFID

Electricity and electromagnetic fields (EMFs) bring countless benefits to society. We cannot live without them, yet we do not know the consequences of long-term exposure to EMFs, if indeed there are any. Therefore, research is needed to understand the risks and set appropriate safety standards.

Scientists researching the health effects of nonionizing electromagnetic fields have to contend with an enormous electromagnetic spectrum, reaching from static electric and magnetic fields to EMFs at frequencies in the terahertz range. They

5. www.cdt.org. The Center for Democracy and Technology is a nonprofit public interest organization.

also have to contend with the complexity of the human organism. The number and nature of the ways in which these two systems interact can at present only be guessed at.

Behind the scenes, there are the dynamic social and political forces at work among an enormous array of stakeholders with different interests in the field: the scientists and research groups themselves and policy makers, politicians, governments, and regulatory bodies, at local, national and international levels. We must not forget health professionals, industrialists, investors, trade associations, trade unions, marketing professionals, users of devices reliant on EMFs, patients, and so on; we should not overlook for a moment the modern media as well.

Today there are hundreds of thousands of RFID scanners and EAS systems in use. All of these systems utilize EMFs to detect and scan tags. According to some sources, RFID systems pose no threat to the health of ordinary people; this might be true but prolonged occupational exposure may occur. The report by The International Commission on Non-Ionizing Radiation Protection (ICNIRP) examines the effects of EM radiation on humans [4]. The report describes mechanisms of thermal and nonthermal interaction between EM fields and biological systems.

Thermal interaction is the heating of tissue which can cause damage. The most notable *nonthermal interaction* is brain stimulation, which means that the membrane potentials may be altered at a cellular level and might have effects on the nervous system. The report states that the high frequencies of EAS/RFID system produce no heating or thermoregulatory stress.

However, EAS and RFID devices may interact with medical devices such as pacemakers, which can cause dangerous situations and indirect health hazard. The report recommends that further studies should be made on this area and that device manufacturers should provide information needed for health risk assessments. There is also a need to continue to collect exposure data, especially for occupational groups. If possible, low-frequency and high-frequency exposure should be differentiated.

It is a well-known fact that strong electromagnetic fields can interfere with electronics. Hospitals contain many devices that are critical, for example, life-support equipment that must not be disrupted. Today many hospitals have banned the use of cellular phones because of the risks of interference. How are RFID applications compared to this? Are they a potential source of interference also?

Since 1993 there has existed an electromagnetic compatibility (EMC) standard for medical devices (MD), the IEC 60601-1-2. This standard, however, was originally only specified for frequencies lower than 1 GHz, and in 2001 it was updated to cover frequencies up to 2.5 GHz. In the standard it was specified that life-supporting equipment must be able to operate in presence of field strengths up to 10 V/m and nonlife-supporting up to 3 V/m [5].

RFID transmitters and similar short-range radio devices are regulated differently in Europe and in the United States. In Europe regulations are country-specific but based on the CEPT regulations. In the United States, radio devices have to conform to the FCC.

Some general types of recommendations are listed here:

- Continue to monitor current research on the health effects of exposure to radio frequencies from RFID equipment to verify whether new information indicates health risks.
- Ensure that any RFID equipment purchased meets the FCC requirements protecting users from any possible thermal effects and includes all possible measures to minimize the risks of emissions from the system interfering with electrically powered active medical devices, such as pacemakers.
- Purchase of RFID equipment that meets the human exposure specifications of nonregulatory agencies such as the ANSI, the IEEE, the NCRP, American Conference of Industrial Hygienists (ACGIH) and the ICNIRP.
- If RFID equipment is installed, conduct measurements of the RF emissions levels generated from systems to insure that the emissions are below those recommended under the OSHA nonionizing standard (CFR 29, Section 1910.97) and by the nonregulatory agencies listed in the previous point.
- If RFID equipment is purchased, ensure that it is installed and maintained according to the manufacturers' specifications.

8.4 Ethical and Moral Dilemmas of Technology

Regardless of whether the discussion is about a prosthetic, monitoring, or enhancement application, what makes microchips and biosensors useful is not simply their implantation into the human body but their integration into an external information and telecommunications environment. In doing so, people can gain greater control over their environments and transform how they interact with each other.

The old question of how and if the law (and government) should intervene to guarantee that technological progress does not result in violations of fundamental rights and fundamental freedoms has been with us for a long time. The answer is undoubtedly tied to the present political, technological, and cultural assumptions, and many issues and dilemmas, beyond just the technical and operational capabilities of any wireless or any other technology must be considered.

Apart from the large number of technical considerations that must be undertaken, there are a few intangible and theoretical considerations such as security, privacy, social, and ethical considerations that have to be discussed as well. Some of them may or may not be acceptable by the today's standards and in today's society, while others could lead to disastrous situations in the future and therefore of interest to the research. A new term *technoethics* was coined to describe an interdisciplinary study of moral and ethical aspects of technology in society [6].

An ethical dilemma is usually defined as "a situation that often involves an apparent conflict between moral imperatives, in which to obey one would result in transgressing another" [7]. This definition recognizes that real ethical dilemmas involve choices among alternatives that have ethical or moral content. This is a multifaceted problem, and finding universally accepted solution will not come easily but the decisions that society makes today will impact future generations and their lifestyle.

It is also critical to realize that technology has no moral value; it is neither good nor evil. Rather, it is the application of the technologies that raises the moral and ethical issues [8].

8.4.1 Basic Concepts of Ethics

Ethics could be defined as a discipline that deals with what is good and bad, as well as with the moral duty and obligation; it can also be regarded as a set of moral principles or values. *Morality* is a doctrine or system of moral conduct. Moral conduct refers to principles of right and wrong in behavior. It is interesting that the words *morale* (Latin origin) and *ethics* (Greek origin) are in fact synonyms [9]. They are used interchangeably to refer to the study of right and wrong behavior.

Concepts of right and wrong are not simple; today they increasingly include the more difficult and subtle questions of fairness, justice, and equity. Two key branches of moral philosophy, or ethics, are descriptive ethics and normative ethics, each taking a different perspective. *Descriptive ethics* is describing, characterizing, and studying the morality of a people, a culture, or a society and compares and contrasts different moral codes, practices, systems, beliefs, and values [10]. Descriptive ethics focuses on the prevailing set of ethical standards in the certain community.

By contrast, *normative ethics* is concerned with supplying and justifying a coherent moral system and seeks to uncover, develop, and justify basic moral principles that are intended to guide behavior, actions, and decisions. Normative ethics, therefore, seeks to propose some principle or principles for distinguishing right from wrong. Normative ethics is concerned with establishing norms or standards by which certain practices might be guided or judged.

It is tempting to observe the prevalence of a particular practice and conclude that because so many are doing it (descriptive ethics), it must be acceptable behavior. Normative ethics demands a more meaningful moral foundation and it would insist that a practice be justified on the basis of some ethical principle, argument, or rationale before being considered acceptable.

Figure 8.2 shows some of the possible influences on a person's conscience and its ethical behavior. Essentially, this diagram demonstrates that the ethical beliefs one holds and how these beliefs are applied are strongly influenced by personality and background. This view assumes that morality depends on a dual consideration of human nature and the human condition [i.e., specific social and cultural circumstances (cultural relativism) defining moral beliefs and practices] [11].

Some of the most important factors that influence ethical behavior are personal behavioral lessons learned as a child; over time, people consolidate these lessons into individual morals, self-imposed standards, and values. These morals, standards, and values also help determine how individuals perceive themselves and how they interact with other members of people in society.

In addition, organizational factors often affected by external forces influence a person's ethical belief system. When people in everyday life are faced with events and situations that go against their ethical foundations, internal conflict can result, and people are not always functioning in a realm of purely right and wrong.

Society may be defined as a community, a nation, or a broad grouping of people having common traditions, values, institutions, and collective activities and interests. The *environment* is a key concept in understanding our relationships.

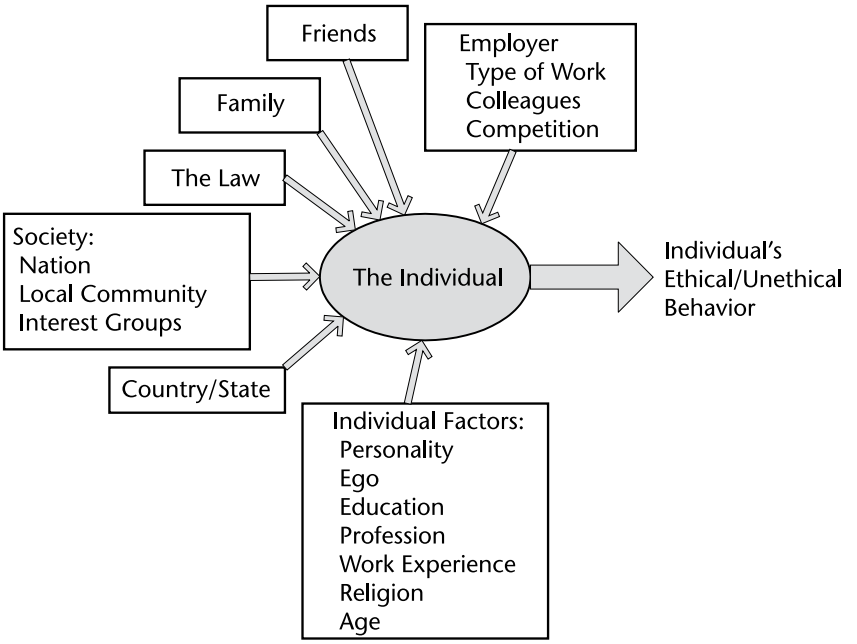


Figure 8.2 Ethical norms and their sources.

The *macroenvironment* is, in a sense, just another way of thinking about society. Some philosophers see the macroenvironment as being composed of four segments: social, economic, political, and technological. The *social* segment (or environment) focuses on demographics, lifestyles, and social values of the society. The *economic* segment focuses on the nature and direction of the economy in which society operates. Variables of interest might include the gross national product, inflation, interest rates, unemployment rates, foreign-exchange fluctuations, and various other aspects of economic activity.

Ethical analysis and decision making occurs at various levels, ranging from the individual to the societal, and even the international. Sociotechnical problems are not solved in the strict sense that purely technical problems are, and quite often a conflict may emerge among choices made at these various levels. Today people realize that advancing the species with information and science technology has great benefits, but at the same time these advancements contradict many individuals' ethical beliefs or feelings.

The *political* segment focuses on the processes by which laws get passed and officials get elected and all other aspects of the interaction between companies, political processes, and government. Of particular interest are the regulatory process and the changes that occur over time in business regulation, various industries, and various issues. Finally, the *technological* segment represents the total set of technology-based advancements or progress taking place in society [12].

Assumption that what is ethical in one country must be ethical in another country may not be necessarily true. Culture, customs, education, language, attitudes, and institutions vary from country to country, and these differences pose sometimes insurmountable obstacles. Ethical issues are difficult when dealing with one culture and once two or possibly more cultures are under consideration, it gets

extremely complicated. It is necessary to deal not only with differing customs, protocol, and ways of operating but also with differing concepts of law or notions of what is acceptable or unacceptable behavior in an ethical sense.

The main concepts of ethics are *responsibility* and *accountability*; responsibility means that as free moral agents, individuals, organizations, and societies are responsible for the actions they take and should be held accountable for the consequences of their actions. The generally accepted view of ethics is that ethical behavior resides above behavior required by the law; however, in many respects the law and ethics overlap. It is important to realize that the law embodies notions of ethics (i.e., the law may be seen as a reflection of what society thinks are minimal standards of conduct and behavior).

Both law and ethics deal with what is deemed right or wrong, but law reflects society's codified ethics. In spite of this overlap, desirable ethical behavior extends beyond what is required by law. Viewed from the standpoint of civilized society, the obedience to the law is a minimum standard of behavior. In other words, what is illegal is definitively also unethical, but what is unethical does not necessarily have to be illegal.

8.4.2 Major Ethical Theories

There are few very influential directions to ethical thinking that most people today would recognize. One involves an emphasis upon seeking the best outcome for self. Another emphasizes considering the consequences of our actions for everyone who will be affected, a third one emphasizes rights and duties, and a fourth one lays stress upon the concept of virtue. Some additional ethical theories will be briefly described due to their importance to the health care industry and medicine.

Egoism[6], also called *ethical egoism*, is a normative ethical position that is becoming increasingly popular, especially in capitalist economic theory. Normative forms of egoism make claims about what one ought to do, rather than describe what one does do [13]. In this context the egoist is someone who seeks the benefit, pleasure, or greatest good for oneself alone [14]. It is quite likely they will be reasonably good neighbors, especially if they have normal friendly human sentiments, and also if they believe their lives will be the better for living in a cooperative community, so they are not as evil as they sound, or as simplistic in their egoism as some economists assume.

An egoist performs in a business or professional environment in a way that causes the lack of trust of other people and this practice could be detrimental in the long run. This point is currently receiving increasing attention in business and professional ethics.

Utilitarianism[7], one basic idea held by utilitarianism in common with egoists, is that what counts above all from a moral point of view is the consequences of the actions. What sets utilitarians apart is the principle that they should try to maximize the total good consequences (and minimize the bad ones), for all those affected by

6. Max Stirner was the first philosopher to call himself an egoist.
7. British philosophers Jeremy Bentham (1748–1832) and John Stuart Mill (1806–1873) were credited with the origins of classical utilitarianism, a moral theory that defines a moral act solely in terms of the outcome or consequences of that act.

the choices [15]. This theory is often expressed in a slogan: "seek the greatest good for the greatest number." It assumes that it is possible to prioritize values in a rank order and understand the consequences of various courses of action. However, it looks like a simplistic cost-benefit analysis, in which the end justifies the means.

One common comment about utilitarianism is that it seems to condone purchasing a maximum of total good at the expense of a small number of victims and that no action is right or wrong in and of itself. Therefore, actions are only judged in light of their consequences.

Consequentialism refers to those moral theories which teach that the consequences of a particular action form the basis for any valid moral judgment about that particular action. Thus, from that standpoint, a morally right action is one that produces a good outcome, or consequence, and the utilitarianism and ethical egoism mentioned earlier are actually varieties of consequentialism.

