An Analysis of the Fundamental Constraints on Low Cost Passive Radio-Frequency Identification System Design

by

Tom Ahlkvist Scharfeld

B.S., Mechanical Engineering (1998)

Northwestern University

Submitted to the Department of Mechanical Engineering
In Partial Fulfillment of the Requirements for the Degree of
Master of Science

at the

Massachusetts Institute of Technology

August 2001
[September 2001]

Signature of Author ....
Department of Mechanical Engineering
August 21, 2001

Certified by ....
Sanjay E. Sarma
Associate Professor of Mechanical Engineering
Thesis Supervisor

Accepted by ....
Ain A. Sonin
Chairman, Department Committee on Graduate Students

An Analysis of the Fundamental Constraints on
Low Cost Passive Radio Frequency Identification System Design

by

Tom Ahlkvist Scharfeld

Submitted to the Department of Mechanical Engineering
On August 21, 2001 in partial fulfillment of the
Requirements for the Degree of Master of Science

Abstract

Passive radio frequency identification (RFID) systems provide an automatic means to inexpensively, accurately, and flexibly capture information. In combination with the Internet, which allows immediate accessibility and delivery of information, passive RFID systems will allow for increased productivities and efficiencies in every segment of the global supply chain. However, the necessary widespread adoption can only be achieved through improvements in performance – including range, speed, integrity, and compatibility – and in particular, decreases in cost. Designers of systems and standards must fully understand and optimize based on the fundamental constraints on passive RFID systems, which include electromagnetics, communications, regulations, and the limits of physical implementation. In this thesis, I present and analyze these fundamental constraints and their associated trade-offs in view of the important application and configuration dependant specifications.

Thesis Supervisor: Sanjay E. Sarma

Title: Associate Professor of Mechanical Engineering

ACKNOWLEDGEMENTS

When I arrived at MIT two years ago, I had little background in the material covered in this thesis. I wish to thank my advisor, Sanjay Sarma, most of all, for allowing me to move into an entirely new domain; one in which there were many better equipped at MIT. Sanjay has allowed me tremendous flexibility and freedom in defining a topic and preparing my thesis. Over the past two years, through the multiple topics I have explored, I have learned a tremendous amount. I owe much of this to Sanjay. Meetings were always motivating. Though Sanjay has been, and will likely continue to be incredibly busy, he has always made himself available for meetings, emails, and phone calls, at home or on the road.

I must also thank Dan Engels. Where Sanjay provided the freedom, Dan provided the push. Had it not been for Dan, I would likely have allowed myself much less time to write. Discussions with Dan were always interesting and insightful. His extreme distaste for gates and functionality has been infectious.

I would like to thank the others at the Auto-ID Center. Kevin Ashton was very helpful in providing the much needed and often ignored insight into the requirements of the end user community. Brooke Peterson, Tracy Skeete, and David Rodriguera provided the support to make the Center an exciting, productive, and friendly place to work.

I must also thank those outside of MIT and the Auto-ID Center whose insights were invaluable. The Alien Technology and Wave-ID team, including John Rolin, Roger Stewart, and Curt Carrender, provided the essential practical and industrial viewpoint. Through them, not only did I learn a great deal about the technology, but also its development. My brief encounters with Matt Reynolds and even briefer encounters with Peter Cole were more helpful than they might imagine.

Finally, I must thank my friends for making MIT a truly enjoyable place to be, my officemates for graciously accepting my constant presence for the last few months, my housemates for graciously accepting my constant absence for the last few months, and my family for supporting me in spite of not knowing what I was doing. Special thanks to Jean Ah Lee for thoroughly reading my thesis, Steve Buerger for offering to read my thesis, my Mother, in spite of her Swedish tongue, for suggesting a few corrections, my Father for painstakingly lumbering through my thesis, Jenny Scharfeld for reading the first paragraph, and Anna Scharfeld for checking to see how long it was.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

Through the existing Internet infrastructure we have instant access to information from a wide variety of sources. It is the means by which this information is recorded which presents an obstacle. In many cases, information exists only in virtual form, whether it be in one's mind or in a machine. It may stay in virtual form. In other cases, however, information describes actual physical objects, states and events. Transforming this physically embodied information to virtual information so that people or machines may access it, presents a problem. Current solutions to this problem typically require manual intervention - people must observe and record. Not only can this be inefficient, but it can also result in inaccurate information. In other solutions, machines with sophisticated and complex vision and sensing systems may observe and record. Even for collecting the most basic information, however, such solutions are often expensive and complex, or require significant conditions and constraints. One potential solution to this problem is radio frequency identification (RFID). Through the attachment of transponders or tags to animate or inanimate objects, and an infrastructure of networked reading devices, physically embodied information can be automatically recorded. A class of RFID systems, passive RFID, allows wireless powering of these tags. Passive RFID systems have the potential to provide extremely low cost and somewhat less constrained automatic capturing of physically embodied information. Passive RFID systems, though capable of offering reduced operational constraints, have several fundamental constraints on their design. It is one objective of this thesis to analyze these constraints.

For a better understanding of this problem, we can consider the global supply-chain. It has been demonstrated that a manufacturer, distributor, retailer, and any other unit in a supply chain, merely knowing what they have and making that information available, will mean tremendous savings and increases in efficiencies [1]. Through the Internet and its associated infrastructure, the problem of delivering information has been solved; the problem of capturing information, however, has not. Current solutions require manual data entry, or manual scanning of barcodes. More sophisticated solutions include automatic scanning of barcodes, or sophisticated machine vision systems. The manual solutions can be costly, time consuming, and plagued by inaccuracies. The current automatic solutions can also be costly, complex, and often require significant constraints on the objects and their environment. Radio-frequency identification (RFID) systems are a potential solution.

Radio-frequency identification systems allow non-line of sight reception of an identification code assigned to an object. The identification code is stored in a tag consisting of a microchip attached to an antenna. A transceiver, often called an interrogator or reader, communicates with the tag. RFID systems alone are worthless. Though they may be able to collect identification codes, those codes must be assigned to an object and they must be made available to an accessible database. A network infrastructure that will support efficient and robust collection and delivery of information captured by RFID systems, or any other system capable of detecting identification codes, is currently being developed [2]. This system should support the capture, storage, and delivery of vast amounts of data robustly and efficiently.

Though each component of the entire system including tags, readers, host systems, databases, and network infrastructure, is equally important, it is the focus of this thesis to analyze and discuss tags, and the portions of the reader that interface directly with tags. The Internet and its associated infrastructure are proven and well developed technologies. RFID systems, though historically older than the Internet, are still relatively young and untested in their development for robust, high volume and low cost operation. Performance is currently satisfactory and constantly improving. Cost is still relatively high (minimum around $0.50/tag, and over $100.00/reader) and with increased improvements and adoption, should drop significantly in the near future. Technology innovations and new applications are driving increases in performance and decreases in costs. Standards, if properly designed, should drive adoption and further decreases in cost. For a better view of RFID and its direction, I will briefly review its history.

1.2 Brief History of RFID

RFID is merely a more recent term given to a family of sensing technologies that has existed for at least the past fifty years. The first commonly accepted use of RFID related technology was during World War II [3]. In their and their allies' aircraft, the British military installed transponders capable of responding with an appropriate identification signal when interrogated by a signal. This technology did not allow determination of exact identification, but rather if an aircraft was their own. This transponder technology, called identify friend or foe (IFF), has undergone continued development and later generations are still used in both military and civilian aircraft.

RFID, as we generally know it today, has also undergone significant development since the early aircraft transponder systems of World War II. In the 60s and 70s, in an effort to safely and securely track military equipment and personnel, various government labs developed identification technology. In the late 70s, two companies were created out of Los Alamos Scientific Laboratories to commercialize this technology. Initial applications included identification and temperature sensing of cattle [4]. In the early 80s railroad companies began using the technology for tracking and identification of railcars.