Rights and duties are a very different way of making moral a choice is to emphasize rights and duties. Aside from many legal rights and duties, some people have thought that there are moral rights and duties as well. One example is the so-called *natural rights*[8] listed in the UN Charter. Natural rights, also called *inalienable rights*, are considered to be self-evident and universal, and they are generally held to be a gift of nature or God that cannot be taken away [16]. Another is the duty to nurture the children, whether the laws of the land require it or not. Those who think in these terms will tend to judge the moral status of a situation not according to the acceptability of its results, but according to whether it came about by the action of proper procedures, such as doing your duty. In contemporary theory, these and other moral claims are referred to as *universal human rights* and form the basis for establishing and/or evaluating ethical standards within the social order. The idea of *human rights* is also closely related to that of natural rights.

The idea implied by the name *natural rights* is that they are moral rights that all people have, whether or not any political process has conferred them in the form of legal rights. Rights, especially *fundamental rights* (in United States, fundamental rights are part of the U.S. Constitution), are a concept favored by politicians rather than philosophers since they have more to do with power than with reason.

A conflict between two people each claiming different fundamental rights cannot be resolved in practice, and one has to look deeper than the concept of rights in order to find a solution. Therefore, it seems that rights cannot be a truly fundamental moral category, and that rights claims can provoke conflict rather than solve it. One way to see the problem is that to claim a fundamental right is to say that nothing else is ever going to matter enough to count against it, not even things you have not thought of yet, sounding almost like a fanaticism.

Cooray [17] argued that "there cannot be a right without a duty," and that "right in one person presupposes a duty in another." These days everyone is claiming their rights to this or that, without any regards to the individual responsibilities; people demand more and more while contributing less and less.

8. Modern notions of natural rights are usually closely associated with the seventeenth-century British philosopher John Locke and his contention that human beings are entitled to life, liberty, and property.

Virtues, another view of ethics, come from asking a different question, so instead of asking, what should I do, the question is, what sort of person should I be, therefore emphasizing the character of the moral agent rather than rules or consequences. Virtue ethics, which has a very ancient lineage, is enjoying a substantial revival in modern times. It provides a way of thinking that is particularly suitable for the task of developing one's character as a good person, hence its title. The idea is that every person is constructing his or her own moral character throughout his for her life by the history of his or her deeds and decisions. In practice, people who think about ethics in this way will often model their behavior on that of exemplars: saints or heroes [18]. This is opposed to consequentialism, which holds that the consequences of a particular act form the basis for any valid moral judgment about that action, and *deontologies*[9], which derives right or wrong from the character of the act itself rather than the outcomes.

Altruism[10] (also called the *ethical altruism*) is an ethical doctrine that holds that individuals have a moral obligation to help, serve, or benefit others, if necessary at the sacrifice of self-interest. Altruism is often seen as a form of consequentialism, as it indicates that an action is ethically right if it brings good consequences to others. Altruism may be considered similar to utilitarianism; however, an essential difference is that the latter prescribes acts that maximize good consequences for all of society, while altruism prescribes maximizing good consequences for everyone except the actor. Altruism could also be a form of deontological ethics asserting a duty, namely to "live for others," as its principle in spite of whether it has good consequences.

Take lying as an example; consequentialists may view lying as wrong because of the negative consequences produced by lying but may agree that in some cases lying is acceptable. A virtue ethicist would focus less on lying in any particular case and instead consider what a decision to tell a lie or not tell a lie says about one's character and moral behavior. A deontologist might claim that lying is always wrong, regardless of any potential positive outcome that might come from lying. In case of conflict between duties, our actual duty becomes to analyze the situation and decide what a right thing to do is (e.g., lying to save the life of an innocent person).

Ethics of care is not a common ethical principle but is discussed extensively in health profession and specifically with respect to nursing ethics [19]. Universal principles are only valid if they can be applied with room for discretionary judgment based on the unique circumstances of each situation. Some of the specific standards or ideals within a caring relationship include caring itself, compassion, concern, and sensitivity to context. Caution is required here since complete rejection of impartiality and ethical principles in favor of sensitivity and emotion may also lead to a rejection or a neglect of otherwise justifiable obligations and rights.

Some other theories of ethics include *absolute values* and *environmental ethics*.

9. Deontological ethics, sometimes referred to as *duty ethics*, places the emphasis on following rules, or doing one's "duty." Which rules to follow is often a point of contention and criticism in deontological ethics. For more information on deontological ethics, refer to the works of Immanuel Kant and Alasdair MacIntyre.
10. The word "altruism" was coined by Auguste Comte, the French founder of positivism, in order to describe the ethical doctrine he supported. His version of altruism calls for living for the sake of others. This doctrine had many opponents, including Friedrich Nietzsche.

8.4.3 Moral Lessons of the Past Research

Many think that the moral lessons learned from the Nuremberg Trials seem so obvious now and that there is no need to study it any further and that it could never happen again. That is a wrong conclusion because moral lessons are quickly forgotten and ethics is more fragile than many people think. Moral reasoning based on defective premises tends to recur in new settings. For example, not all of the Nazi physicians were mentally deranged; they actually believed they were doing the right thing.

The first principle of the Nuremberg Code is that "the voluntary consent of the human subject is absolutely essential." However, this principle was compromised almost immediately after the Nuremberg trials; the Helsinki Declaration, which superseded the Nuremberg Trials, weakened the provision by placing too much emphasis on the advancement of science and not enough on the integrity of the subject. Katz in his work [20] faulted the U.S. Department of Health and Human Services Rules and Regulations for a lack of a similar failure fully to protect human research subjects.

In the United States, a series of highly publicized abuses in research led to the enactment of the 1974 National Research Act (Public Law 93-348), which created the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The National Commission was charged with several tasks; one task was to identify the ethical principles[11] that should govern the conduct of biomedical and behavioral research with human subjects. Another task was to develop guidelines to ensure that specific investigations would be designed and conducted in accordance with these principles.

The Commission established that informed consent is required; to protect human subjects who participate in research, they should be given the opportunity to choose what will and will not happen to them. The essential components of informed consent are information and comprehension. Information is to be given to research subjects about the procedure, their purposes, anticipated risks and benefits, alternative procedures, and opportunity to ask questions (*veracity*, the principle of telling the truth). The subject's comprehension of material is a result of the manner in which the information is provided as well as the subject's intelligence, cognitive capability, and language. More information can be found at the Web site of the U.S. Department of Health and Human Services, Office for Human Research Protections (OHRP) [21].

The requirement for truly informed consent is a major reason behind current moves to strengthen regulatory mechanisms regarding research involving humans. Examples of earlier cases of unethical research behavior that have occurred since the Nuremberg Trials are the Tuskegee Syphilis Study, the Willowbrook Hepatitis Study, U.S. radiation experiments, the Jewish Chronic Disease Hospital Study, and

11. The term *principle* can be defined in several ways; a principle may refer to a basic truth, law or assumption. *Ethical principles* can refer to a generalization that can be used in moral reasoning or a specific rule of good conduct.

the lysergic acid study[12] supported by the U.S. Central Intelligence Agency (CIA), and probably many others that have not been brought to light. In some of these cases the initial premise is that law takes precedence over ethics, that the good of the many is more important than the good of the few, that national emergencies supersede ethics, and that some persons (prisoners in this case) can lose their claim to humanity [22].

The lesson here is that moral premises must be valid if morally valid conclusions are to be drawn and that a morally repulsive conclusion results from a morally wrong and inadmissible premise.

These and some other cases stirred a highly emotional debate on ethics. In addition, lessons learned can be gathered from some mandatory programs as well. A few very well-known examples of mandatory systems/programs targeting human subjects in United States that were implemented without the consent of the individual(s) on which they were performed:

- The DoD's mandatory vaccinations against the biological germ weapon called anthrax; in 1998, 16 military members (14 Navy and 2 Air Force) while stationed in the Gulf refused to take the anthrax inoculations. As a result, two sailors were discharged from the Navy for "disobeying a lawful general order." The others were given 30 days restriction to the ship, 30 days extra duty, reduction in pay for 1 month and/or reduction in rank. The reason given for the two sailors' dismissal was the inclusion of previous discipline problems where they had "demonstrated a pattern of misconduct" [23].
- Several mandatory vaccination programs throughout U.S. history.
- President Franklin Roosevelt signed the Selective Training and Service Act of 1940, creating the country's first peacetime draft.

Since implementation of the National Research Act in 1974, the process of protecting human subjects in the United States has been evaluated periodically by various methods. In 1998, the National Institutes of Health (NIH) sponsored evaluation which concluded that the institutional review board (IRB)-based human subjects' protection system has been implemented in accordance with regulations.

The NIH, the nation's medical research agency, includes 27 institutes and centers and is a component of the U.S. Department of Health and Human Services. It is the primary federal agency for conducting and supporting basic, clinical, and translational medical research, and it investigates the causes, treatments, and cures for both common and rare diseases.

12. Lysergic acid diethylamide (LSD or LSD-25), also known as lysergide and colloquially as acid, is a semisynthetic psychedelic drug. LSD is nonaddictive and well known for its psychological effects. Introduced by Sandoz Laboratories, with the trade name Delysid, as a drug with various psychiatric uses in 1947, LSD became a therapeutic agent. In the 1950s the CIA thought it might be applicable to mind control and chemical warfare; the agency's MKULTRA research program propagated the drug among young servicemen and students.

8.4.4 Ethics and Technology Today

One of the major problems in the development, implementation, and operation of new technologies is the failure to address social factors, especially ethical issues [24]. Most technologies have unintended and sometimes unexpected side effects, which, if known in advance, might have altered decisions about their adoption.

Two schools of thought have developed. One is well known for the dilemma of technological (and risk) assessment: in the early development of a technology, when it is relatively easy to control its direction, society inevitably lacks the knowledge to exercise reasonable control. By the time there is more experience and a better understanding of the risks, control has become difficult, if not impossible.

The other school of thought argues that persons should not be subjected to any technological risks until they have a clear understanding of that the risk and have granted their consent without being constrained by economic or other external pressures. The contention is that the concept of free and informed consent as applied in the field of medicine is applicable to technology in general and ought to be a part of what guides morally grounded public policy.

What is ethical and what is not ethical could be a difficult question. Different people will have different answers to the same question; the same people could even have different answers today from the answers given five years ago and completely different answers five years in the future. As recently as 50 years ago, a person would have been considered totally insane to say that in a few years a human being would walk on the Moon; jet planes were just becoming commonplace, and no rocket had ever been launched, but Sputnik in 1957 started the competition in outer space, and the first footstep on the Moon on July 20, 1969 (only 12 years later) proved that same person to be a visionary.

8.4.4.1 Engineering Ethics

Whether engineering ethics is an independent field of applied ethics or if it can be included in an already existing field has been argued by traditional ethicists and advocates of the uniqueness theory.

Traditional ethicists do not think that there is anything unique about the moral problems, for example, privacy, free speech, and intellectual property, which are considered by engineering ethicists. These new moral problems, which are associated with engineering, medicine, and business, according to the traditional ethicists can be analyzed by using the traditional ethical theories and categories of morality, as discussed in [25].

According to John C. Maxwell, an internationally respected leadership expert, speaker, and author [26]:

> There's no such thing as business ethics—there is only ethics. People try to use one set of ethics for their professional life, another for their spiritual life, and still another at home with their family. That gets them into trouble. Ethics is ethics. If you desire to be ethical, you live it by one standard across the board. Educators, philosophers, theologians, and lawyers have taken what really is a simple matter and made it very confusing. Living an ethical life may not always be easy, but it need not be complicated.

The opposite side in this debate is the advocates of the uniqueness theory; they find that some aspects of engineering ethics are unique since data communications ethics issues did not exist before the introduction of computing. Some advocates mean that the moral problems associated with data communications cannot be analyzed with the traditional morality; instead, there is a need of a new computer ethical theory or a new framework of morality.

Of course, different fields can have their specific ethical challenges (professional ethics): medicine, biotechnology, engineering, and business, but they are all intertwined under the common umbrella of ethics. A professional faces several situations on a daily basis, where the ethical values have to be considered, for example, decision-making, where the consequences are significant for both the person who makes the decision and the people who are affected by it.

In the case of wireless body implants there are a number of overlapping professional ethical views (medicine, bioengineering, computer engineering) that may be contradicting each other, which makes the situation and decision-making process quite challenging. Before the discussion can start on the ethical issues in medicine, specifically wireless body implants, a few general terms will have to be defined, such as ethics, morality, and conscience, which will be pivotal for this discussion.

In order to address the issues of health and safety that were of concern to engineering during the industrial revolution, a new ethical framework was required. The establishment of professional engineering societies in the nineteenth century, including the American Society of Civil Engineers (1852), the American Society of Mechanical Engineers (1880), and the American Institute of Electrical Engineers (1884), gave ethics a forum in which conflicts could be resolved [27]. The modern engineering ethics code development in United States began after World War II, as engineers became increasingly aware of the social impact of their work and of corresponding social responsibilities. The key characteristic of this period was the rise to codified prominence of a new principle recognizing the importance of public safety, health, and welfare.

Beginning in the 1970s, individual engineers, especially professors of engineering, became more involved, often working in tandem with academic philosophers. This new phase was stimulated by a series of widely publicized cases perceived as examples of engineering negligence or improper subordination to economic interests and by federal funding for engineering ethics research [28]. The fast-paced changes in technology leave decision-makers, legislators, and other major stakeholders[13] with no time for in-depth analysis, and the international community is often faced with immediate policy choices that carry serious moral and ethical consequences.

Only recently have philosophers started making direct and sustained study of ethics and technology, following two philosophical traditions, each marked by distinct differences: the continental or phenomenological tradition and the Anglo-American or analytical tradition.

The continental approach sees technology as a special subject of ethics; because technology has fundamentally transformed the human condition, generating prob-

13. Every ethical, social, and political issue has stakeholders (i.e., players in the game who have an interest in the outcome and usually have vocal opinions).

lems of global magnitude extending into the indefinite future, it calls for a new approach to ethics.

According to Hans Jonas, a twentieth-century German-born philosopher, ethical prospective technology could, in the past, remain in the background because technology itself had no high moral purpose. Unlike politics or religion, for instance, technology used to be treated as a marginal aspect of human life, one limited in both power and effect; by contrast, during the modern period technology became a lot more important.

The Anglo-American tradition does not tend to deal with technology as a whole, but looks separately at particular technologies, such as computing, engineering, and medical and biological sciences. It is based on the concepts and principles of traditional ethical theory at least as a starting point for analyses. Although each of the technologies has a unique set of problems, certain themes, such as responsibility, risk, equity, and autonomy, are common to almost all. As opposed to the continental approach, philosophers following an Anglo-American analytical tradition organize ethical discussions around particular technologies.

As discussed in the *Routledge Encyclopedia of Philosophy* [29], biomedical ethics studies the ethical implications of the use and development of advanced medical technologies; information technology ethics (i.e., computer ethics) examines social and ethical significance of computers and high-speed digital networks; engineering ethics studies the professional responsibilities of engineers; and environmental ethics evaluates the effects of various technologies on the natural environment.

Invention and application of advanced technologies is commonly evaluated in terms of liberty and autonomy. In biomedicine, for example, in the name of liberty and autonomy, there is an effort to work out the exact parameters of free and informed consent and then to propose ways to institutionalize them. In the area of information technology (IT), liberty and autonomy are key to explaining normative dimensions of privacy as well as prescribing the extent of freedom of speech over the digital networks.

During the last half of the twentieth century, new fields of ethical reflection emerged to deal with the nuclear weapons (*nuclear ethics*), chemical transformation of the environment (*environmental ethics*), biomedical advances (*bioethics*), and computers and IT (*computer ethics*) so the ethics of scientific research and of the engineering practice became specialized areas of study themselves. At the beginning of the twenty-first century, ethical and political challenges became global in scope and intensified by the terrorist use of technology and science.

Ethics, generally speaking, is concerned with identifying proper means and distinguishing good and bad ends. Science, technology, and ethics interactions in these broad senses have, furthermore, been examined from multiple historical and cultural perspectives. The Continental European tradition, for instance, tends to focus more globally on science and technology as a whole, whereas in the Anglo-American tradition dominate [30] the ethics of particular technologies, for example, medical ethics or computer ethics, ethics in areas of professional practice such as engineering and business, or issues such as equity, privacy, and risk.

8.4.4.2 Efforts to Regulate the Future

In a process of making ethical decisions, it is important to realize that solution often means that a reasonable compromise among various requirements is reached. The best solution may not be the best technical, best economic, best political, or best social, although it should consider all of them.

New technologies and their impact on society was discussed in report prepared for the conference prior to EXPO 2000 in Hannover, Germany [31], and over the last 10 years of the twenty-first century many of the assumptions and predictions started to materialize.

Many of the new technologies are directly or indirectly related to human body, health, and performance, with biomedicine and bioengineering being one of the most often discussed. The U.S.-government-sponsored workshop in 2001 resulted in a detailed document [32] that predicted some of the technology and technology-related developments over the next quarter-century (*NBIC*, short for *nano-bio-info-cogno*):

- Fast, broadband (human) brain-to-machine interfaces for civilian and military purposes;
- Robots and software operating on principles compatible with human goals, awareness, and personality;
- More durable human body, more resistant to different kinds of stress (physical and/or emotional), aging, biological threats;
- Control of the genetics of humans, animals, and plants;
- International consensus on ethical, legal, and moral issues.

The US-NBIC initiative is technology-driven and is influenced by a perspective of national security that emerged after 9/11 and by a vision of competitive security and defense for the United States. Considerations of ethical, legal, or social and moral issues related to NBIC are to be addressed a posteriori. Also missing are considerations of risks and issues. The technocratic understanding of society and culture and the vision of a perfect future are also cited as problem areas [33]. One crucial subject for the development of the ideas presented in the NBIC report is the development of neural implants.

Following the U.S. NBIC Conference, the EU decided to constitute a high-level expert group (HLEG). In December 2003, a 25-member group was drawn from a variety of countries and disciplinary backgrounds. The group met formally four times and submitted its report in July 2004. The European Union developed its own definition of converging, EU-HLEG; the definition is broad, but nanotechnology, biotechnology, and information technology have central roles.

8.4.5 Enhancing Humans

8.4.5.1 History of Human Enhancement

The racist doctrine of human perfection promoted by Nazis was based on *eugenics*[14], defined by nineteenth-century inventor Sir Francis Galton. Hitler's *lebensborn* or "life spring" project was supposed to increase the number of blue-eyed, tall, blond Aryans[15] by mating racially screened women with SS men and officers in the German regular army [34]. In the early part of the Nazi era, enforced sterilization and legal bans on the intermarriage of superior and inferior humans were the preferred means of eliminating bad hereditary features; later, death camps were used as a more efficient method.

Even long before Galton, a classic Greek philosopher Plato believed that human reproduction should be monitored and controlled by the state. Eugenics was also practiced in different and less deadly forms in other parts of Europe and in the United States[16].

Today scientists know a lot more about genetic engineering, cloning, and so forth. In [35] Satava pointed out that humans are the only species on this planet with the capacity to direct their own evolution at their own accelerated pace and no longer confined by neither the will of nature nor the excruciating slow pace of evolution. This may be the ultimate challenge for the next generation of scientists and humans in general. It is also critical to realize that technology has no moral value; it is neither good nor evil. Rather, it is the application of the technologies that raises the moral and ethical issues.

Supersensory sight could see radar, infrared, and ultraviolet images; augmented hearing could detect lower- and higher-frequency sounds. Enhanced smell could intensify our ability to discern scents, and an amplified sense of touch will enable discernment of environmental stimuli such as changes in barometric pressure. These capacities would change the usual standards for humans (i.e., as the numbers of so-called *enhanced humans* increase, today's normal might be seen as subnormal, so the important questions revolve around whether there should be limits placed upon modifications of essential aspects of the human species [36]).

For example, *kleptomania* is the compulsive need to steal what one could afford to buy, and it used to be considered a moral defect requiring counseling or punishment or behavior change. Recent research found the biochemical site on the brain which is the source of kleptomania so the condition is without question the result of a biochemical lesion.

14. Eugenics is the "applied science or the biosocial movement which advocates the use of practices aimed at improving the genetic composition of a population," usually referring to human populations. Eugenics was popular in the early decades of the twentieth century, but has fallen into disfavor after having become associated with Nazi Germany.
15. The term *Aryan* originates from the Sanskrit word ārya, in origin an ethnic self-designation, in classical Sanskrit meaning "honorable, respectable, or noble." The Aryan race is a concept historically influential in Western culture in the period of the late nineteenth century and early twentieth century, starting with the idea that the original speakers of the Indo-European languages and their descendants up to the present day constitute a distinctive race or subrace of the larger Caucasian race. While originally meant simply as a neutral ethno-linguistic classification, it was later used for ideologically motivated racism in Nazi and neo-Nazi doctrine and hence also in other currents such as occultism and white supremacism.
16. Charles Davenport, an American scientist, stands out as history's leading eugenicist. In the early twentieth century he took eugenics from a scientific idea to a worldwide movement implemented in many countries.

Every mental characteristic, whether it is a matter of personality, cognition or emotionality, will eventually be identified as a biochemical process which itself is largely genetically determined and hence can be altered. These alterations may be pharmaceutical or they may be genetic, acoustic, visual, or by some type of electronic implants.

Adderall, a medication for Attention Deficit Hyperactivity Disorder (ADHD)[17], has become popular among college students who actually do not have the disorder. Adderall is the combination of dextroamphetamine[18] and amphetamine and works like cocaine; those who use it can stay focused and awake for many hours. Students with prescriptions commonly sell it or give it away to others. This medication is in a class of medications called *central nervous system stimulants*, and it works by changing the amounts of certain natural substances in the brain [37].

The U.S. Air Force has been using dextroamphetamine for over 60 years and gives them to pilots on long missions to help them remain focused and alert, but in December 2003, a new compound, *modafinil*, was approved for the same purpose [38]. It is interesting to note that, although approved by FDA, the exact mechanism of how modafinil works is still unknown.

Ritalin (Methylphenidate)[19] is a psychostimulant drug approved for treatment of ADHD, postural orthostatic tachycardia syndrome, and narcolepsy. It may also be prescribed for off-label use in treatment-resistant cases of lethargy, depression, neural insult, obesity, and sometimes for other psychiatric disorders such as obsessive-compulsive disorder. Based on the pool results published in the journal *Nature* in 2009, one in five people reported using Ritalin not as medically intended but to increase their brain power (*brain boosting*).

The first brain implants were surgically inserted in 1974 in the state of Ohio and also in Stockholm, Sweden. Brain electrodes were inserted into the skulls of babies in 1946 without the knowledge of their parents. According to Esther Morales, in the 1950s and 1960s, electrical implants were inserted into the brains of animals and humans, especially in the United States, during research into behavior modification and brain and body functioning [39].

Influencing brain functions became an important goal of military and intelligence services, and mind control methods were used in attempts to change human behavior and attitudes. Today implants are small enough to be inserted into the neck or back and also intravenously in different parts of the body during surgical operations, with or without the consent of the subject, and it is almost impossible to detect or remove them.

Aside from the ethics of implanting devices in a human body, there is also an ethical dilemma of who is responsible for removing or turning off these devices [40]. Implantable devices have a long history in medicine, with artificial hips being

17. Attention deficit hyperactivity disorder (ADHD) is a neurobiological disorder in which a person has more difficulty focusing, controlling actions, and remaining still or quiet than other people who are the same age. It affects 5% of children before the age of 19 but also adults.
18. Dextroamphetamine is classified as Schedule II, the most restrictive category possible for a drug with recognized medical use.
19. Methylphenidate was synthesized by Ciba chemist Leandro Panizzon. His wife, Marguerite, had low blood pressure and would take the drug as a stimulant before playing tennis. He named the substance Ritaline, after his wife's nickname, Rita.

implanted since 1925, pacemakers since 1957, Starr-Edwards heart valves since 1961, artificial hearts since 1982, and ventricular assist devices since 1991. The ethics of deactivation or removal of these devices were not an issue until the use of Implantable Cardioverter Defibrillator (ICD) device, as the ICD can produce considerable distress from defibrillation shocks in end-of-life patients.

8.4.5.2 Today's Definition of Human Enhancement

The definition of what exactly is *enhancement* is not very clear, since it is not clear what constitutes "normal." In the United States, the President's Council on Bioethics defines human enhancement as going "beyond therapy" (i.e., not just becoming healthy again but exceeding the normal, healthy state). Humans have always sought to enhance themselves; ancient Olympians ate mushrooms for improved success in their events, and Renaissance women wore corsets to slim their waists [41]. These days a plastic surgery offers a number of surgical procedures with a goal of improving the appearance of men and women.

Today, the human performance modification (enhancement) includes many different scientific disciplines such as learning, psychology, neurology, and pharmacology, as well as focused research in sleep and cognition, and development of prosthetics and treatments for spinal cord damage. In spite of the serious reasons for these areas of research, there is the potential for abuses, as well as serious concerns about a point at which changing natural humanity begins. Such ethical considerations should and will appropriately limit the types of activities and applications in human performance modification.

In the future, brain technology might go beyond helping sick, for example, offering behavior modifications for the people who are short-tempered, have no sense of humor, or are too emotional. Beyond that is the possibility of enhancing people's cognitive processes, enabling them to think more clearly, to have better command of arithmetic, to have a better memory for faces, to be more generous and loving, and so forth. Of course, relief from these conditions will not be accessible to everyone since they probably will not be cheap; this is one of the most important ethical challenges of all.

These changes in human nature could become even more emphasized in the case of children. Rich parents in an intensely competitive society may be able to secure implants for their children and most likely change the already unequal opportunities in life. The major concern should be the social impact of implementing a technology that widens the divisions not only between individuals, but also between rich and poor nations, as well as the possibility that it could facilitate control of humans. The ethical principle of equity is suggesting that everybody should have fair access to the benefits under consideration.

On the other side of the debate are those who claim that using new technologies to enhance native human capacities is not different from the old technologies [42], such as using clothes to enhance human skin or shoes as enhancement of human feet.

Using these new technologies, commercial interests or governments could control and monitor citizens. In a free society this possibility may seem remote, although it is plausible to project initial compulsory usage for children, for the military or for criminals (as it is already happening with some simple RFID applications).

Policy decisions will arise about this usage and also about mandating implants to affect specific behaviors. The main question is who will control the technology and what will be programmed so the prospects for invasions of liberty and privacy are alarming.

From Birth to Death and Bench to Clinic: The Hastings Center Bioethics Briefing Book for Journalists, Policymakers, and Campaigns contains 36 overviews of issues in bioethics of high public interest, such as abortion, health care reform, human and sports enhancement, organ transplantation, personalized medicine, medical error, and stem cells. In addressing the question of potential threats that may arise from adversarial activities in human performance modification, the single most important factor is awareness, with the ability to assess the significance of the developing applications, the Hastings Center, created in 1969, is a nonpartison research institution dedicated to bioethics and public interest.

In bionics, the future is in the creation of artificial muscles and nerves, making it possible for progressive technological integration into the body, eventually replacing or augmenting the structures that mediate the various physical and mental attributes that people normally consider "natural" to human beings, including emotion, natural sensory modes, rational thought, and properties of imagination.

It is clear that the severely disabled are often the first to appreciate the fruitful couplings of humans and machines. A brief discussion with anyone who has a pacemaker, a new hip, a hearing aid, an artificial heart, or any one of a host of bionic devices will show that. The neural prosthetic and interface technologies of today can be broken down into three major areas: auditory and visual prosthesis, Functional Neuromuscular Stimulation (FNS), and prosthetic limb control via implanted neural interfaces. So far, the most successful implants have been in the realm of hearing, and there are thousands of people worldwide outfitted with cochlear implants. Although current versions of these devices may not match the fidelity of normal ears, they have proven very useful.

In a *Medical Ethics* magazine [43], Ellen M. McGee, a retired adjunct professor of philosophy and presently an ethics consultant, described the future that may include the reality of science fiction's cyborgs, persons who have developed some intimate and/or necessary connection with a machine. It is likely that computer chips implanted in our brains and acting as sensors or actuators may soon not only assist the blind and those with failing memory, but even provide fluency in a new language, enable recognition of previously unmet individuals, and provide instantaneous access to encyclopedic databases. Some others claim that even today humans already are cyborgs of a sort.

U.S. federal agencies are involved in national security and medical research as well. For example, one of them is DARPA, with its funding of unclassified studies conducted at various universities, described by Jonathan D. Moreno, a professor of medical ethics, history and sociology of science, and philosophy, in [44]. Included in such work is a conventional prosthetics research but also an effort to have machines and brains communicate in order to improve real-time human performance in battle.

A similar program is *AugCog* (augmented cognition), which tries to eliminate the boundary between artificial and natural intelligence. The helmets would sense a soldier's emotions and then transmit appropriate combat orders based on telemetric findings. These devices could potentially rob combatants of autonomous behavior

(i.e., free will) conceivably depriving them of making ethical choices. These efforts are linked to other DARPA programs to improve the soldier's endurance, alertness and vigilance, fear response, and pursuit of a hypercourageous, but probably judgment-impaired, soldier.