These early uses were typically at higher UHF frequencies (900 MHz/1.9 GHz). Through the 80s, several companies in the United States and Europe began developing technologies for operation at different frequencies, with different power sources, memories, and other functions. In the mid-to-late 80's, as the large semiconductor companies became involved, there was a shift towards performance improvement, size reduction, and cost reduction.

In the late 80s, and through the 90s, as performance improved, and size and cost decreased, new applications emerged. Some of these include automatic highway tolling, access control and security, vehicle immobilizers, airline baggage handling, inventory management and asset tracking, and the closely related smart-cards [5-8]. For continued adoption in applications demanding high performance, small size, and low cost, improvements in design and manufacturing must be achieved. Much work is continuing today and companies continue to emerge with innovative, low cost, and high performance technologies. With the advent of the Internet and increased development in information technology, RFID systems have been provided with the infrastructure necessary to achieve penetration into new applications.

For wide-scale global adoption of RFID technology, not only must performance improve and cost decrease, but also systems of one vendor must be compatible with those of another, and must additionally operate under both foreign and domestic regulations. Efforts to develop standards for RFID and various applications are continuing. Standards, though able to drive increased adoption and subsequent reduced costs, run the risk of stifling competition, innovation, and thus continued improvements in the technology.

It is a further objective of this thesis to understand the fundamental constraints on RFID systems for purposes of proper standards development.

1.3 Classifications of RFID Systems

Over the several stages of RFID system development, several types of systems have emerged. These can be classified in several ways. The term RFID implies a rather broad class of identification devices. All RFID systems have readers and tags. Readers capture the information stored or gathered by the tag. Readers are loosely attached to particular environments whereas tags are attached to objects. Because tags are attached to objects, they suffer the most stringent specifications related to performance, size, and cost. The various classifications of RFID systems are based around these specifications. One broad classification is that of "chip" versus "chipless" tags. Chip tags have an integrated circuit chip, whereas chipless tags do not. Another classification, which is a subset of chip tags, is that of passive, semi-passive, or active. Passive tags have no on-tag power source and no active transmitter, semi-passive tags have an on-tag power source, but no active transmitter, and active tags have both an on-tag power source and an active transmitter. Yet another classification is that of read-only or read-write. Read-only tags have either read-only, or write once read many memory. Read-write tags allow writing and re-writing of information. In this thesis, I will focus on read-only "chip" passive tags. I will briefly describe these classifications and justify this focus.

1.3.1 Chipless versus Chip

As our focus is on extremely low cost tags that provide the minimum of functionality – simply a read-only device with a permanent unique identification number – it might seem that chipless tags would be optimal. One would avoid the silicon costs and the intricate manufacturing process required for integrated circuitry. However, for two reasons, I will focus on the use of chips.

- The tag must hold enough memory to hold an identification number from a scheme designed to uniquely identify massive numbers of objects.
- The reader must be able to read multiple tags in its field.

In order to uniquely identify all manufactured items, a numbering scheme should allow for a sufficient number of unique codes. A length of 64-bits may be suitable in the near term, while 96-bits is optimal [9]. Most chipless tags at present allow up to 24 bits or less, though some may allow 64 bits [10]. These, however, come at a higher cost.

Because an ever increasing number of objects, ever decreasing in size, will be tagged, it is necessary that a reader be able to read multiple tags within its range and in close proximity to each other. At present, the best way to successfully accomplish this is through some intelligence on the tag itself. Methods of spatially isolating a single tag among many others, though potentially feasible, are not yet available. I will discuss the various multiple tag identification schemes in Chapter 7.

Though chipless tags show tremendous promise in achieving larger memories and improved anti-collision functions, and in the future, circuits may be printed directly onto non-silicon substrates, chip tags offer the most near-term promise in satisfying the demands of the majority of object tracking and identification applications.

1.3.2 Passive, Semi-Passive, and Active

The difference between passive, semi-passive, and active tags is the power source and the transmitter. Passive tags have no on-tag power source and no on-tag transmitter. Semi-passive tags have an on-tag power source, but no on-tag transmitter. Active tags have both an on-tag power source and an active transmitter.

Active tags offer the best performance. Ranges can be on the order of kilometers and communications are fast and reliable. This comes at a high cost, however. Semi-passive tags offer increased operating ranges over passive tags (on the order of tens of meters) and because of their power source may allow increased functionality. These additions, however, also increase cost. Passive tags have ranges of less than 10 meters and are most sensitive to regulatory constraints and environmental influences. They however have the most potential for extreme low cost.

1.3.3 Read-only and Read-write

Any chip tag may be read-only or read-write. Typically though, only passive tags are read-only. Read-only tags have an identification code recorded at the time of manufacture, or when allocated to a particular object. Memory is either read-only memory (ROM), or write-once-read-many (WORM). I will discuss these memories further in Chapter 5.

Read-write tags may be written to many times throughout their life. They also typically have an identification code or serial number recorded at manufacture. Because a variety of information can be written, read-write tags offer additional functionality. This, however, comes at an increased cost.

Read-only tags offer the greatest potential for low cost, and in combination with a well-designed networked distributed database, provide essentially the same functionality as read-write tags.

1.4 Assumptions

Before discussing the specifications, I will reiterate the assumptions. I will assume that tags are read-only and passive. Range will typically be no more than 6 meters, and there is an infrastructure of readers, either at key locations or distributed throughout a facility. In addition, I will assume that there is a networked infrastructure through which identification codes unique to specific objects can be linked to information about those objects [2]. Identification codes will be logically partitioned with fields for a header, manufacturer code, product code, and serial number [9].

1.5 Specifications

As we have seen, through history, RFID applications have changed along with the technology itself. These applications each have their own specific demands. To properly analyze RFID technology, both for gaining an understanding of its functional characteristics, and for making improvements in design, we must evaluate its function on the basis of specifications. Because of the wide and varied applications, this is difficult. There are however some broad characteristics we should consider. These include cost, size, and performance. Performance concerns include reading range, speed, integrity of the communications, and compatibility between systems of different vendors and within differing worldwide regulation domains.

The more extreme and demanding near-term applications are those involving low-cost disposable and consumable items. Tagging of mail [11] and grocery items are two examples. In such applications, performance is important. Readers must be able to quickly read large numbers of tags at ranges on the order of meters in relatively unconstrained environments. Cost, however, is extremely important. Because the tagged items are consumable, and enormous numbers of items will be tagged, minute increases in the cost of a single tag will mean tremendous overall costs. Depending on the size of the object, size can also be extremely important. I will briefly discuss the performance, cost, and size specifications.

1.5.1 Performance

The main performance specifications include: range, speed, integrity, and compatibility. Exact specifications require knowledge of the exact application and configuration in which RFID systems are to be used. Though the broad application is low cost consumable item tagging, this can be subdivided into almost countless applications each with multiple configurations. For example, we can consider each unit in the supply chain including factories, distribution centers, retail stores, homes and waste management centers, and transportation between them.

In factories there are assembly, sorting, and shipping applications. Distribution centers might have receiving, sorting, picking, packing, and shipping applications. Retail stores might have receiving, sorting, and other inventory management applications. Homes may also have inventory management applications. Waste management centers might have sorting applications. Transportation providers and cross-docking providers will have additional inventory and sorting applications. Reverse logistics operations offer additional applications.

It is likely that if consumable items are tagged, the boxes and totes that hold them, the pallets that hold the boxes, and the containers that hold the pallets will also each be tagged. Pallet and container tags could likely be reusable, so cost constraints would not be so stringent. Boxes are fewer in number than the items inside, so cost demands will be more stringent than those for pallets, but not as stringent as those for items.