There are different thoughts on the possibility of the future intelligent machines and/or cyborgs. In his paper J. Sorrs Hall [45] argued that a society including super-ethical machines would not only be better for people to live in, but stronger and more dynamic than human society is today, so not only should humans give consciences to the machines whenever possible, but creating machines that exceed humans morally as well as intellectually is required.

8.4.5.3 The Limitations of Human Enhancements

Over the last few years, human enhancement has become an important topic of debate in applied ethics. In previous chapters of this book, technical achievements of modern biotechnology were discussed in which medicine, science, and engineering are working together to extend our longevity as well as to enhance life's quality. Living to be 120 years old and be able to play tennis at that age is an ultimate goal of the life-enhancing techniques; sitting for 30 years in a wheelchair, maybe without even recognizing one's closest family members any longer, is definitely not what society should be striving for.

Longevity without the quality of life is meaningless. In addition, all phases of life should be prolonged, not just the retirement age; it is hard to imagine people having the same length of their youth and working through a productive middle age, but having an extended retirement and old age to last 50 to 60 years. In either case, it will require a very different social structure of the society in which people will be living.

Although small steps towards enhancing our lives and extending it have to be analyzed very critically and very carefully due to the safety and security issues, the wheel of progress cannot be stopped. Over the last few thousand years, humans have significantly improved and extended lives more than doubly; it is hard to imagine lives without surgeries, medications, prosthesis, and other modern marvels of technology and medicine, without which humans would probably go back to the life expectancy of 25 or maybe 30 years of age. From that prospective, the human race has already been enhanced.

The environment in which people are living and the way they (and all other creatures on Earth) are living their lives determines the longevity; they are not determined just by nature (or God) itself, if at all. Therefore, it is not arrogant or unnatural or unethical to strive towards further improvements and enhancements of our bodies and our beings.

This is a very new and promising technology for improving the quality of people's lives as well as extending them. On the other hand, a vision of body organs, including the heart and brain, being able to communicate with the outside world and exchange information and being radio controlled by the outside controller, can be frightening. Here, less of the concern lies in the conspiracy or government-run

programs than in the potential misuse and abuse by the greed of corporations and criminals.

The Latin saying *noli me tangere* (a warning against touching or interfering) was still very much practiced in 1944, when Dr. Alfred Blalock performed his first heart surgery. The heart was something not to be touched while people were dying from all kinds of heart-related ailments. Many of his colleagues at the Johns Hopkins hospital in Baltimore, Maryland, did not even wish to assist Dr. Blalock in performing such a groundbreaking surgery. Open heart surgeries, heart transplants, artificial hearts, and heart pace-makers followed over the next 60 years, ensuring that humans live longer and better. So why should the brain not be the next logical step?

The brain is a much more complex organ than heart; the heart can be analyzed (and modeled) as a little bit more complicated mechanical device, perhaps as a pump, while the brain is a combination of numerous and very complex electronic circuits about which scientists still do not know much. Therefore, electronic chip implants are not only required as a therapeutic tool but also as a research instrument that will help scientists to better understand the human brain. In view of these potentially great implications of the implantable brain chip, its development and implementation should definitely be regulated, after an open dialogue that will ensure that everyone's needs and concerns are being addressed.

Even the experts disagree; while some think that there should be no limitation on how people can choose to modify themselves, others think that the technology should be regulated, treated as research on human subjects, and closely monitored for its effects. The common concern is the use in the military, prisons, and for children or other individuals whose power to make decision (informed consent) might be limited or completely nonexistent. Being able to decide when and if something will be done to our bodies, as well as being able to offer the same opportunity to everyone, are crucial parts of the process.

8.4.6 Ethical Decision-Making Process

8.4.6.1 Basic Principles

We might say, according to John Bordley Rawls, a twentieth-century American philosopher and a leading figure in moral and political philosophy, that the principal aim of ethics is the formulation of justifiable principles that may be used in cases wherein there are conflicting interests to determine which one of them should be given preference [46]. When a decision is made about what is ethical (right, just, fair) using the conventional approach, three key elements described next go into such a decision.

First, observing the behavior, act, or practice that has been committed or is about to happen and that the comparison of the practice with prevailing norms of acceptability, that is, society's or some other standard of what is right or wrong (Figure 8.3). In other words, ethical principles provide a common framework that individuals who hold differing normative positions can adopt when debating an issue.

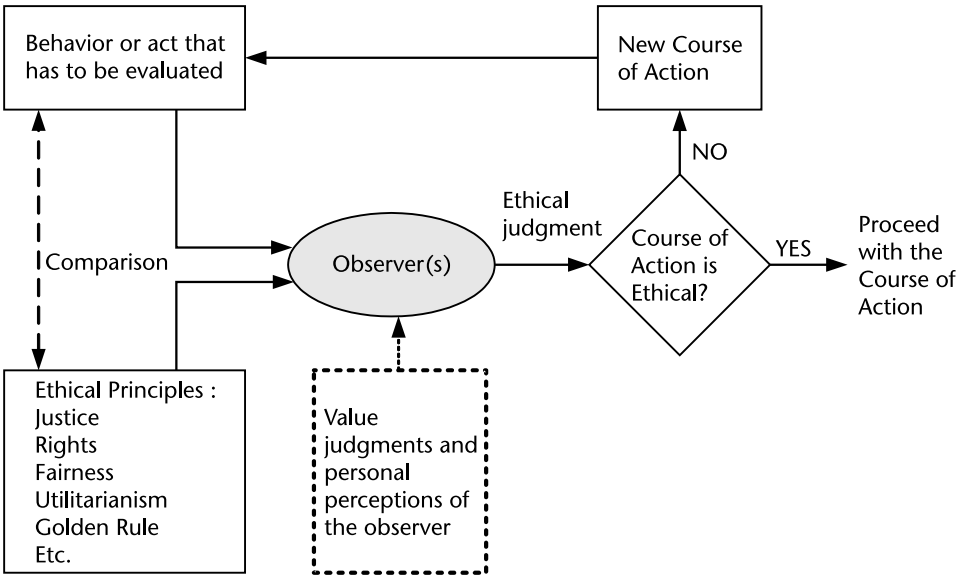


Figure 8.3 Ethical judgment process.

It is important to recognize that value judgments are being made by someone with his or her personal opinion and bias towards the judgment outcome (subjectivism). This means that two different people could look at the same behavior, compare it with their concepts of what the prevailing norms are (i.e., different ethical principles), and reach different conclusions as to whether the behavior was ethical or not, and none of them has to necessarily be wrong. This becomes quite complex as perceptions of what is ethical inevitably lead to the difficult task of ranking different values against one another.

Decisions that include ethical dilemmas can quickly become very complex. Generally speaking, members of society generally agree at a very high level that certain behaviors are wrong; however, the consensus tends to disintegrate when faced with the not abstract but specific, real-life situations. The most serious danger is that of falling into an *ethical relativism* in which people pick and choose which source of norms they wish to use based on what will justify their current actions or maximize their freedom.

There are numerous implications and consequences of advanced technologies, and these issues fall into several categories: scientific (is the science really safe?), social (what are the societal implications?), behavioral (how will individuals' behavior change?), political (how will the legal and regulatory systems react?), and philosophical (what fundamental moral and ethical values are challenged?).

According to LaRue Hosmer, professor emeritus of corporate strategy [47], five important points should be made about the character and nature of ethics and decision making:

- They have extended consequences;
- They have multiple alternatives;
- They have mixed outcomes;

- They have uncertain consequences;
- They have personal implications.

It is a very complex task to decide whether a new and advanced technology is ethical and morally acceptable, since not all of the implications of the new technologies can be foreseen or predicted at the time of their development. The ethical implications of application of wireless body implants in neuroscience are new challenge, since the ability to intervene in the brain in many ways is a recent undertaking. Therefore, it is critical to identify the components likely to challenge our conventional thinking and investigate their social, behavioral, political, moral, and ethical implications.

Some critics would claim that some of the body implant technologies cannot be accomplished and/or perfected in the next two to three decades, but the issues are so important that even longer time spans may be inadequate for us to prepare for the potential consequences.

The intimate relation between bodily and psychic functions is basic to person's personal identity. Neurosciences are developing very quickly, and the brain implants developed to alleviate tremors in Parkinson's disease are only one example. They show that body implants may influence the nervous system and particularly the brain and thus human identity, as well as individual subjectivity and autonomy.

8.4.6.2 Organizational Ethics Decision-Makers

The creation of hospital ethics committees in the 1980s reflected a need on the part of hospitals to establish some mechanism to deal with those clinical dilemmas and policy issues that revolved around the appropriate use of medical technology. There is general agreement on the (usually) threefold function of such committees: case consultation, policy recommendation, and education, although institutions vary in their approach to these functions.

Some committees are heavily represented by hospital administration or hospital counsel and maintain a defensive posture for the institution, while others actively exclude such individuals. Some committees are largely physician committees and their deliberations represent that perspective. Maybe the best approach is to attempt to establish a broad-based, multidisciplinary approach and also include nurses, social workers, clergy, and academic ethicists from philosophy or the social sciences [48].

The most common function of ethics committees is to provide clinical case consultation. There is a wide variation in approach (i.e., style and methodology) from one committee to another. While it is still common for full committees to interview patients and their families and/or other parties, smaller consultation teams or subcommittees are becoming more common. In fact, the solitary ethics consultant model has been used in a number of institutions with some success.

Regardless of which model is used, an important question remains about authority and expertise. Moral authority in a committee setting is often a result of a consensus; the authority of an individual consultant is a reflection of his or her training and expertise and these will remain vague and elusive concepts until such time that there are accreditation organizations to evaluate such individuals.

Professional organizations such as the American Society for Bioethics and Humanities[20] [49] should play an important role in the development of such standards.

Over the past few years businesses have also become more proactive in seeking to protect themselves from future scandals, and sometimes it seems that many companies are obsessed with ethics programs and compliance; the increased awareness of and need for ethical behavior are making a difference. Business leaders supporting ethical behaviors and communicating values are critical to a company's ethics. If a company's underlying values do not reward ethical behaviors, then an opportunity has been missed that sets the company up for future problems.

A good ethical decision-making process requires the ability to explore all the aspects of a decision and then to weigh the options surrounding a course of action. Ask yourself these questions: If you had to explain your decision on television, would you be comfortable doing so, and if you had to do the same thing over again, would you do anything differently? In a corporate world, this is the role of the *chief ethics officer* (sometimes called the chief ethics and compliance officer) who acts as the point person to steer all levels of employees toward integrating ethics into decision-making processes and codes of conduct.

The reporting structure is very important when it comes to ethics; the ethics officer must be independent in order to be effective, so ideally, the chief ethics officer should report directly to the CEO (but not CFO) or the board's audit committee.

The analysis of the interests and opinions could be a very effort-making and time-consuming process. In the case of wireless body implants, there are a number of interested parties (stakeholders) involved in the ethical decision-making process. Organizations developing wireless body implants and/or applied technology should have a person onboard to coordinate these efforts.

8.4.6.3 Stakeholders and Their Involvement

The society typically desires the benefits of science without exposing themselves to any risk; however, there are few human activities that are completely risk-free. Effort is needed to improve public education about minimal risk approaches as opposed to no risk approaches and about risk analysis. It is important to recognize that every scientific advance will be questioned for its value in both "good" and "bad" applications, regardless of how these are judged.

It is difficult to envision the situation where a complete consensus will ever be obtained; there are potential or actual hazards to every scientific discovery and/or technological application and a certain compromise has to be made. One way to address this conflict is to require that safeguards be put in place to protect the public from the malicious use and/or the well-intentioned but incautious use (i.e., misuse) of science and technology.

All species in nature strive to do things better and faster, but there is one important difference that set humans apart from all other species in nature, the capability to restrain themselves and make a conscious decision whether to do something

20. The American Society for Bioethics and Humanities is a professional society of more than 1,500 individuals, organizations, and institutions interested in bioethics and humanities. It was founded in 1998 by the merger of the Society for Health and Human Values, the Society for Bioethics Consultation, and the American Association of Bioethics.

or not. Just because humans can do something, it does not mean they should. William "Bill" Ernest McKibben is an American environmentalist and writer who frequently writes about global warming and alternative energy and advocates for more localized economies. In his books and articles he covers this topic extensively from the environmental position [50].

McKibben belongs to a group of so-called *bioconservatives* who are generally opposed to the use of technology to modify human nature. A focal idea in bioconservativism is that human enhancement technologies will undermine human dignity. On the other side of the spectrum are so-called *transhumanists,* who believe that human enhancement technologies should be made widely available, that individuals should have broad discretion over which of these technologies to apply to themselves, and that parents should normally have the right to choose enhancements for their future children [51].

Langdon Winner, a professor of humanities and social sciences and a political theorist who focuses upon social and political issues that surround modern technological change, offers three guidelines about how to start "real life" (sometimes called "lifeworld") discussions [52]:

- No engineering without political deliberation. And that would necessitate ethical and social deliberation as well. Who gets what and why are questions that need to be addressed.
- No means without ends. Clearly established and meaningful goals need to be discussed and agreed upon before any design, development, production, or distribution gets under way.

This is a good starting point for a discussion about new technology and its development, but it needs further explanation. Winner emphasized the importance of incorporating all stakeholders in a decision-making process; determining who the stakeholders are and how to incorporate them are sometimes difficult questions to answer and could lead to a completely new discussion. In addition, sometimes inventions are created without unique purposes (i.e., meaningful goals) in mind, but once accepted in real life, users come up with appropriate applications.

8.4.7 Mathematical Modeling of Ethical Decisions

According to William Starbuck, "Decision implies an end of deliberation and the beginning of action." Typically, decision-making process consists of five steps: the initiation step (i.e., the recognition of the need), the problem-defining step, the system-modeling step (mental, graphic, physical, or mathematical), the analysis and evaluation step, and the implementation (and reevaluation) step. Without reevaluation this is an *open-loop process,* while the process that includes observation of the results and return to problem-defining step is called a *closed-loop process* [53].

There are two mental (cognitive) approaches typically used in decision-making and problem-solving processes: intuitive and analytical [54]. In many cases, *intuitive decisions* are not supported by data and documentation and may appear arbitrary, and many decisions (probably too many) today are of the intuitive type. The person charged with making the decision collects information, probably biased by his or her own values, and after some thinking makes a decision. Over the years,

researchers from different scientific disciplines have contributed to the understanding of the human actions and decisions[21].

Some of the reasons why decision-makers should not rely just on the subjective and intuitive decision-making process are:

- Humans are not very good at making complex, unaided decisions;
- Individuals respond to complex decisions by using intuition and/or personal experience to find the easiest solution;
- Teamwork may not be helpful, and groups can devolve into entrenched positions resistant to compromise;
- The temptation is to think that honesty and common sense will suffice, and that may or may not be the case.

On the other hand, analytical decision-making, when used collectively, leads to decisions with which most stakeholders can agree. Generally speaking, every decision-making problem is defined by the presence of multiple alternatives; when the feasible set of alternatives of a decision consists of a finite number of elements that are explicitly known in the beginning of the solution process, it creates an important class of problems called the *multicriteria evaluation problems*.

8.4.7.1 Decisions and Decision-Making

The goal of the *multicriterial decision analysis* (MCDA) methods is to define priorities between alternatives (actions, scenarios, projects) according to multiple criteria, allowing for more in-depth analysis of the problem because they consider various aspects. MCDA methods use partly mathematical hypotheses and partly information gathered from the decision makers and stakeholders.

MCDA methods utilize a decision matrix to provide a systematic analytical approach for integrating risk levels, uncertainty, and valuation, which enables evaluation and ranking of many alternatives. Within MCDA, almost all methodologies share similar steps of organization and decision matrix construction, but each methodology synthesizes information differently. Most problems involve four separate hierarchies: benefits, costs, opportunities, and risks.

There is a large number of methods used for the multicriterial decision analysis, for example, the analytic hierarchy process, Bayesian analysis, multiattribute utility/value theory (MAU(V)T), and outranking methods. Different methods require different types of value information and could follow various optimization algorithms [55].

Deliberative multicriteria evaluation (DMCE) combines the advantages of multicriteria analysis in providing structure and integration in complex decision problems with the advantages of deliberation and stakeholder interaction and conflict resolution provided by the use of the community reference panel or citizens' jury [56].

21. Sigmund Freud's work on the unconscious suggests that people's actions and decisions are often influenced by causes hidden in the mind; in their book on game theory, John von Neumann and Oskar Morgenstern describe a mathematical basis for economic decision-making.

Some techniques rank options, some identify a single optimal alternative, some provide an incomplete ranking, and others differentiate between acceptable and unacceptable alternatives. Unfortunately, many methods assume independence of criteria and alternatives in their formulation; often real-life problems involve mutual dependencies that cannot be neglected by always assuming independence.

The choice of which model is most appropriate depends on the problem at hand and may be to some extent dependent on with which model the decision-maker is most comfortable. It is important to remember that applying different MCDA methods on a same problem and using the same data may yield different results.

8.4.7.2 Analytic Hierarchy Process

It is important to take a systematic and well-organized approach to making serious decisions in problem solving that also involves ethics. When selecting a methodology to address an ethical dilemma related to a new technology, as in wireless body implants, there is an aim to achieve several objectives. The method should improve understanding of the problem as well as simplify the judgment process so that the complexity of the dilemma does not compromise the quality of the decision. In addition, it should enhance the ability to examine and explain and/or defend the decision.

The Analytic Hierarchy Process (AHP) was developed by Thomas L. Saaty (born 1926 in Iraq), an American mathematician, in the 1970s, and published in 1980 and 1994, as a decision-making tool involving building a hierarchy (i.e., ranking) of decision elements and then performing comparisons between each possible pair in each cluster as a matrix and initially described in [57]. This gives a weighting for each element within a cluster (or level of the hierarchy) and also a consistency ratio, useful for checking the consistency of the data. The AHP provides a means of making decisions or choices among alternatives, particularly where a number of objectives have to be satisfied (i.e., multiple criteria or multiattribute decision-making).

Looking at the hierarchy from top to bottom (Figure 8.4), the AHP structure comprises goals (systematic branches and nodes), criteria (evaluation parameters), and alternative ratings (measuring the adequacy of the solution for the criterion). Each branch is then further divided into an appropriate and desired level of detail. At the end, the iteration process transforms the unstructured problem (sometimes with subjective data only) into a manageable problem organized both vertically and horizontally, under the form of a hierarchy of weighted criteria.

The selection of the criteria for the project estimation is a very important part of the proposed method. When the number of criteria increases, the importance of each criterion is reduced. Level 0 is the goal of the analysis, while Level 1 is a multiple criteria that consist of several factors. There could be more levels of subcriteria (Level 2) and subsubcriteria added. The last level, Level 3, is the alternative choices.

An important aspect in structuring a hierarchy is that any element in a level can be compared with respect to some elements in the level immediately above. The hierarchy does not have to be complete (i.e., an element at an upper level need not function as a criterion for all the elements in the lower level). It can be partitioned into nearly disconnected subhierarchies sharing only a common topmost element.

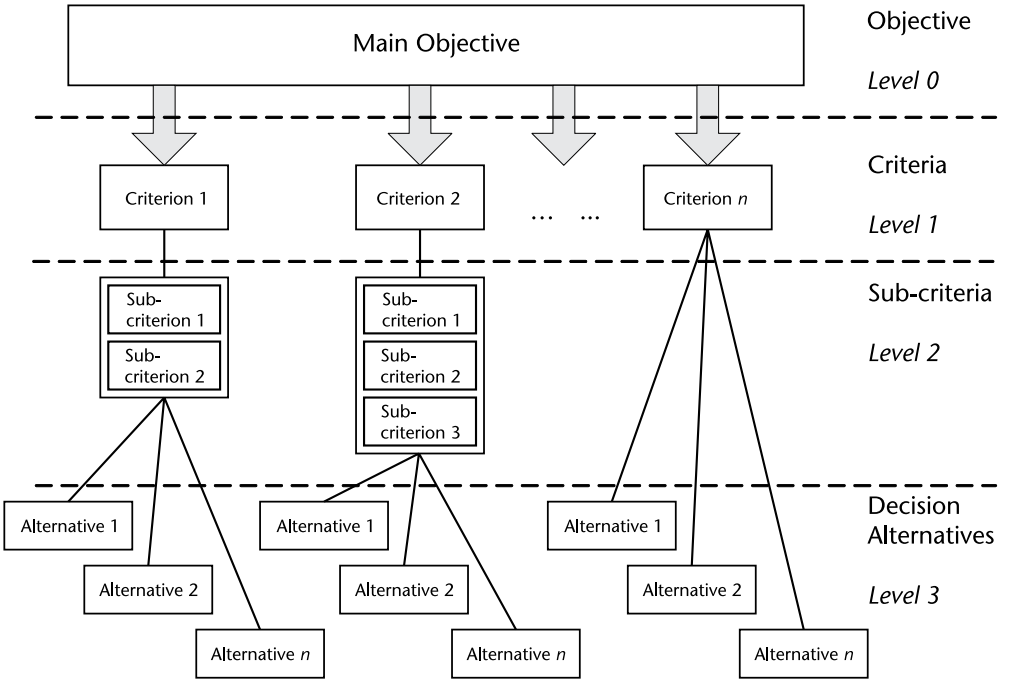


Figure 8.4 Analytic hierarchy process block diagram.

The analyst can insert and delete levels and elements as necessary to clarify the task or to focus on one or more areas of the system.

Here n elements are considered with the goal of providing and quantifying judgments on the relative weight (or importance) of each element with respect to all the other elements. The first step is to set the problem as a hierarchy, where the topmost node is the overall objective of the decision, while subsequent lower-level nodes consist of the criteria used in arriving to this decision. The alternatives, from which make a pick, are bottom level of the hierarchy (i.e., the n elements that are to be compared).

One of the major strengths of the AHP is the use of pairwise comparisons to derive accurate ratio scale priorities, as opposed to using traditional approaches of assigning weights, which can also be difficult to justify. According to Ishizaka and Labib [58], psychologists claim that it is easier and more accurate to express one's opinion on only two alternatives than simultaneously on many alternatives; thus, the pair-wise comparisons are made by answering the question: Of two elements i and j, which is more important (or larger) with respect to the given factor and how much more?

Tobias Dantzig (1884–1956), who was born in Latvia and studied mathematics with Henri Poincaré in Paris, in his book first published in 1930, claimed that most people (and even some animals) could compare two objects and determine that one is larger than the other [59][22]. Most people can easily compare two objects against each other (ability of human mind referred to as *fuzzy differentiation*, also

22. This sense for numbers should not be confused with counting, which involves a different and rather intricate mental process.

called *qualitative assessment*), but they might have difficulty in effectively comparing many objects in a pairwise fashion against multiple criteria.

One of the important features of the AHP is that it deals with opinion and hunch ("gut feeling") as well as other intangibles as easily as with facts; in other words, both qualitative and quantitative criteria can be compared using informed judgments to derive weights and priorities. Although other decision-makers can adopt other weights and even other criteria, the sources of disagreements can be clarified, debated, and negotiated in a more coherent manner by relying on an explicit model.

In the AHP, each element in the hierarchy is considered to be independent of all the others, meaning that the decision criteria are considered to be independent of one another, and the alternatives are considered to be independent of the decision criteria and of each other. However, in many real-world applications, there is interdependence among the elements and the alternatives. The *analytic network process* (ANP) is a more general form of the AHP or, in other words, the AHP is a special case of the ANP. It is also known as the systems-with-feedback approach, as described in [60] by Saaty.

Medical applications have been quite common during the last decade; for example, in [61] the performance of two medical treatment options, using AHP, was evaluated. A few years earlier, at the Hospital for Sick Children in Toronto, AHP was used in a pilot study to help determine who should receive an organ transplant [62]. Using pairwise comparisons, a set of criteria for children receiving organ transplants was evaluated using the AHP. The study was able to take into account ethical, qualitative, and quantitative factors to determine who should receive organ transplants.

The study determined that such factors as the patient's ability to pay[23], medical insurance, or financial or economic status should not be considered in making a transplant decision and the physical limitations, such as being disabled, should not be a determining factor for an organ transplant. The conclusion of the study was that the most important criteria would be the patient's ability to survive the difficult transplant process, accept the difficult transition process following an organ transplant, and lead a relatively normal life after the organ transplant.

One of very few applications of AHP dealing with ethical decision-making was published in [63]. This paper briefly described a simplified process of selecting the most ethical course of action, taking into consideration consequentialism, consistency, and duty. Consequentialism was subdivided into minimal harm and egoism, while duty was subdivided into be-true and right-to-know. Accommodating multiple stakeholders and modeling uncertainty were also addressed in the same paper.

8.4.7.3 Group Decision-Making

A group is two or more individuals who are connected to each other by some type of social relationship. *Group dynamics* is the study of groups and also a general term for group processes. The scientific study of groups began around the 1890s as part of the new growing field of social psychology. Because they interact and

23. This is most likely a result of Canada's national health care system, assuring health care for all Canadians.

influence each other, groups develop a number of dynamic processes that separate them from a random collection of individuals. These processes include norms, roles, relations, development, need to belong, social influence, and effects on behavior.

Group decision is usually understood as aggregating different individual preferences on a given set of alternatives to a single collective preference. It is assumed that the individuals participating in making a group decision face the same common problem and are all interested in finding a solution or an answer to that problem. A group decision situation usually involves multiple decision-makers, each with different skills, experience, and knowledge relating to different aspects (criteria) of the problem. Consensus forming is usually a critical part in the group decision-making process [64].

Not all the groups are wise and capable of making smart decisions; according to Dr. James Michael Surowiecki, an American journalist [65], these key criteria separate wise groups from irrational ones:

- *Diversity of opinion:* Each person should have his or her own opinion;
- *Independence:* People's opinions should not be determined or influenced by the opinions of those around them;
- *Decentralization:* People are able to specialize and draw on local knowledge;
- *Aggregation:* Some mechanism must exist for turning private judgments into a collective decision.

Usually if the group is working together in a decision-making session and some judgment or judgments seem to be on the opposite side from the majority, mathematical models and methods are of little help. If the final outcome of the analysis differs from the expectations, which are usually based on intuition and/or experience, the cause of deviation has to be researched. Neither intuition nor mathematical model is necessarily wrong, but perhaps some important factors are not considered or given sufficient attention [66].

It was said [67] that: "Consensus is good, unless it is achieved too easily, in which case it becomes suspect." In other words, some disagreement between parties involved is a good thing, since it can lead to a healthy discussion and improve odds for correct decision. When the group is not on the same page in terms of understanding the problem, or maybe they are on opposite sides of the issue, conflict resolution, negotiation, and facilitation are required.

Feedback among the participants in the decision-making process could be achieved by conducting a brainstorming, using the structured communications process known as a nominal group technique (NGT)[24] or applying a Delphi technique[25].

24. The nominal group technique is a decision-making method with the idea to make the decision quickly, but with everyone's opinions taken into consideration. First, every member of the group gives his or her view of the solution, with a short explanation. Then duplicate solutions are eliminated from the list of all solutions, and the members proceed to rank the solutions: first, second, third, fourth, and so on. This technique was originally developed by Delbecq and VandeVen in 1970s.
25. Delphi technique was developed by the research group at the RAND Corporation, Santa Monica, California, in 1969 to obtain the most reliable consensus of opinion from a group of individuals about a dilemma that had no objective resolution.

8.5 Review Questions and Problems

1. Last week, Edward B. Someone tried out a brand-new wireless LAN card on his laptop at work. He did not expect anything to happen, because his organization's wireless LAN was not up and running yet. But to his surprise, he was able to connect without any trouble to the network of an office down the street. Oops!

   Discuss the fact that space around us is full of radio frequencies containing data and proprietary, as well as personal, information. Was Edward committing a crime by accessing someone else's network? Discuss the different scenarios and outcomes.
2. Think about and discuss the following statement:
   "How would you like it if, for instance, one day you realized your underwear was reporting on your whereabouts?" (California State Democratic Senator Debra Bowen, at a 2003 hearing)

   With what other famous (and erroneous) statements in the history of science and technology can you compare this statement? Two similar examples are given here:
   - "There is no reason anyone would want a computer in their home." (Ken Olson, president, chairman and founder of Digital Equipment Corporation, 1977)
   - "640k memory ought to be enough for anybody," (Bill Gates, Microsoft founder, 1981)
3. One concern of RFID technology is that UHF tags can be read wirelessly up to distances of 30 feet (HF tags can be read at less distance, but still wirelessly). Tags can obviously be removed and destroyed after purchase, but that makes returning a product or recalls more difficult. The Gen 2 protocol offers a kill command to deactivate tags, but tags that are killed cannot be revived.