Given that containers, pallets, and boxes can be individually tagged and identified, their contents should be known. Identification of individual consumable items will occur at locations where they are no longer contained. Such situations occur in factories and distribution centers, but more frequently in stores, homes, and waste management centers. Applications in stores, particularly at checkout, are especially demanding. Readers and tags may have to cope with interference from other readers in the vicinity, and reflections or influences from metal and people in the environment. Large numbers of random distributions of items may need to be identified accurately and quickly. Maximum required range will likely need to be on the order of meters. Exact requirements, however, will depend on the desired configuration, of which a number exist. Readers may be installed around gates or portals, in mobile carts or containers, or stationed throughout an area. The environment may be somewhat constrained, or completely unconstrained – objects may need to be separated from other objects and passed through a reader individually, or simply carried in bulk at full walking speed through the door.

1.5.2 Cost and Size

Without doubt, from the standpoint of the end user, cost should be minimized. Given that tags are consumable and disposable, they are a recurring cost. Minute savings in the cost of a single tag will mean tremendous savings to the end user. Current targets are in the range of 5 cents or lower at significant volume. Minimization of cost at this scale presents major challenges to design. Every aspect of the system must be optimized first for low cost and second for performance – not only tags with their integrated circuit and antennas, but also all the methods used to fabricate, assemble, and apply the final product. Design must creatively exploit the fundamental constraints on the system.

Size, of course, is also extremely important. At the very minimum, tags should be smaller than the tagged object. The antenna will always be the constraining factor as tag microchips with dimensions less than 0.5 mm square have already been manufactured. Because performance, in particular range, is dependent on size and shape of antennas and antenna size is largely dependent on operating frequency, tags on very small objects will not be able to achieve the same performance of a tag with a larger antenna more suited to the operating frequency. At 5.8 GHz for example, an optimal antenna length is 2.5 cm, whereas at 915 MHz, an optimal antenna length is 16 cm. Antennas with lengths smaller than these optimal lengths may still have suitable performance, however it will degrade. Blindly increasing frequency is not an option as there are other overriding constraints such as propagation, power transfer efficiency, and regulations. As with the other applications, size is highly application and configuration dependant. In some instances only extremely close range is required. In such cases, sizes may be very small.

Regardless, the main concerns remain cost, size, and performance related issues, including, range, speed, integrity and communications. It is these specifications, in combination with the fundamental constraints, that determine the ultimate design of the RFID system.

1.6 Functions, Constraints and Organization

Given the assumptions and the important specifications, we must consider the fundamental constraints. As the organization of this thesis is based around these constraints, I will summarize the organization in parallel with the constraints. First, however, it will be helpful to reconsider the components and functions of an RFID system.

1.6.1 Components and their Functions

In general, as shown in Figure 1.1, an RFID system consists of four main components: a host, a reader, multiple tags, and a channel through which the reader and tags communicate.

I will consider only the components of the system that interact directly with the function of the tags, namely the tag interface portion of the reader, the tags themselves, and the communications channel. The functions performed by these components along with the specifications and constraints that govern their design, are shown in Figure 1.2.

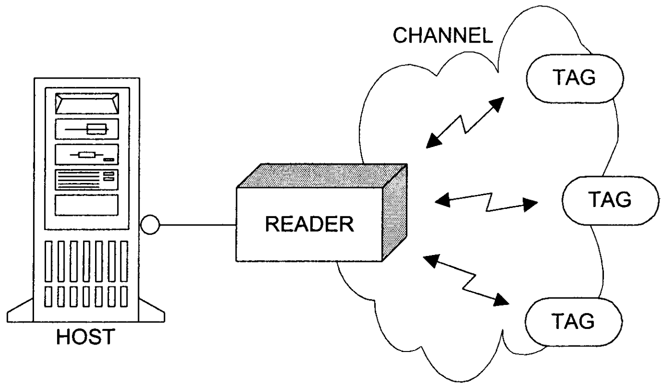


Figure 1.1: The four main components of an RFID system: host, reader, channel, and tags.

From Figure 1.2, we see that the operations of the tag and the reader are complementary. The internal operations of both the tag and the reader include the higher-level algorithms and command protocols necessary for identification of a single tag or multiple tags within the reading range of a reader. Commands are generally based around algorithms called anti-collision algorithms as they imply the ability to reduce the occurrence of multiple simultaneous, colliding responses to a query from the reader.

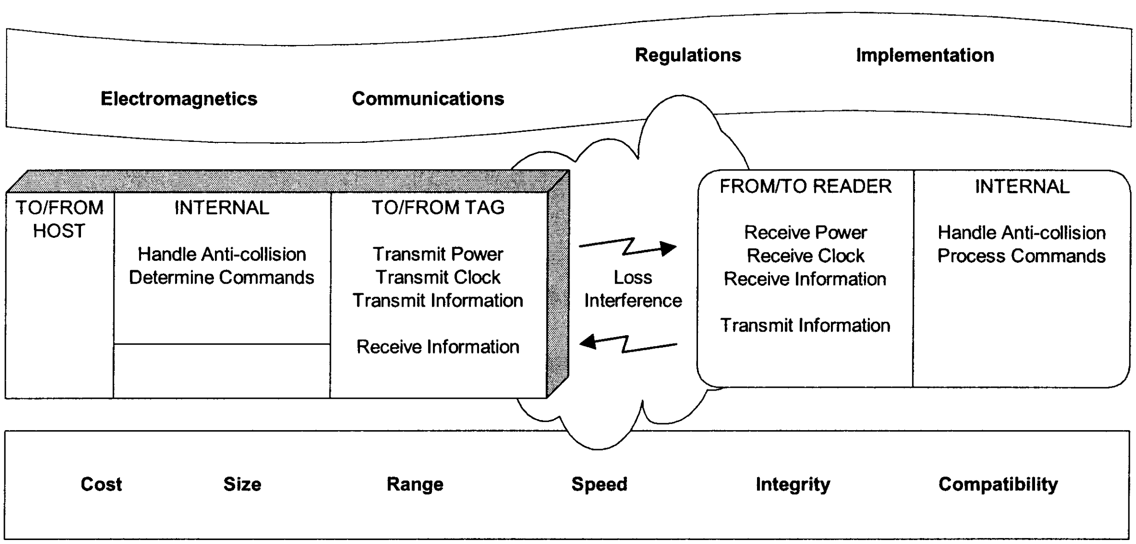


Figure 1.2: Functions of the reader, channel, and tag and their governing specifications and constraints.

The interface operations involve the lower level functions necessary to satisfy the requirements of the higher-level commands. The reader must transmit power, information, and a clock. The information transmitted, is essentially what is required for the multiple identification, or anti-collision protocols. The clock is necessary for driving the digital circuitry of the tag. Depending on the frequency of operation, the clock is either generated directly from the carrier frequency, or through modulations of the carrier.

The tag must receive and process the power, information, and clock from the reader-transmitted signal. After the higher-level internal processing of the received information, the tag may choose to transmit information, likely its identification code, or portions of it, back to the reader. In passive RFID, this is done through modulations of the reader's signal. The reader must, in turn, receive this information, and sned it back to the higher-level internal functions.

1.6.2 Constraints and Organization

The design and implementation of these various functions is determined by both the specifications and fundamental constraints. These constraints include the electromagnetics and communications, which define the methods by which the tag and reader can communicate, and the regulations and hardware constraints, which impose limits on these methods.

In Chapter 2, I will discuss the electromagnetic fields and waves that link the reader to the tag. I will describe how they are created by antennas and how their characteristics and behavior vary with operating frequency and distance from the antenna. I will describe how RFID systems exploit these characteristics to achieve power and communications coupling. Finally, I will cover the various environmental influences on the behavior of fields and waves, and the coupling between reader and tag.

In Chapter 3, I will discuss the communications between the reader and tag and the theory that describes it. I will discuss and analyze the characteristics of the various coding and modulation schemes commonly used in RFID. I will consider detection methods and concerns related to the integrity of communications and error detection.

In Chapter 4, I will consider the regulations that govern both the electromagnetics and communications. I will consider regulations imposed by bodies in three different regions of the world, Europe, the United States, and Japan. I will discuss their various methods of specifying regulations and the implications this has for RFID systems.

In Chapter 5, I will discuss constraints related to physical implementation of the hardware. I will discuss those properties of antennas that affect performance, cost, and size. I will also consider the integrated circuit and how its design and implementation affects performance and cost.