   One novel idea is to use a so-called *clipped tag* that gives consumers the option of privacy protection by allowing them to tear off perforated sections that hold parts of the tags' antennas, reducing their read ranges to only a few inches. This provides a visible means of enhancing privacy protection and allows for later use of the tag for returns and recalls. Discuss this idea and list its pros and cons.
4. Does RFID enable tracking people by satellite? Do you think that someone could covertly read the contents in a shopping bag?
5. Consumers have concerns about the privacy and use of their personal information. Many of the companies and organizations promoting the use of RFID have or are developing policies addressing consumer notice and privacy. However, it is important to remember many applicable consumer protections are already written into law. For example, retailers are already restricted in the sale or distribution of consumer information, and secure computer systems and data encryption schemes are already in place for the electronic transfer of private data.

   There are already federal guidelines in place that address many of these concerns, including the Privacy Act (1974), the Electronic Communications

Privacy Act (1986), the Telecommunications Act (1996), Health Insurance Portability and Accountability Act (1996), and the Financial Modernization Act (Gramm-Leach-Bliley Act of 2000), among others. Do you think that we need new, RFID-specific, legal protections?

6. AIM Global has instituted an image called the *RFID emblem* that acts as a visual indicator to consumers and retail workers to help them find and identify the presence (and type) of RFID tag in a label, tag, or item. Some think that this is a sufficient self-regulation of the industry; what is your opinion? What other measures, if any, would you suggest in order to further protect consumers privacy?
7. Besides consumer goods, RFID technology can be beneficial in other applications (preventing drug counterfeiting and reducing tampering, alerting staff to wandering patients in retirement homes, locating lost children in public places, and finding stray animals). Do you believe these positive uses can counter skepticism about the privacy invasion of RFID? Explain your answer.
8. Here are a few examples, mentioned earlier in this chapter, of systems/programs in the United States that were implemented without the consent of the individual(s) involved:
   - The DoD's mandatory vaccinations against the biological germ weapon called anthrax;
   - Several mandatory vaccination programs throughout U.S. history;
   - President Franklin Roosevelt's signing of the Selective Training and Service Act of 1940, which created the country's first peacetime draft.

   These examples stirred a highly emotional debate on ethics among many people and a few lessons learned can be gathered from these mandatory programs.

   Speculation continues to grow as to which groups of the human population will have microchips introduced into their bodies to test the feasibility of the concept and for further future implementation. Some have eluded to the fact that eventual human microchip implantation is coming and is possible. These people believe that it will first be on a volunteer basis and then the government will intervene, making it mandatory in the penal system and the military and later in the general public as well. Discuss the topic in detail.
9. It has been long predicted that RFID and sensors would be combined, whereby a sensor gathers environmental information that is stored on an accompanying RFID tag. In February 2007, the Web site http://www.NewScientist.com discovered a recently filed patent application by Kodak that outlines a new application for RFID ingestible tags that act as monitors for health characteristics within the human body. The idea is that the RFID tag antenna could be composed of organic material that would dissolve as a result of certain chemical reactions within the human body. Once dissolved, the tag antenna, and therefore the tag itself, would stop transmitting a signal, indicating that the targeted chemical reaction had occurred. Kodak calls them *fragile tags*.

For example, imagine an RFID reader-equipped drug dispenser installed in the home bathroom of a patient. The patient is prescribed to take a pill every day, which is issued by the dispenser. Once the pill is issued, the dispenser's RFID reader activates and begins polling for the signal of a tag, which is physically attached to the dispensed pill. In this uningested state, the tag functions properly, responding to the RFID reader's interrogation, which, in turn, informs the dispenser that the day's dosage has not yet been taken.

Once the patient ingests the pill/tag, the organic tag antenna is subjected to chemicals within the patient's stomach. The tag antenna was designed to rapidly dissolve in the presence of normal stomach chemicals, so after only a few minutes it does so, and the tag ceases to respond to the RFID reader signal, which the dispenser interprets as the patient having taken his or her daily medication.

The concept could be applied to changes in mechanical states as well. Discuss an application in which a tag may be affixed to an artificial or natural body part and when worn on the body part, for example, an artificial hip, has proceeded to a predetermined level, the tag is rendered useless, thus alerting the remote query that the body part has achieved an unsatisfactory level of wear.

10. Could RFID ever lead to massive layoffs of workers?
11. Analyze and discuss the following statement, taken from [68]:

> The Smart Human Environment will completely change the way we interact with our environment and with each other. People will communicate with their technological environment naturally, using a variety of modalities and devices. The environment will be aware and will understand the user's social, physical and situational context. The environment will be able to smartly assist the user in his tasks, based on this context awareness and knowledge of the user's behavioral profile as well as common sense knowledge. It will exhibit pro-active behavior for recurring tasks and provide personalized information services.

12. What are some of the most important ethical issues engineers could face during the development of a medical device? Should engineers be concerned about ethical issues regarding medical implants they are working on? If not, who should be the one to analyze the problem and make a decision on how to proceed? If yes, how should engineers approach this dilemma? Please elaborate your answers.
13. Many may regard the employment of neural implants as a technological progress, accept the new technology, and welcome it, while others may be horrified just thinking about it. People from different cultures and ethnical backgrounds may also respond differently. Discuss the opinion of the society where you live and their possible reaction to this very radical, new technology.
14. Who will drive forward the technological development of wireless body implants, and who will control its implementation? Should it be left in the hands of commercial concerns, which could easily be dangerous to society, or should political entities be responsible? Should military and/or

government deployment be regulated and overseen? An issue of misuse by large corporations and criminals comes to mind as well. Discuss the topic.

15. Research different methods for the conflict resolution in a group decision-making situation. What would the selection of the appropriate method depend on? What are the factors we have to consider when deciding who will be involved in the making the final decision regarding the problem at hand?
16. The development of ethics for implanted devices is critical as the technological advances in implantable devices will soon produce a huge number of patients with implanted devices in an end-of-life situation. Who has the right to turn off the device; does the doctor have the right to turn it off, against the patient's wishes, for reasons of futility; does the patient have the right to turn it off, against the doctor's wishes, and if so, for what reasons? Who else could be authorized to turn off the implanted devices?
17. Why do you think that it is important that the chief ethics officer in an organization reports directly to the CEO (but not the CFO) or the board's audit committee? Where do you see conflict of interests arising?

References

[1] Ohkubo, M., et al., "RFID Privacy Issues and Technical Challenges," *Communications of the ACM,* Vol. 48, No. 9, September 2005.

[2] Garfinkel, S., et al., "RFID Privacy: An Overview of Problems and Proposed Solutions," *IEEE Computer Society*, 2005.

[3] Federal Trade Commission Report, *Radio Frequency Identification Applications and Implications for Consumers*, March 2005.

[4] "Possible Health Risks to the General Public from the Use of Security and Similar Devices," *International Commission on Non-Ionizing Radiation Protection*, 2002.

[5] Lindqvist, P., "RFID Monitoring of Healthcare Routines and Processes in Hospital Environment," M.S. Thesis, Master's Programme in Bioinformation Technology, Helsinki University of Technology, Department of Electrical and Communications Engineering, August 10, 2006.

[6] Galvan, J., "On Technoethics," *IEEE-RAS Magazine*, Vol. 10, 2003-2004, pp. 58–63.

[7] Zerbe, W., et al., (eds.), *Emotions, Ethics and Decision-Making*, Bingley, U.K.: Emerald Group Publishing, 2008.

[8] Satava, R. M., "Biomedical, Ethical, and Moral Issues Being Forced by Advanced Medical Technologies," *Proceedings of the American Philosophical Society*, Vol. 147, No. 3, September 2003.

[9] Schweizer, A., "Kultur und Ethik," 1996 reprint of an original work published in 1923, Verlag C.H. Beck, 1996.

[10] DeGeorge, R. T., *Business Ethics*, 4th ed., Upper Saddle River, NJ: Prentice-Hall, 1995.

[11] Bennett-Woods, D., "Ethics at a Glance," Regis University, Rueckert-Hartman School for Health Professionals, 2005.

[12] Carroll, A., and A. Buchholtz, *Business and Society: Ethics and Stakeholder Management*, 7th ed., Cincinnati, OH: South-Western College Publishing, 2008.

[13] http://plato.stanford.edu/entries/egoism/ (accessed December 17, 2010).

[14] http://webs.wofford.edu/kaycd/ethics/egoism.htm (accessed December 17, 2010).

[15] http://www.iep.utm.edu/bentham/ (accessed December 17, 2010).

[16] http://www.naturalrights.us/, Fuerle, Richard D., "Natural Rights: A New Theory" (accessed December 17, 2010).

[17] http://civilisationis.com/cooray/btof/index.htm, (accessed, December 2011).

[18] Battye, J., et al., *Ethical Issues of New and Emerging Technologies*, Report No. 104, ISSN 1171-0101, Published by the Ministry of Research, Science and Technology, New Zealand, October 1999.

[19] http://caae.phil.cmu.edu/Cavalier/80130/part2/sect8.html (accessed December 17, 2010).

[20] Katz, J., "The Consent Principle of the Nuremberg Code: Its Significance Now and Then," in G. J. Annas and M. A. Grodin, (eds.), *The Nazi Doctors and the Nuremberg Code: Human Rights in Human Experimentation*, New York: Oxford University Press, 1992, pp. 231–233.

[21] http://www.hhs.gov/ohrp/45CFRpt46faq.html (accessed July 17, 2010).

[22] The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *Report and Recommendations: Research Involving Prisoners*, *DHEW Publication,* No. (OS) 76-131, 1976.

[23] Haynes, D. L., "Implications of User Identification Devices (UIDS) for the United States Navy," Master Thesis, Naval Postgraduate School, September 2001.

[24] Oram, D., "Designing for Sustainability: Negotiating Ethical Implications," *IEEE Technology and Society Magazine*, Vol. 29, No. 3, Fall 2010.

[25] Horniak, V., "Privacy of Communication: Ethics and Technology," Department of Computer Science and Engineering, Mälardalen University, Master Thesis in Computer Engineering, 2004.

[26] Maxwell, J. C., *There Is No Such Thing as Business Ethics*, New York: Warner Business Books, 2003.

[27] Moriarty, G., "Three Kinds of Ethics for Three Kinds of Engineering," *IEEE Technology and Society Magazine,* Fall 2001.

[28] Mitcham, C., "A Historico-Ethical Perspective on Engineering Education: From Use and Convenience to Policy Engagement," *Engineering Studies,* Taylor and Francis Group, Vol. 1, No. 1, March 2009, pp. 35–53.

[29] *Routledge Encyclopedia of Philosophy*, Version 1.0, London and New York, 1998.

[30] Mitcham, C., (ed.), *Encyclopedia of Science, Technology, and Ethics*, Farmington Hills, MI: Thomson Gale, a part of The Thomson Corporation, 2005.

[31] Organisation for Economic Co-operation and Development, "21st Century Technologies—Promises and Perils of a Dynamic Future," *EXPO 2000*, 2000.

[32] NSF/DOC-sponsored report, *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, Arlington, VA, 2002.

[33] Government of Canada, *Toward Understanding Science and Technology Convergence*, Science & Technology Foresight Directorate, Office of the National Science Advisor, Privy Council Office, September 2005.

[34] Agar, N., *Liberal Eugenics: In Defense of Human Enhancement*, Cambridge, MA: Blackwell Publishing, 2004.

[35] Satava, R. M., "Biomedical, Ethical, and Moral Issues Being Forced by Advanced Medical Technologies," *Proceedings of the American Philosophical Society,* Vol. 147, No. 3, September 2003.

[36] McGee, E. M., and G. Q. Maguire, Jr., "Ethical Assessment of Implantable Brain Chips," *Twentieth World Congress of Philosophy*, Boston, MA, August 10–15, 1998.

[37] http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0000166 (accessed September 3, 2010).

[38] http://www.af.mil/news/story.asp?id=123007615 (accessed September 3, 2010).

[39] Morales, E., "Wireless Neural Implants," Raytheon, SAS, Software Engineering Center, 2008.

[40] Eugene B Wu, "The Ethics of Implantable Devices," *J. Med. Ethics*, 2007.

[41] Frankel, M. S., and C. J. Kapustij, "Enhancing Humans," The Hastings Center, 2008.

[42] Savulesku, J., and N. Bostrom, *Human Enhancement*, New York: Oxford University Press, 2008.

[43] McGee, E. M., and G. Q. Maguire, Jr., "Implantable Brain Chips: Ethical and Policy Issues," *Medical Ethics,* Winter 2001.

[44] Moreno, J. D., *Mindwars: Brain Research and National Defense*, Washington, D.C.: Dana Press, 2006.

[45] Hall, J. S., "Ethics for Machines," 2000, http://autogeny.org/ethics.html.

[46] Rawls, J., "Outline for the Procedure for Ethics," *Philosophical Review*, Vol. 60, No. 2, April 1951.

[47] Hosmer, L. T., *The Ethics of Management*, Homewood, IL: Richard D. Irwin, 1987.

[48] http://wings.buffalo.edu/bioethics/man-comp.html (accessed September 28, 2010).

[49] www.asbh.org (accessed July 15, 2010).

[50] McKibben, B., *Enough*, New York: Henry Holt, 2003, p. 205.

[51] Bostrom, N., "In Defense of Posthuman Dignity," *Bioethics,* Vol. 19, No. 3, 2005, pp. 202–214.

[52] Winner, L., "Artifact/Ideas and Political Culture," *Whole Earth Review,* No. 73, Winter 1991, pp. 23–24.

[53] Chankong, V., and Y. Haimes, *Multiobjective Decision Making: Theory and Methodology*, North-Holland Series in System Science and Engineering, Amsterdam, Netherlands: North-Holland Publishing, 1983.

[54] Umiker, W. O., "Intuitive Decision Making and Problem Solving," *Medical Laboratory Observer*, April 1989.

[55] Cho, K. T., "Multicriteria Decision Methods: An Attempt to Evaluate and Unify," School of Systems Management Engineering, Sungkyunkwan University, Suwon, Korea, 1995.

[56] Proctor, W., and M. Drechsler, "Deliberative Multi-Criteria Evaluation: A Case Study of Recreation and Tourism Options in Victoria, Australia," *European Society for Ecological Economics, Frontiers 2 Conference*, Tenerife, 2003.

[57] Saaty, T. L., *The Analytic Hierarchy Process*, New York: McGraw-Hill, 1980.

[58] Alessio, I., and L. Ashraf, "Analytic Hierarchy Process and Expert Choice: Benefits and Limitations," *ORInsight*, Vol. 22, No. 4, 2009, pp. 201–220.

[59] Dantzig, T., *Number: The Language of Science*, New York: Pi Press, 2005.

[60] Saaty, T. L., *Decision Making with Dependence and Feedback: The Analytic Network Process*, Pittsburgh, PA: RWS Publications, 2001.

[61] Hummel, M., et al., "A Multicriterial Decision Analysis of Augmentative Treatment of Upper Limbs in Persons with Tetraplegia," *Journal of Rehabilitation Research & Development*, Vol. 42, No. 5, 2005, pp. 635–644.

[62] Koch, T., et al., "A Pilot Study on Transplant Eligibility Criteria," *Pediatric Nursing*, March 13, 1997, pp. 160–162.

[63] Millet, I., "Ethical Decision Making Using the Analytic Hierarchy Process," *Journal of Business Ethics*, Vol. 17, 1998, pp. 1197–1204.

[64] Choudhury, A. K., "Consensus-Based Intelligent Group Decision-Making Model for the Selection of Advanced Technology," *Decision Support Systems*, 2005.

[65] Surowiecki, J., *The Wisdom of Crowds*, New York: Anchor, 2005.

[66] Saaty, L. T., and K. Peniwati, *Group Decision Making: Drawing Out and Reconciling Differences*, Pittsburgh, PA: RWS Publications, 2008.

[67] Buchanan, L., and A. O'Connell, "A Brief History of Decision Making," *Harvard Business Review*, 2006.

[68] Sipilä, M., (ed.), "Communications Technologies, The VTT Roadmaps," Helsinki, Finland: *VTT Research Notes*, 2146, ESPOO 2002, 2002.

Appendix

Common Unit Conversion

1 mile = 1.609 km = 5,280 feet = 63,360 inches

1m = 1.0936 yards = 3.2808 feet = 0.001 km = 6.2137×10^{-4} mile = 39.3701 inches

1 foot = 12 inches = 0.3048m = 0.3333 yard

1 $m^2 = 10^6$ mm^2 = 1,550 in^2 = 10.7639 ft^2 = 1.1960 yd^2

1 $m^3 = 10^9$ mm^2 = 61,023.7 in^3 = 1.30795 yd^3 = 35.3147 ft^3

1 km/hr = 0.27778 m/s = 0.6214 mile/hr = 3.281 ft/s

1 mile/hr = 1.6093 km/hr

1 lb = 0.4536 kg

1 kg = 2.2046 lb

1N = 1 mkg/s^2 = 0.1020 kp = 0.2248 lbf

1 $N/m^2 = 10^{-5}$ bar = 0.0209 lbf/ft^2 = 0.1020 kp/m^2 = 9.8692×10^{-6} atm

$$°C = \frac{5}{9} \cdot (°F - 32)$$

$$°F = \frac{5}{9} \cdot °C + 32$$

1 Henry = 1H = 1 Vs/Am

1 Farad = 1F = 1 As/V = 1 Ωs

1 Tesla = 1T = 1 $Vs/m^2 = 10^4$ Gauss

1 Weber = 1 Wb = 1 Vs

1T = 1 Wb/m^2

1 Maxwell = 10^{-8} Weber

Frequency and Wavelength

$$\text{Wavelength: } \lambda = \frac{c}{f}$$

where

f = frequency [Hz]

c = speed of light in vacuum ($c \approx 3\times10^8$ m/sec or 9.84×10^8 ft/sec)

Basic Rules of Exponents

$$a^m a^n = a^{m+n} \qquad \left(\frac{a}{b}\right)^n = \frac{a^n}{b^n} \qquad \frac{a^m}{a^n} = a^{m-n}$$

$$(ab)^m = a^m b^m \qquad \left(a^m\right)^n = a^{mn}$$

$$a^{-n} = \frac{1}{a^n} \qquad a^n = \frac{1}{a^{-n}}$$

$$a^{\frac{m}{n}} = \sqrt[n]{a^m}$$

Basic Rules of Logarithms

$$\log ab = \log a + \log b$$

$$\log \frac{a}{b} = \log a - \log b$$

$$\log a^n = n \log a$$

$$\log \frac{1}{a} = -\log a$$

$$a = x^b \Rightarrow b = \log_x a$$

x is the base of a logarithm and base 10 is most commonly used in radio communications. Base 2 and the base of natural logarithms e are used sometimes as well.

Complex Numbers

A complex number, $Z = R + jX$, has a real part R, and imaginary part X. Here, j is the imaginary unit, $j = \sqrt{-1}$. In mathematics, i is used for the imaginary unit, but in electrical engineering j is used in order to avoid confusion with the alternating current, whose symbol is also i. Complex numbers can be shown in a two-dimensional complex plane.

Vectors and Vector Operations

Vectors are three-dimensional quantities. The *dot product* of two vectors **A** and **B** can be defined as follows:

$$\mathbf{A} \bullet \mathbf{B} = |\mathbf{A}||\mathbf{B}|\cos\theta$$

If the angle θ is zero, **A** and **B** are in parallel and the dot product has a maximum value. If the angle θ is 90°, **A** and **B** are orthogonal and the dot product is equal to zero.

The cross-product of two vectors **A** and **B** can be defined as follows:

$$\mathbf{A} \times \mathbf{B} = \hat{n}|\mathbf{A}||\mathbf{B}|\sin\theta$$

The resulting vector is orthogonal to **A** and **B**, and $\hat{n}$ is a *unit vector* normal to the plane in which **A** and **B** reside. If the angle θ is 0° or 180°, meaning that **A** and **B** are parallel, the cross-product is zero. If the angle θ is 90°, meaning that **A** and **B** are orthogonal, the cross-product of these two vectors has a maximum value.

Decibels

The decibel unit is commonly used in radio communications because the direct relationship between radio-related power levels covers a wide range of numerical values. The logarithmic nature of the relationship between two power levels results in values that are easy to handle.

The abbreviation for decibel, *dB*, has a capital B since a bel was derived from Alexander Graham Bell's last name. Addition or subtraction operations can be easily performed on logarithmic values, simplifying the handling of amplification and attenuation. In addition, humans perceive differences in the sensory impressions of varying intensity in a logarithmic fashion.

The decibel is used to compare one power (or voltage level) to another:

$$\text{Gain (or Loss) in decibles} = 10\log_{10}\frac{P_2}{P_1}$$

where

P_1 = input power
P_2 = output power

A 3-dB gain represents a doubling of power, while a –3-dB loss represents half of the power (see Table A.1).

Table A.1 Power Gain and Loss in Decibels

dB Gain (Power)	*Factor*	*dB Loss (Power)*	*Factor*
0 dB	1 (the same)	0 dB	1 (the same)
1	1.25	–1	0.8
3	2	–3	0.5
6	4	–6	0.25
10	10	–10	0.10
12	16	–12	0.06
20	100	–20	0.01
30	1,000	–30	0.001
40	10,000	–40	0.0001

dBm and dBW

dBm is a power relative to 1 mW. A 100-mW signal is equivalent to a 20-dBm signal.

$$P_{dBm} = 10\log_{10}\frac{P_{mW}}{1mW}$$

dBW is a power relative to 1W:

$$P_{dBW} = 10\log_{10}\frac{P_{W}}{1W}$$

dBm units can be converted into dBW units by subtracting 30 dB:

$$P_{dBW} = P_{dBm} - 30$$

so, for example:

0 dBm = 1 mW
30 dBm = 1W
30 dBm = 0 dBW
–30 dBm = 0 dBm
40 dBm = 10 dBW

A word of caution about dBm: *We cannot add dBm to dBm.*

The two powers must first be converted to milliwatts and then added and the sum reconverted to dBm. For example:

$$P_{total} = 10\,dBm + 10\,dBm = ?$$
$$\text{since } 10\,dBm = 10\,mW$$
$$10\,mW + 20\,mW = 20\,mW$$
$$P_{total} = 10\log_{10}\left(\frac{20}{1}\right) = 13\,dBm$$

We know that doubling the power means 3-dB increase in decibels. Thus, as a shortcut, we could just add 3 dB to 10 dBm and get a resulting 13 dBm.

In case we are adding different powers, for example, 15 dBm and 20 dBm, there is no way around and we have to use the conversion. The result is:

$$15\,dBm + 20\,dBm = 10\log\left(10^{\frac{15}{10}} + 10^{\frac{20}{10}}\right) \approx 21.2\,dBm$$

Signal-to-Noise Ratio

The Signal-to-noise Ratio (S/N or SNR) is the amount by which a signal level exceeds the noise level:

$$S/N_{dB} = 10\log_{10}\frac{Signal\,Level_{mW}}{Noise\,Level_{wW}} = Signal\,Level_{dBm} - Noise\,Level_{dBm}$$

EIRP

Effective Isotropic Radiated Power (EIRP) describes the performance of a transmitting system:

$$EIRP_{dBm} = Tx\,Power_{dBm} + Antenna\,Gain_{dBi} - Line\,Loss_{dB}$$

Fade Margin

Fade Margin (FM) is an "extra" signal power added to a link to ensure its continued operation if it suffers from signal propagation effects. Antenna gain and hardware losses (cables, connectors, branching units, and so on) on both ends are added together.

$$FM_{dB} = System\,Gain_{dB} + Antenna\,Gain_{dBi} - Propgation\,Losses_{dB} - Hardward\,Losses_{dB}$$

System Gain

System gain is the total gain of the radio system without considering antennas and cables:

$$System\,Gain_{dB} = Tx\,Power_{dBm} - Rx\,Threshold\,Sensitivity_{dBm}$$

Free-Space Loss

Free-space path loss is the signal energy lost in traversing a path in free space only, with no other obstructions or propagation issues. It can be expressed with Friis' formula in logarithmic form:

$$FSPL_{dB} = 96.6 + 20\log_{10}\left(d_{miles}\right) + 20\log_{10}\left(f_{GHz}\right)$$
$$FSPL_{dB} = 92.4 + 20\log_{10}\left(d_{km}\right) + 20\log_{10}\left(f_{GHz}\right)$$

Glossary

Access control An RFID application in which RFID tag–equipped badges are used to provide secured access to a facility.

Active tag A type of RFID tag that contains an internal power source and, in some cases, also a radio transceiver. These additional component(s) are used to enhance the effective read/write range, and rate of data transfer characteristics of the RFID tag.

Adaptive frequency agility Technique that allows an interrogator to change its frequency of operation automatically from one channel to another.

AIM Global The worldwide trade association of components and systems providers for automatic identification, data collection, and data integration in management information systems. Its members are manufacturers or service providers of identification technologies, such as RFID, bar code, smartcard, and biometrics.

Amplitude modulation (AM) Representation of data or signal states by the amplitude of a fixed-frequency sinusoidal carrier wave. When data is in binary form, the modulation involves two levels of amplitude and is referred to as amplitude shift keying.

Amplitude shift keying (ASK) Representation of binary data states, 0 and 1, by the amplitude of a fixed-frequency sinusoidal carrier wave. When the amplitudes are determined by the carrier being switched on and off, the process is known as on-off keying (OOK).

Antenna A conductive structure specifically designed to couple or radiate electromagnetic energy. In a driven mode the structure is a transmitter antenna. In receiver mode the structure is a receiver antenna. Antenna structures, often encountered in RFID systems, may be used to both transmit and receive electromagnetic energy, particularly data-modulated electromagnetic energy.

Antenna gain The measure of the amount of signal that the antenna radiates or receives. It is given as a decibel ratio, compared to a theoretical omnidirectional antenna called an isotropic antenna. All other things being equal, a high-gain antenna will transmit and receive weaker signals farther than a low-gain antenna. Omnidirectional antennas, such as a dipole, will have a lower gain than directional antennas because they distribute their power over a wider area. A half-wave dipole antenna will have a gain of near 1, or nearly equal the isotropic antenna.

Anticollision (anticontention) A facility for avoiding contention at the reader/interrogator receiver for responses arising from transponders simultaneously present within the read or interrogation zone of an RFID system and competing for attention at the same time.

Anticollision capability An RFID technology characteristic that allows for multiple RFID tags to be identified while present in an RF portal.

Application identifier (AI) A metadata element used to define the meaning of the data that follows.

Assigned frequency band Frequency band within which the emission by a device is authorized.

Auto-ID Labs Currently headquartered at the Massachusetts Institute of Technology (MIT) in Cambridge, Massachusetts, and further based at six other leading universities worldwide: the University of Cambridge in the United Kingdom; the University of Adelaide in Australia; Keio University in Tokyo, Japan; Fudan University in Shanghai, China; the University of St. Gallen in Switzerland; and the Information and Communications University (ICU) in Daejeon, Republic of Korea.

Backscatter modulation A process whereby a transponder responds to a reader/interrogation signal or field by modulating and reradiating or transmitting the response signal at the same carrier frequency.

Bandwidth The range or band of frequencies, defined within the electromagnetic spectrum, that a system is capable of receiving or delivering.

Batch reading Capability of an RFID reader/interrogator to read a number of transponders present within the system's interrogation zone at the same time. This is an alternative term for *multiple reading*.

Battery-assisted tag Transponder that includes a battery to enhance its receive performance and power its internal circuitry.

Batteryless tag Transponder that derives all of the power necessary for its operation from the field generated by an interrogator.

Battery-powered tag Transponder that uses the power from its battery to perform all of its operational functions (*see* Passive tag).

Baud A unit of signaling or transmission speed representing the number of signaling events per unit time. When the signal event is a single-bit, binary state representation, the baud is equivalent to the bit rate, expressed in bits per second (bps).

Biocompatibility Acceptance of an artificial implant by surrounding tissues and by the body as a whole. The implant should be compatible with tissues in terms of mechanical, chemical, surface, and pharmacological properties.

Biomedical telemetry device An intentional radiator used to transmit measurements of either human or animal biomedical phenomena to a receiver.

Biometrics The biological identification of a person, including characteristics of structure and of action such as iris and retinal patterns, hand geometry, fingerprints, voice responses to challenges, and dynamics of handwritten signatures. Biometrics

are a more secure form of authentication than using cards or typing passwords; however, some forms have relatively high failure rates. Biometric authentication is often a secondary mechanism in two-factor authentication (two-factor authentication is a system wherein two different methods are used to authenticate; using two factors as opposed to one implies a higher level of security).

Bluetooth A radio technology developed by Ericsson and other companies built around a new chip that makes it possible to transmit signal over short distances between phones, computers, and other devices without use of wires. More information is available at http://www.bluetooth.com.

Broadband A classification of the information capacity or bandwidth of a communication channel. It is generally taken to mean a bandwidth higher than 2 Mbps.

Capture field/area/zone (also interrogation zone/area/volume) The region of the electromagnetic field, determined by the reader/interrogator antenna, in which the transponders are signaled to deliver a response.

Collision An event in which two or more data communication sources compete for attention at the same time and cause a clash of data, inseparable without some means of anticollision or contention management.