In Chapter 6, I will discuss and analyze anti-collision methods and the command protocols necessary to drive the identification of multiple tags within the range of a reader.

In Chapter 7, I will relate the fundamental constraints to the important performance specifications, and discuss their interactions.

Finally, in Chapter 8, I will conclude and present directions for future work.

Chapter 2

ELECTROMAGNETICS AND ANTENNAS

2.1 Introduction

In RFID systems, readers are linked to tags by electromagnetic fields and waves occupying or propagating through the environment. In order to understand and design these systems we must understand how these fields and waves are created, manipulated, and received. In this chapter, I will review the principles and behaviors of electromagnetic fields and waves. We seek to understand,

- How fields are created, their size, available power, their variation with range, angle, orientation, and polarization, and how they are transmitted and received.
- How antenna type, size, and shape affect the properties of the field.
- How we can maximize the reception of power and information through tuning and matching.
- How reactive and radiating link conditions vary with the channel and its environment.

These issues will aid us in understanding the physical constraints on RFID systems and the fundamentals governing their function. In addition, they will help us in understanding other fundamental constraints on RFID systems, including communications theory, and regulations. In the end we will consider how these various factors influence performance and cost.

2.2 Maxwell's Equations and Electromagnetics Fundamentals

When a single charge emitting static electric field lines is accelerated in some direction, an electric field is radiated. If that charge is continuously oscillated it will radiate a continuously oscillating electric field. A time and space varying electric field has an associated magnetic field. Maxwell's equations describe the behavior of these electromagnetic fields at every point in space and instant in time relative to the position and motion of charged particles [12]. We can represent these equations in time-harmonic (sinusoidal) form with frequency ω, through the relationship $\partial/\partial t = -j\omega t$:

$$\nabla \times \mathbf{E} = j\omega \mathbf{B} \tag{2.1}$$

$$\nabla \times \mathbf{H} = \mathbf{J} - j\omega \mathbf{D} \tag{2.2}$$

$$\nabla \bullet \mathbf{D} = \rho \tag{2.3}$$

$$\nabla \bullet \mathbf{B} = 0 \tag{2.4}$$

where:

$\mathbf{E}$ = electric field (V/m)

$\mathbf{H}$ = magnetic field (A/m)

$\mathbf{B}$ = magnetic flux density (T)
$\mathbf{D}$ = electric displacement (C/m^2)
$\mathbf{J}$ = electric current density (A/m^2)
ρ = electric charge density (C/m^3)

The continuity equation for the conservation of charge and current is given as

$$j\omega\rho + \nabla \cdot \mathbf{J} = 0 \tag{2.5}$$

Together with Maxwell's equations, they form the fundamental equations of electromagnetics.

The time averaged power per unit area delivered through a surface by electric and magnetic fields is given by Poynting's vector:

$$\mathbf{S} = \tfrac{1}{2}\mathrm{Re}\{\mathbf{E}\times\mathbf{H}^*\} \tag{2.6}$$

where $\mathbf{H}^*$ is the complex conjugate of $\mathbf{H}$, and $\mathbf{S}$ is a power density having units of W/m^2.

In a region with no charge or current densities, an electric field, coupled with and orthogonal in polarization to a magnetic field, will propagate through a given medium in a direction perpendicular to both fields. This is an electromagnetic wave. An antenna is designed to radiate electromagnetic fields and waves.

In further discussion we will use several parameters to describe fields and waves. These parameters can be given in general form for describing interactions with any material. We will focus on those for loss-less media and free-space. The free-space wavenumber is given by

$$k_0 = \omega\sqrt{\mu_0\varepsilon_0} = \frac{\omega}{c} = \frac{2\pi}{\lambda_0} = \beta \tag{2.7}$$

where:

c = speed of light $\cong 3\times10^8 (m/s)$
ε_0 = free space permittivity $= 8.8542\times10^{-12} (F/m)$
μ_0 = free space permeability $= 4\pi\times10^{-7} (H/m)$

The permittivity relates the electric field $\mathbf{E}$ to the electric displacement $\mathbf{D}$ through $\mathbf{D} = \varepsilon_0\mathbf{E}$. The permeability relates the magnetic field $\mathbf{H}$ to the magnetic flux $\mathbf{B}$ through $\mathbf{B} = \mu_0\mathbf{H}$.

Another useful parameter is the free-space impedance

$$\eta_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} \tag{2.8}$$

Through the fundamental equations of electromagnetics and these various parameters, we can understand how field and waves are created.

2.3 Antennas and Their Surrounding Regions

For a better sense of the relationship between electromagnetic fields and waves, it is helpful to examine how they are formed by an antenna. We can derive the electric and magnetic fields created and radiated by any antenna using Maxwell's equations. For purposes of illustration, we will consider two antennas considered to be electrically small in that their maximum dimension is much less than the wavelength of oscillation. We will consider the ideal dipole, also known as a Hertzian dipole, and the small loop [13]. The ideal dipole is an infinitesimal element that carries current with a uniform amplitude and phase over its length. The small loop is a closed current loop with a perimeter of less than about a quarter of the oscillation wavelength. The small loop is the magnetic dual of the electric ideal dipole.

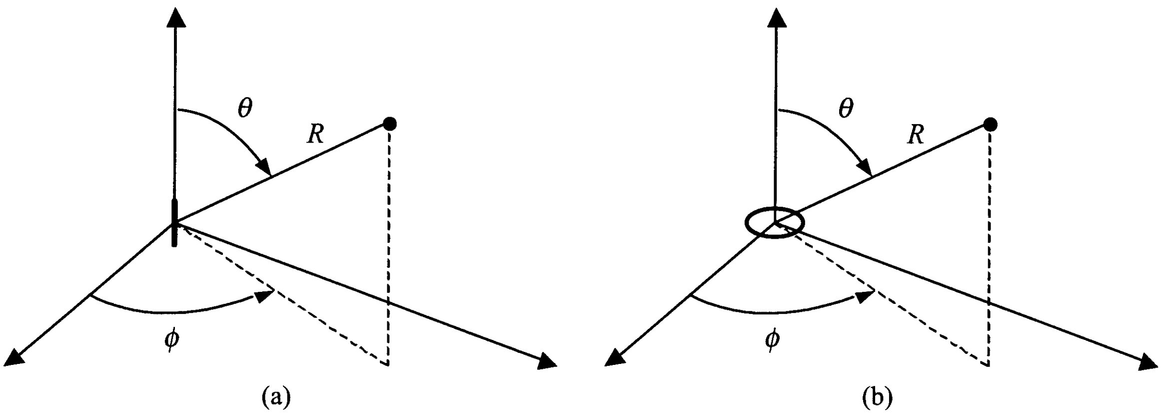


Figure 2.1: (a) Ideal dipole and (b) small loop.

A general approach to deriving the radiated electric and magnetic fields involves first calculating a vector potential based on the current density. The electric field is found from the vector potential and the magnetic field is subsequently found from the electric field. Fields created by an oscillating ideal dipole with length *dl* can be written as [14]

$$\mathbf{E} = -\frac{Id\ell}{4\pi}\eta_0\beta^2 2\cos\theta\left[\frac{1}{(j\beta r)^2}+\frac{1}{(j\beta r)^3}\right]e^{-j\beta r}\hat{\mathbf{r}} - \frac{Id\ell}{4\pi}\eta\beta^2\sin\theta\left[\frac{1}{j\beta r}+\frac{1}{(j\beta r)^2}+\frac{1}{(j\beta r)^3}\right]e^{-j\beta r}\hat{\boldsymbol{\theta}} \tag{2.9}$$

$$\mathbf{H} = -\frac{Id\ell}{4\pi}\beta^2\sin\theta\left[\frac{1}{j\beta r}+\frac{1}{(j\phi r)^2}\right]e^{-j\beta r}\hat{\boldsymbol{\varphi}} \tag{2.10}$$

Examining the electric and magnetic field equations, one should note the dependence on distance r from the antenna. When $\beta r << 1$ (or, $r << \lambda/2\pi$), the third order terms will dominate. At this distance, the electric field strength decays as $1/r^3$, and the magnetic field strength decays as $1/r^2$. We will refer to this region as the near-field.