Contention (clash) Simultaneous transponder responses capable of causing potential confusion and misreading within a reader/interrogator system unequipped with anticontention facilities.

Dense reader (interrogator) mode The situation when the number of readers operating is large in comparison to the number of available channels (e.g., 20 readers operating in 20 available channels).

De-tuning The change in the performance of transponders and readers caused by the presence of metal or ferromagnetic materials.

Electronic article surveillance (EAS) An RFID application in which the exit of items out of some physical environment is monitored electronically; typically used for theft and loss prevention in the retail industry.

Electronic Industries Association (EIA) An association that specifies electrical transmission standards, including those used in networking.

Electronic Product Code (EPC) The EPC numbering system uniquely identifies objects and facilitates tracking throughout the product's life cycle.

Encryption The reversible transformation of data from the original (the plaintext) to a difficult-to-interpret format as a mechanism for protecting its confidentiality, integrity, and sometimes authenticity; uses an encryption algorithm and one or more encryption keys.

EPCglobal Inc. A nonprofit corporation jointly established in September 2003 by European Article Numbers (EAN) International, a European distribution standardizing organization, and Uniform Code Council, Inc. (UCC), an organization responsible for the distribution and management of bar codes. EPCglobal is leading the development of industry-driven standards for the EPC to support the use of RFID.

European Article Numbering (EAN) system The international standard bar code for retail food packages.

European Telecommunications Standards Institute (ETSI) A body formed by the European Commission in 1988, which included vendors and operators. ETSI's purpose is to define standards that will enable the European market for telecommunications to function as a single market.

Federal Communications Commission (FCC) The U.S. government agency responsible for allocation of radio spectrum for communication services that regulates interstate communications: licenses, rates, tariffs, standards, and limitations. In Canada, the same function is conducted by Industry Canada.

Field strength The strength of the electromagnetic signal at a specified distance from the transmitting antenna. The legal field strength limits vary with country. Units of measurement include milliamps per meter (mA/m), millivolts per meter (mV/m), decibel microvolts per meter (dBμV/m), decibel microamperes per meter (dBμA/m), microvolts per meter (μV/m), and microamps per meter (μA/m).

Frequency modulation (FM) Representation of data or signal states by using different transmission frequencies; where data is in binary form the modulation constitutes two transmission frequencies and is referred to as frequency shift keying (FSK).

Gen 2 The Generation 2 standard from EPCGlobal that defines the coding and air interface for UHF tag operation.

Group selection A mode of operation whereby an interrogator can search for and identify unique tags within an RF portal or an RF field of view.

International Standardization Organization (ISO) A global network of the national standards institutes of 150 countries, on the basis of one member per country. The organization has developed more than 13,000 international standards, including ISO 9000, ISO 14000, ISO 18000, and so forth.

Interrogator A device that is used to read and/or write data to RFID tags; another name for an RFID reader.

ISO 14443 A four-part international standard for contactless smart cards operating at 13.56 MHz in close proximity with a reader antenna with a read range distance up to 10 cm. The advantage products utilizing ISO 14443 have over those utilizing ISO 15693 is that the transaction speed is faster, making security and transaction speed superior for large packets of information such as biometric templates; 14443A has grown to be the leading standard for access control and transportation and 14443B for banking.

ISO 15693 International standard regulating contactless, vicinity technology, typically representing a distance over 10 centimeters. The advantage that ISO 15693 has over ISO 14443 is greater convenience due to longer read ranges and less power consumption.

International Telecommunication Union (ITU) Created in 1934 to facilitate global cooperation in the development of all forms of telecommunication. It inherited, combined, and enlarged the functions of the International Telegraphic Union (formed in 1865) and the International Radiotelegraphic Union (formed in 1908). In 1947 the ITU became a specialized agency of the United Nations. Its legal "charter" is the International Telecommunications Convention, a treaty ratified by 191 countries that thereby joined the ITU. The ITU's Constitution is part of the convention.

Item management Processes for the identification, tracking, and tracing of goods or items that are being manufactured, stored, transported, or discarded.

Label Sometimes called an inlay; a tag which is thin and flexible.

Listen-before-talk (LBT) Under European regulations, a reader must first listen on a particular channel for other signals before it is allowed to transmit; otherwise, it must select another channel. In addition, every 4 seconds the reader must release the channel for 0.1 second to allow other readers a chance to occupy that channel (also known as *listen before transmit*).

Manchester coding A biphase code format in which each bit in the source-encoded form is represented by 2 bits in the derived or channel encoded form. The transformation rule ascribes 01 to represent 0 and 10 to represent 1.

Manufacturers Tag ID (MfrTagID) A reference number that uniquely identifies the tag.

Modulation The methods of altering the signal between reader and transponder, in order to carry the encoded information are quite varied. In some cases, the modulation can be different between the reader and the transponder, and between the transponder and the reader. Some of the methods used are amplitude modulation (AM), phase modulation (PM), frequency modulation (FM), pulse width modulation (PWM), and continuous wave modulation (CW).

Multitechnology reader A reader utilizing two or more technologies, such as proximity (125 kHz) and contactless (13.56 MHz).

Near field An operating specification for an RFID tag to be near or in close proximity to an interrogator's antenna. Near-field capable interrogators and corresponding RFID tags typically have a read/write range of 4–6 inches.

Near-field communication (NFC) A standards-based, short-range wireless connectivity technology that enables simple and safe two-way interactions among electronic devices, developed in conjunction between Philips and Sony to compete with Bluetooth wireless communication.

Passive tag A type of RFID tag that does not contain an internal power source. This type of tag design is less complex and is usually of a single or dual chip design. It is said to be "beam powered" using the electromagnetic energy of an interrogator.

Pulse oximetry A simple noninvasive method of monitoring the percentage of hemoglobin, which is saturated with oxygen. Pulse oximeters are now a standard part of perioperative monitoring, which gives the operator a noninvasive indication of the patient's cardiorespiratory status.

Radiation resistance Portion of the antenna's impedance that results in power radiated into space (that is, the effective resistance that is related to the power radiated by the antenna). Radiation resistance varies with antenna length. Resistance increases as the wavelength increases.

RF absorption A radio phenomenon that occurs when transmitted RF signal energy is consumed by some material in the pathway of the RF transmission.

RF cancellation A radio phenomenon that occurs where a transmitted RF signal is neutralized by competing RF interference.

RF portal A defined physical area of RF signal presence. Also known as an RF depth of field, and or physical RF field of view.

RF reflection A radio phenomenon that occurs when a transmitted RF signal is echoed off of another RF radiator placed within the pathway of the RF transmission.

RFDC An implementation of automated data collection whereby portable ADC reader devices are connected to a host computer via RF so that interactive data transfers can occur.

RFID A method of storing and retrieving data via electromagnetic transmission to an RF-compatible integrated circuit.

RFID carrier frequency A defined RF to transmit and receive data; RFID frequencies include 2.45 GHz, 915 MHz, 13.56 MHz, and 125 kHz.

RFID site survey A comprehensive analysis to determine or confirm that a proposed RFID solution meets the intended application requirements and technology specifications of use. It also defines the equipment needed to implement a proposed RFID system and outlines the responsibilities of each party involved with the system implementation.

RS232 A common physical interface standard specified by the EIA for the interconnection of devices. The standard allows for a single device to be connected (point-to-point) at baud rates up to 9,600 bps, at distances up to 15m. More recent implementations of the standard may allow higher baud rates and greater distances.

RS422 A balanced interface standard similar to RS232, but using differential voltages across twisted pair cables; more noise immune than RS232 and can be used to connect single or multiple devices to a master unit at distances up to 3,000m.

RS485 An enhanced version of RS422, which permits multiple devices (commonly 32) to be attached to a two-wire bus at distances of over a kilometer (close to a mile).

Savant Servers that act as local repositories for data and associated information, supporting sophisticated, flexible middleware for serving database/XML queries.

SAW A technology used for automatic identification in which low power microwave RF signals are converted to ultrasonic acoustic signals by a piezoelectric crystalline material in the transponder. Variations in phase shift in the reflected signal can be used to provide a unique identity.

Smart label A passive RFID data carrier structured into a flexible label-like form that allows overprinting with text, graphics, or data carrier symbols such as linear bar codes, multirow bar codes, or matrix code symbols. Alternatively, it is used to describe a passive chip or chipless RFID devices that are used to emulate and extend a printed label function.

Spurious emissions Unwanted harmonic outputs. For example, the type approval testing of the RFID reader includes measurement of the harmonics produced by the RFID reader, to ensure they are within the prescribed limits.

Standing waves The combination of forward and reverse traveling waves produce a standing wave, which is so called because the positions of maximum and minimum signal do not vary with time. The actual shape of this standing wave is a function of the load impedance.

Supply chain A grouping of at least four distinct management business processes that define the planning of a product, the sourcing of a product's components, the making of a product, and the delivery of a product.

Synchronization A mechanism that allows multiple readers to operate in close proximity by synchronization of their transmissions.

Transmitter (exciter) The electronics that drive an antenna; together with the antenna and a receiver, they are called a reader or scanner.

Transponder A type of integrated circuit designed to store data and respond to RF transmissions of a given frequency; another name for a RFID tag.

Transportation management A term to reference several RFID applications within the transportation industry. These include electronic toll and traffic management (ETTM), rail and intermodal tracking, fleet management, and vehicle parking/security access control.

Unique identifier (UID) A number that uniquely identifies the transponder and is used for addressing each transponder individually.

Unlicensed National Information Infrastructure (U-NII) A 5-GHz microwave band that does not require licensing (in United States, at least).

Wiegand format The most common data format in an access control system consisting of 26 bits of information.

Write The transfer of data to a tag. The tag's internal operation may include reading the data in order to verify the operation.

Write broadcast capability An RFID technology characteristic that allows data to be written to multiple tags while those tags are within an RF portal.

Write once read many (WORM) A transponder that can be partially or totally programmed once by the user and thereafter only read.

Write rate The rate at which data can be transferred to a tag, written into the tag's memory, and verified as correct. It is measured in bits (or bytes) per second.

Acronyms

ABR	Advanced biomechanical rehabilitation
ADC	Automated data collection
ADHD	Attention deficit hyperactivity disorder
AHP	Analytic hierarchy process
AI	Application identifier
AIDC	Automatic identification and data collection
AIM	Automatic identification manufacturers
AM	Amplitude modulation
AMA	American Medical Association
AmI	Ambient intelligence
ANP	Analytic network process
ASK	Amplitude shift keying
ASICs	Application-specific integrated circuits
ATA	American Telemedicine Association
BAN	Body area network
BCI	Brain-controlled interface; brain-computer interface
BSN	Body sensor network
CENELEC	European Committee for Electrotechnical Standardization
CEPT	European Conference of Postal & Telecommunications Administrations
DARPA	Defense Advanced Research Projects Agency
DBS	Deep brain stimulation
EAS	Electronic article surveillance
EC	European Commission
ECG/EKG	Electrocardiogram

ECoG	Electrocorticography
EEG	Electroencephalogram
EGE	European Group on Ethics and New Technologies
EMC	Electromagnetic compatibility
EMG	Electromyogram
EPC	Electronic Product Code
EAN	European Article Numbering
EIA	Electronic Industries Association
EOG	Electro-oculogram
ETSI	European Telecommunications Standards Institute
EuMHA	European MHealth Alliance
EUT	Equipment under test
FCC	Federal Communications Commission
FDA	U.S. Food and Drug Administration
GPRS	General packet radio service
HIMSS	Healthcare Information and Management System Society
IBCOM	Intrabrain communication
ICNIRP	International Commission on Non-Ionizing Radiation Protection
IMEC	Interuniversity MicroElectronics Center
IMD	Implantable medical devices; implantable microelectronic devices
IPS	Indoor positioning solution
ISO	International Standardization Organization (ISO)
ITU	International Telecommunication Union
LBT	Listen-before-talk
MCDA	Multicriterial decision analysis
MEMS	Microelectromechanical system
MICS	Medical implant communications service
MRP	Manufacturing resource planning
MST	Microsystem technology
NFC	Near-field communication
PAN	Personal area network
PCD	Proximity coupling device (reader/writer)
PES	Personal environment service

PICC	Proximity integrated circuit card
POS	Point of sale
PUPI	Pseudo unique PICC identifier
QoS	Quality of service
RFDC	Radio frequency data collection
RFID	Radio frequency identification
SAR	Specific absorption rate
SAW	Surface acoustic wave
SCS	Spinal cord stimulation
TE	Technoethics
UID	Unique identifier
UMTS	Universal Mobile Telecommunications System
U-NII	Unlicensed National Information Infrastructure
UWB	Ultrawideband
WBAN	Wireless body area network
WBI	Wireless body implants
WIMS	Wireless integrated microsystems
WLSA	Wireless-Life Science Alliance
WMTS	Wireless medical telemetry service
WORM	Write once read many
WPAN	Wireless personal area network
WSN	Wireless sensor networks
WWHI	West Wireless Health Institute

About the Author

Harvey Lehpamer completed his primary education and technical high school specialized in electronics in Zagreb, Croatia. He graduated from the School of Electrical Engineering (also called the Electrotechnical Faculty) of the University of Zagreb, Croatia, and received his master's and doctoral degrees from the same institution.

Dr. Lehpamer has 30 years of experience in the planning, design, and deployment of the wireless and wireline networks including microwave, fiber optic, and other transmission (transport) systems in Europe, North America, Africa, South America, and other parts of the world. He also has experience in project management, proposals and sales support, teaching, conducting seminars, electronic circuit design, and manufacturing and testing environment. Dr. Lehpamer is the author of the books *Transmission Systems Design Handbook for Wireless Networks* (Artech House, 2002), *Microwave Transmission Systems: Planning, Design, and Deployment* (McGraw-Hill, 2004), and *Microwave Transmission Systems: Planning, Design, and Deployment, Second Edition* (McGraw-Hill, 2010).

Dr. Lehpamer has worked for such companies as Ericsson Wireless Communications, San Diego, California, Qualcomm, San Diego, California, Clearnet, Toronto, Canada, Ontario Hydro, Canada, Lucas Aerospace, Microwave Technologies Division, Canada, and Electroproject Consulting Engineers, Zagreb, Croatia. He is a Licensed Professional Engineer of the Province of Ontario, Canada.

Currently, Dr. Lehpamer is an owner and the principal engineer of HL Telecom Consulting, a consulting company in San Diego, California (http://www.HLTelecomConsulting.com); he may be contacted at: HL_2@Hotmail.com or HarveyLehpamer@HLTelecomConsulting.com.

Index

N

O

P

Z