When the distance r is much greater than $\lambda/2\pi$, the first order terms dominate, and both the electric field and magnetic field strengths decay as $1/r$. We will refer to this region as the far-field.

Examining these equations further, we notice that in the near-field where $\beta r \ll 1$, not only does the third order term dominate, but $e^{-j\beta r}$ also approaches 1. The electric and magnetic fields reduce to

$$\mathbf{E}^{\mathrm{nf}} = j\frac{Id\ell}{4\pi\beta r^3}\eta_0\left(2\cos\theta\hat{\mathbf{r}}+\sin\theta\hat{\boldsymbol{\theta}}\right) \tag{2.11}$$

$$\mathbf{H}^{\mathrm{nf}} = \frac{Id\ell}{4\pi r^2}\sin\theta\hat{\boldsymbol{\varphi}} \tag{2.12}$$

Beyond the dissimilar relationship with distance r indicating dissimilar decay, we note that the electric field is imaginary, indicating it is $\lambda/4$ (90 degrees) out of phase with the magnetic field. This is an indication of a reactive field – energy is essentially stored and released between the two fields. Evaluating the Poynting vector for this case further reveals that there is no real power flow. The electromagnetic fields in the near-field are essentially decoupled from each other and quasi-static. Further simplification will reveal that the small dipole is essentially behaving as a static electric dipole.

In considering the far-field where $\beta r \gg 1$, the first order terms dominate and the field equations reduce to

$$\mathbf{E}^{\mathrm{ff}} = j\frac{Id\ell}{4\pi r}\eta_0\beta e^{-j\beta r}\sin\theta\hat{\boldsymbol{\theta}} \tag{2.13}$$

$$\mathbf{H}^{\mathrm{ff}} = j\frac{Id\ell}{4\pi}\beta e^{-j\beta r}\sin\theta\hat{\boldsymbol{\varphi}} \tag{2.14}$$

We immediately see that the electric and magnetic fields are in phase, orthogonal in polarization, and related in magnitude by $E_\theta/H_\varphi=\eta_0$, the intrinsic impedance of free space. Both fields decay as $1/r$. Evaluation of the Poynting vector reveals a real power density, indicating propagation through a surface. The electromagnetic fields in the far-field constitute an electromagnetic wave.

We can also consider the fields for a small loop antenna,

$$\mathbf{E} = \frac{Id\ell}{4\pi}\eta_0\beta^2\sin\theta\left[\frac{1}{j\beta r}+\frac{1}{(j\beta r)^2}\right]e^{-j\beta r}\hat{\boldsymbol{\varphi}} \tag{2.15}$$

$$\begin{aligned}\mathbf{H} = &-\frac{Id\ell}{4\pi}\beta^2 2\cos\theta\left[\frac{1}{(j\beta r)^2}+\frac{1}{(j\beta r)^3}\right]e^{-j\beta r}\hat{\mathbf{r}}\\ &-\frac{Id\ell}{4\pi}\beta^2\sin\theta\left[\frac{1}{j\beta r}+\frac{1}{(j\beta r)^2}+\frac{1}{(j\beta r)^3}\right]e^{-j\beta r}\hat{\boldsymbol{\theta}}\end{aligned} \tag{2.16}$$

The near-field components are

$$\mathbf{E}^{\mathrm{nf}} = -\frac{Id\ell}{4\pi r^2}\eta_0 \sin\theta\hat{\boldsymbol{\varphi}} \tag{2.17}$$

$$\mathbf{H}^{\mathrm{nf}} = j\frac{Id\ell}{4\pi\beta r^3}\left(2\cos\theta\hat{\mathbf{r}} + \sin\theta\hat{\boldsymbol{\theta}}\right) \tag{2.18}$$

and the far-field components are

$$\mathbf{H}^{\mathrm{ff}} = j\frac{Id\ell}{4\pi r}\beta e^{-j\beta r}\sin\theta\hat{\boldsymbol{\theta}} \tag{2.19}$$

$$\mathbf{E}^{\mathrm{ff}} = -j\frac{Id\ell}{4\pi r}\eta_0\beta e^{-j\beta r}\sin\theta\hat{\boldsymbol{\varphi}} \tag{2.20}$$

Noting the similarities between the ideal dipole and the small loop field equations, we see that the small loop is the dual of the ideal dipole.

There is yet another region is of concern, often called the Rayleigh region. In this region, it becomes reasonable to approximate a spherical electromagnetic wave as a uniform plane wave. Plane waves are those with an electric field with the same direction, magnitude, and phase in infinite planes, perpendicular to the direction of propagation. They do not exist in practice, as they would require infinitely sized sources. However, at a distance far enough from the source, this approximation becomes reasonable. This region is given by

$$r > \frac{2D^2}{\lambda} \tag{2.21}$$

where r is the distance from the antenna, and D is the length of a radiating line source. When we discuss electromagnetic waves, we will not only assume we are in the far-field, but also in the Rayleigh region.

In addition to understanding the relationship between electromagnetic fields and waves, we also see that electromagnetic fields exhibit radically different behavior in the near-field zone as compared with the far-field zone. In the near-field, fields are reactive and quasi-static, while in the far-field they constitute radiated waves. This result is particularly important to RFID systems. Those systems operating at lower frequencies where the near-field encompasses the operating range, must achieve coupling through the quasi-static fields. RFID systems operating at higher frequencies typically operate in the far field and achieve coupling through electromagnetic waves.

We should note that the transition point between near-field and far-field is actually dependent on the antenna geometry. However, we will use the transition point of $\lambda/2\pi$ as the standard definition.

2.4 Impedance of Space, Antennas, and Circuits

Before further analyzing the behavior of RFID systems operating in the near-field and far-field, it will be helpful to discuss the concept of impedance as it relates to free-space, antennas, and circuits. In general, impedance describes the relationship between an effort and a flow. In electromagnetic field and wave theory, impedance is the relationship between the electric field and the magnetic field.

$$Z = \frac{E}{H} \tag{2.22}$$

In an antenna or electrical circuit, it is the relationship between the voltage and the current.

$$Z = \frac{V}{I} \tag{2.23}$$

Regardless of domain, it is an extremely useful parameter capable of characterizing the behavior of fields and waves, radiation and reaction of an antenna, and power transfer between an antenna, transmission line, and load.

Figure 2.2a shows the magnitude of the free space impedance of the fields generated by both an ideal dipole, and a small loop. We see that at the near-field to far-field transition point ($\lambda/2\pi$), the impedances converge and become constant. We note that in the far field, the impedances of the two different antennas are identical and equal to the intrinsic impedance of free space, η. In the near-field however, they are different. Figure 2.1b shows the phase of the impedances of the two antennas. We note that in the near-field, the phases are opposite at positive and negative $\pi/2$, while in the far-field they converge to 0.

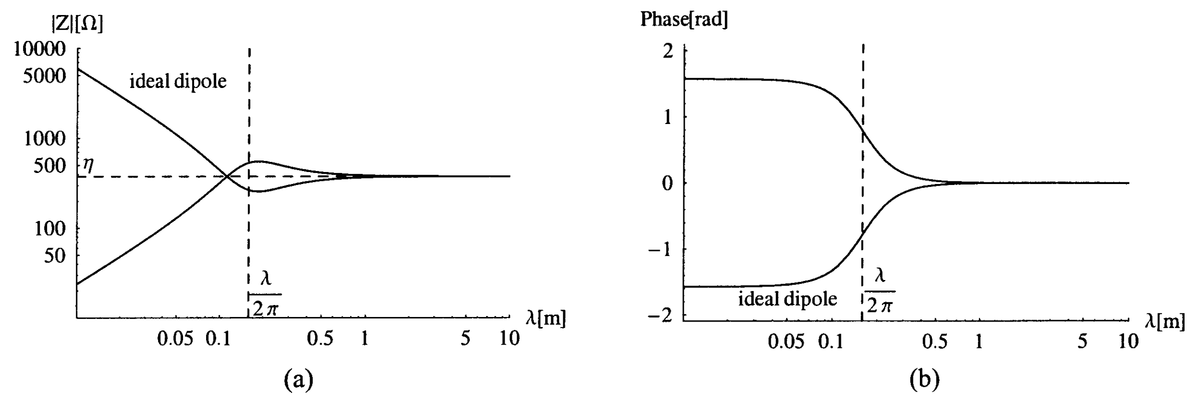


Figure 2.2: (a) Plot of impedance magnitude versus wavelength for an ideal dipole and small loop. (b) Plot of impedance phase versus wavelength for an ideal dipole and small loop. The ideal dipole impedance is labeled; the small loop impedance is unlabeled.

The antennas themselves have an input impedance at their terminals. The real portion represents a combination of actual radiation R_{rad} and ohmic losses R_{ohmic}, while the reactive portion X indicates energy stored in the field around the antenna.

$$Z_{in} = R_{rad} + R_{ohmic} + jX \tag{2.24}$$

The smaller the antenna relative to a wavelength, the lower the radiation resistive component, and the higher the reactive component. This characterizes an inefficient radiator. If we refer back to Figure 2.2 showing the impedance magnitude and phase of the fields produced by the ideal dipole and the small loop, we see that the small loop has a large positive reactive component, while the ideal dipole has a large negative reactive component. A positive reactive component represents inductance while a negative reactive component represents capacitance. This is particularly important to RFID systems that operate in the near-field.

As the antenna size increases relative to the wavelength, the radiation resistance component increases, while the reactive component decreases. At a length or perimeter of a half-wavelength the reactive component approaches zero while the resistive component reaches its maximum. At this dimension, the antenna is resonant and radiates efficiently. Far-field systems typically use resonant antennas. Increasing the size further, however, results in an increase in the reactive component and a decrease in the resistive component, until at a dimension of a wavelength, the impedance is similar to what it was at an infinitesimal dimension. This cycle repeats for every multiple of a wavelength.

Transmission lines and electric circuits, too, have an impedance. As described by transmission line theory, for maximum power transfer, impedances must be conjugate matches. Real components should be equal, while reactive components should be equal and opposite. We will see in the following sections how this affects the operation of RFID systems.

2.5 Coupling in the Near Field and Far Field

We will now consider the antennas and fundamental principles necessary for communications and energy coupling for operation in either the near-field or the far-field. Due to the difference in electromagnetic field behavior in each region, methods are significantly different and I will treat them separately. In the near-field, coupling is between a source and a sink as opposed to the far-field, where it is between a transmitter and receiver.

2.5.1 Near-field Coupling

As I have previously shown, electromagnetic fields in the near-field are reactive and quasi-static in nature. The electric fields are decoupled from the magnetic fields, and depending on the type of antenna employed, one will dominate the other. In the case of the ideal dipole, the electric field dominates, while in the case of the small loop, the magnetic field dominates. Coupling may be achieved capacitively through interaction with the electric field, or inductively through interaction with the magnetic field. Among near-field RFID systems, inductively coupled systems are more widely available than capacitively coupled systems. Thus I will emphasize inductively coupled systems.

In this section, I will describe the principles of and present relations governing both inductive and capacitive coupling. I will describe how energy is coupled, and how communications can be attained.

2.5.1.1 Inductive Coupling

Antennas for systems employing inductive coupling rely on interaction with the quasi-static magnetic field. Essentially these systems are transformers where current running through a primary coil induces a magnetic field, which then induces a current and voltage in a secondary coil. In the case of RFID, the primary coil is attached to the reader while the secondary coil is attached to the tag. A small loop antenna is preferred to a dipole antenna because the near-field magnetic field created by the small loop dominates that created by the dipole.

Here it is important to clarify the difference between a small loop and a large loop. As mentioned previously, the perimeter of a small loop antenna should be less than approximately a quarter wavelength. Large loop near-field and far-field radiation characteristics are significantly different from those of small loops. In a large loop the current distribution varies significantly over the perimeter of the loop, while in the small loop it is approximately uniform. In addition, the small loop field equations assume that the distance from the loop is much larger than the size of the antenna. This may not always be the case.

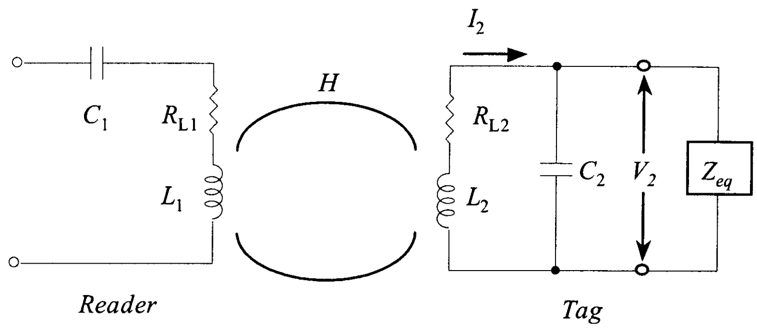


Figure 2.3: Equivalent circuits for reader and tag.

In order to determine the power induced in the secondary loop, we must first examine the near-field magnetic field equations for the loop antenna. Because the magnetic field equations for the small loop assume that the distance from the loop is much larger than the radius of the loop, and direct derivation of the equations for a large loop is more complex, we can use the Biot-Savart Law to derive a reasonable approximation for the magnetic field. The Biot-Savart Law relates the magnetic field H directly to the current distribution without requiring calculation of the vector potential. Assuming a uniform current throughout the loop and N turns, the Biot-Savart Law takes the form

$$\mathbf{H} = \frac{NI}{4\pi} \oint_C \frac{\mathbf{dl} \times \mathbf{R}}{R^3} \qquad (2.25)$$

Using this form of the equation, when distance from the antenna is on the order of the its radius, we can calculate the magnetic field at a point on the axis normal to the plane of the loop antenna. Based on the near-field small loop magnetic field equation, we know that this will give us the highest strength for a given distance, and hence provide the maximum range:

$$\mathbf{H} = \frac{N_1 I b^2}{2\left(z^2 + b_1^{\,2}\right)^{3/2}} \hat{\mathbf{z}} \qquad (2.26)$$

where b is the loop radius. Given the magnetic field, we can determine its influence on a secondary antenna. For maximum power coupling, the secondary antenna should capture as much of the magnetic field as possible. The mutual flux between the two antennas when the magnetic field is constant through some surface with area A, is given by

$$\mathbf{\Phi}_{12} = \int_{S_2} \mathbf{B}_1 \bullet d\mathbf{s}_2 = \mu H_1 A_2 \cos\psi \qquad (2.27)$$

where ψ is the angle between the field lines at the surface normal. A mutual inductance L_{12} can be expressed in terms of the mutual flux:

$$L_{12} = \frac{N_2 \Phi_{12}}{I_1} \tag{2.28}$$

As given by Lenz's Law, the negative time rate of change of this mutual flux induces a voltage in the secondary loop:

$$V_{1\to 2} = -N_2 \frac{d\Phi_{12}}{dt} = j\omega N_2 \Phi_{12} \tag{2.29}$$

Because the secondary loop and the attached circuit has some equivalent impedance, the voltage $V_{1\to 2}$ generates a finite current I_2. This current produces an additional magnetic flux which opposes the initial flux due to the secondary loop's self inductance L_2. This flux causes a voltage drop $V_{2\to 2}$, where

$$V_{2\to 2} = -L_2 \frac{dI_2}{dt} = j\omega L_2 I_2 \tag{2.30}$$

The self inductance of a small circular loop is given by [13]

$$L_2 = \mu b N^2 \left[\ln\left(\frac{8b}{a}\right) - 1.75 \right] \tag{2.31}$$

where b is the loop radius, a is the wire radius, and $a << b$.

The secondary loop also has an associated resistance that produces an additional voltage drop.

$$V_{R_{L_2}} = I_2 R_{L_2} \tag{2.32}$$

where ohmic resistance for a small circular loop antenna is given by [13]

$$R_{L_2} = N^2 \frac{b}{a} \sqrt{\frac{\omega\mu}{2\sigma}} \tag{2.33}$$

Combining the mutually induced voltage, the self induced voltage, and the resistive component of the voltage gives the total voltage across the secondary loop V_2.

$$V_2 = V_{1\to 2} - V_{2\to 2} - V_{R_{L_2}} \tag{2.34}$$

Substituting (2.27), (2.29), (2.30), and (2.32) into (2.34), we find

$$V_2 = j\omega N_2 \mu_0 H_1 A_2 \cos\psi - I_2 \left(j\omega L_2 + R_{L_2} \right) \tag{2.35}$$

The same relations hold for induction from the secondary to the primary, in RFID, the tag to the reader.

$$V_1 = j\omega N_1 \mu_0 H_2 A_1 \cos\psi - I_1\left(j\omega L_1 + R_{L_1}\right) \tag{2.36}$$

2.5.1.1.1 Resonance and Q

To maximize the power available to the secondary circuit at a given frequency, it is necessary to create LC resonant circuits in both the tag and the reader. The resonant frequency of an LC circuit is given by

$$\omega_{LC} = \frac{1}{\sqrt{LC}} \tag{2.37}$$

In the case of the reader we wish to maximize the output current for maximum field strength, so we add a capacitance in series with a loop to create an impedance that approaches zero at the resonant frequency. In the case of the tag, we wish to maximize voltage to drive the tag circuitry, so we add a capacitance in parallel with the loop to create an impedance that approaches infinity at the resonant frequency. Because there is resistance inherent in the loops, connections, and other components, there is some quality factor Q associated with both the tag and the reader. Q is defined as the ratio of energy stored to energy dissipated or as the ratio of the center frequency of operation to the 3dB bandwidth. It can be expressed as:

$$Q = \frac{\omega L}{R} \quad \text{or} \quad Q = \frac{f_o}{\Delta f_{3\text{dB}}} \tag{2.38}$$

To maximize Q, we may either reduce the resistance, increase the reactance associated with the loops, or both. For a single frequency, maximum Q is desirable for maximum power coupling. But, because a high Q implies a narrow bandwidth, it must be low enough to allow for sufficient communications. Bandwidth as related to communications will be discussed in Chapter 3. This tradeoff will be revisited in Chapter 5.

2.5.1.1.2 Load Modulation

These relations between magnetic field strength and voltage induction, apply to both reader to tag power transfer and communications, and tag to reader communications. Tag to reader communications is achieved through variation of the tag's impedance. By varying its impedance, the tag varies the current through its loop, which alters the magnetic field. This change is subsequently detected by the reader.

The tag varies its impedance by either switching resistances or capacitances; hence there is both ohmic load modulation and capacitive load modulation [6]. In ohmic modulation, a resistance in parallel with the load is switched on or off, either in time with the data stream, or at some higher frequency (subcarrier). The parallel resistance reduces the effective resistance, which causes a higher current in the tag. This results in a voltage drop at the reader. The signal is essentially amplitude modulated.

In capacitive load modulation, a capacitance in parallel with the load is switched on or off just as with the resistor in ohmic modulation. These changes are detected by the reader as a combination of amplitude and phase modulation.

2.5.1.1.3 Voltage Available to Load

With the resonance capacitor in parallel with the equivalent impedance of the attached circuitry, we can simplify the expression for voltage in the tag. Referring to Figure 2.3, first we find the current I_2 through the tags loop as a function of the parallel equivalent impedance of the resonance capacitor and the additional circuitry equivalent impedance.

$$I_2 = \frac{V_2}{1 \Big/ \left(1/Z_{C_2} + 1/Z_{eq} \right)} = V_2 \left(\frac{1}{Z_{eq}} + j\omega C_2 \right) \tag{2.39}$$

Substituting this into Eq. (2.35) we find

$$V_2 = \frac{j\omega N_2 \mu_0 H_1 A_2 \cos\psi}{1 + \left(j\omega L_2 + R_{L_2} \right) \left(\frac{1}{Z_{eq}} + j\omega C_2 \right)} \tag{2.40}$$

This is the voltage available to a load in parallel with the loop antenna. Current through the load is simply found by Ohm's Law based on the impedance of the load. Power consumed by the load can then be determined. It is clear from Eq. (2.40) that the voltage increases with number of turns, the magnetic field strength, loop area, and orientation. The effect of the circuit parameters is not as clear. Together they define the resonant frequency and bandwidth. It should be apparent that the equivalent load impedance also influences the resonance.

2.5.1.2 Capacitive Coupling

In capacitively coupled systems, antennas create and interact with quasi-static electric fields. In these systems it is the distribution of charges rather than currents that determines the field strength and hence influences the coupling strength. Because coupling strength is dependent on amount of collected charge, rather than current, conductivity can be less important than in inductive systems [15]. However, capacitively coupled systems can suffer due to environmental affects, as I will discuss later in the chapter.

A wire or a plate dipole is a suitable antenna for capacitively coupled systems since the electric field dominates the magnetic field. If considering a dipole antenna for both primary and secondary antennas, a rough approximation for the electric field can be made by using that of the ideal dipole. However, this would assume that the distance between the primary and secondary antennas would be much greater than the length of the antenna. This may not be the case. Thus, we can derive the electric field for a wire dipole antenna of length L, with uniform charge distribution ρ, from the following equation:

$$\mathbf{E} = \frac{\rho}{4\pi\varepsilon} \int_L \frac{\mathbf{R}}{R^3} d\mathbf{L} \tag{2.41}$$

The charge collected on a surface S can be found from a derivation of Ampere's Circuital Law:

$$Q = \varepsilon \int_S \mathbf{E} \cdot d\mathbf{S} \tag{2.42}$$

A time varying electric field will induce a current through the secondary antenna. The secondary antenna has its own impedance consisting of capacitance across its electrodes, and resistance associated with the antenna and other materials. Given some equivalent impedance within the antenna and the attached circuit, the induced current will generate a voltage potential across each element.

2.5.1.2.1 Resonance and Load Modulation

Just as inductively coupled systems require resonant circuits for maximum coupling, so do capacitively coupled systems. Since the antenna has its own capacitance, inductance is added in parallel on the tag and in series in the reader. In addition, like in the inductively coupled systems, the tag can communicate with the reader through varying its impedance.

Whether a system is inductively coupled or capacitively coupled, we wish to maximize coupling for purposes of power transfer and communication. In the case of coupling through transmission and reception of electromagnetic waves, the goals are the same, though the mechanisms are quite different. I will discuss these mechanisms in the next section.

2.5.2 Far-field Coupling

As is common in most wireless communications systems, coupling in systems operating in the far-field is achieved through transmission, propagation, and reception of electromagnetic waves. After presenting some useful relations for electromagnetic fields, I will discuss the performance parameters necessary for describing the radiating properties of an antenna. I will next discuss the transmission and reception of electromagnetic waves, focusing on the power available to a receiving antenna and its attached load.

2.5.2.1 Antenna Parameters

We have already considered electrically small dipole and loop antennas. These antennas are noted for their high reactance (whether it be capacitive or inductive), inefficient radiation characteristics, and difficulty in matching. While suitable for operation in the near-field, they are generally not suitable for far-field operation where transmission and, particularly, reception should be efficient. For this reason, resonant antennas characterized by a dimension on the order of one half the wavelength of transmitted frequency, are commonly used. Resonant antennas offer more efficient radiation and reduced reactivity. Bandwidth, however, can be narrow. Types commonly used in RFID systems include half-wave dipole and microstrip patch antennas. Certain physical characteristics of these antennas will be discussed in Chapter 5.

The waves radiated by all antennas consist of electric and magnetic fields related by the free-space impedance. The field magnitudes vary with antenna type and output power, but both decay with the inverse of distance from the source. The angular distribution of radiation, however, varies with the type of antenna.

We use the radiation pattern to describe the angular distribution of an antenna's radiation. An antenna radiation pattern is given in terms of the normalized electric field distribution over a constant distance r. In the case where a z-directed source has an electric field with only a theta component, the radiation pattern can be given as,

$$F(\theta,\phi)=\frac{E_\theta}{E_\theta(\max)} \tag{2.43}$$

In the case where there is phase variation, the phase is generally set to 0 at the point where the electric field is at its maximum. In this way relative phase can be described.

Certain antennas may be able to concentrate their fields into a narrower beam of radiation. In doing so the power density will increase relative to distance, allowing transmission at longer ranges. The term *directivity* is used to describe how an antenna concentrates its energy in one direction as compared to every other direction. It is defined as the ratio of radiation intensity in a certain direction to the average radiation intensity.

Directivity is solely based on an antenna's radiation pattern. It is often useful, however, to describe not only the directive properties of an antenna but its efficiency in transforming some input power to radiated output power. The term *gain* quantifies this. Typically it is defined as 4π times the ratio of radiation intensity in a given direction to the net power input of the antenna.

$$G(\theta,\phi)=\frac{4\pi U(\theta,\phi)}{P_{in}} \tag{2.44}$$

When referring to gain simply as G, we will be referring to maximum gain.

Gain is often described by comparing the maximum radiation intensity of one antenna to the maximum radiation intensity of some standard reference antenna. Typically we will describe an antenna's gain relative to an isotropic radiator that radiates energy in all directions uniformly. Also, in certain instances gain will be described relative to a half-wave dipole. Gain is typically given in units of decibels (dB). An isotropic radiator has a gain of 0 dB, while a half-wave dipole antenna has a gain of 2.15 dB. When describing the gain relative to an isotropic radiator we describe this by units of dBi. When describing gain relative to a half-wave dipole antenna, we use units of dBd. If we consider an antenna with a gain of 6 dB, it can be described by a gain of 6 dBi, or 3.85 dBd [13].

$$G\ [\text{dB}]=G\ [\text{dBi}]=G_d\ [\text{dBd}]+2.15\ \text{dB} \tag{2.45}$$

Unless otherwise specified, when referring to gain, we will be referring to gain relative to an isotropic radiator.

Just as gain can be described relative to some standard reference antenna, so to can be the radiated power. We often use *effective (or equivalent) isotropically radiated power*, EIRP, defined as the net input power to an antenna multiplied by its gain relative to an isotropic antenna.

$$EIRP=G_tP_t \tag{2.46}$$

We may also use *effective radiated power*, ERP, which is simply the net input power to an antenna multipled by its gain relative to a half-wave dipole antenna:

$$ERP=P_tG_{td} \tag{2.47}$$

EIRP is related to ERP by

$$EIRP = ERP \times 1.64 \tag{2.48}$$

With these parameters we can now consider the transmission and reception of waves.

2.5.2.2 Transmission and Reception

In considering the operation of RFID tags, we must determine the power available at the tag antenna. The tag will absorb some of this power for powering itself and detecting information. It will also scatter some of this power for transmitting information back to the reader. To understand this process we will first determine the power available to the tag given the transmitted power and gain at the reader. We will then determine the power delivered to the tag's load. Finally, we will consider the scattering of power back to the reader for communications.

2.5.2.3 Reader to Tag Transmission

Given an electromagnetic wave incident on a receiving antenna, the electric field will induce a voltage V_A across the antenna,

$$V_A = \mathbf{E}^{\mathbf{i}} \cdot \mathbf{h}^* \tag{2.49}$$

where $\mathbf{E}^{\mathbf{i}}$ is the incident electric field, $\mathbf{h}$ is the antenna *vector effective length*, and $\mathbf{h}^*$ is its complex conjugate. The tag will generally store charge and use it when necessary. However, there must be sufficient power available at the antenna to keep the tag charge storage full. If sufficient power is not available at a given distance, range will be reduced.

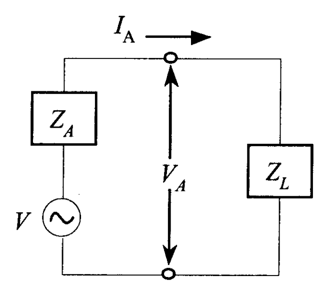


Figure 2.4: Equivalent circuit of a far-field tag where we model the antenna as a voltage source and impedance.

If we consider a transmitter that sends some power P_t through an antenna with gain G_t, we can compute the power density S incident on a receiving antenna at some distance R from the transmitting antenna:

$$S = \frac{G_t P_t}{4\pi R^2} \tag{2.50}$$

This power density is derived from Poynting's vector given by Eq. (2.6), and computed from the electric and magnetic fields.

From this power density, it is useful to define an effective aperture A_e, which is based on the antenna's own gain. It essentially can be thought of as a power capture area:

$$A_e(\theta,\phi)=\frac{\lambda^2}{4\pi}G(\theta,\phi) \tag{2.51}$$

When simply denoted as A_e without the dependence on angle, it represents the maximum effective area.

Simply multiplying the effective aperture by the power density should give the power received by the receiving antenna. However, we must also consider potential polarization mismatch issues. If the antenna has some polarization that is different from that of the incoming wave, only a fraction of the power will be received. For this reason, we define a polarization mismatch factor, p, to account for the potential polarization mismatch [13]. It is given as,

$$p=\frac{\left|\mathbf{E^i}\cdot\mathbf{h}^*\right|^2}{\left|\mathbf{E^i}\right|^2\left|\mathbf{h}\right|^2} \tag{2.52}$$

where $\mathbf{E^i}$ and $\mathbf{h}$ are defined as in (2.49). When there is no polarization mismatch, p is 1.

We can now compute the power received by the receiving antenna:

$$P_r=pSA_{er}(\theta,\phi)=p\frac{G_tP_tA_{er}(\theta,\phi)}{4\pi R^2} \tag{2.53}$$

Substituting for A_{er} in terms of G_r using Eq. (2.51) and Eq. (2.53) becomes:

$$P_r=pP_t\frac{G_tG_r\lambda^2}{(4\pi R)^2} \tag{2.54}$$

This gives the power received by the antenna, yet as there is some load in the attached circuitry, we must consider the power available to the load. For maximum power transfer from the antenna to the load, we must match the impedance of the antenna with the load equivalent impedance. In cases where a conjugate match between the load and antenna impedances is achieved, the power available to the load is that received by the antenna. In other circumstances we include an impedance mismatch factor q [13]:

$$q=\frac{P_D}{P_{D_{\max}}}=\frac{4R_AR_L}{(R_A+R_L)^2+(X_A+X_L)^2} \tag{2.55}$$

When a conjugate match is achieved between the antenna impedance Z_A and the load impedance Z_L, q is 1. With q we can find the power delivered to the load resistance R_L:

$$P_D=qP_r=qpP_t\frac{G_tG_r\lambda^2}{(4\pi R)^2} \tag{2.56}$$

The fraction of P_t not delivered to the load (1-q) is scattered. In the following section, I will discuss scattering.

2.5.2.4 Backscatter Modulation

Much as near-field RFID tags achieve communications with a reader by varying their load impedance, so do far-field tags. In the far-field, however, variation of the tag's load impedance causes an intended mismatch in impedance between the tag's antenna and load. This causes some power to be reflected back through the antenna and scattered, much like the antenna is radiating its own signal. The return scattered signal is detected and decoded by the reader. This form of communication is called backscatter modulation. After briefly reviewing the general concept of scattering, I will focus on scattering by antennas. I will then relate the power available at the antenna and delivered to a load to that scattered back to the source.

2.5.2.4.1 Scattering

When an electromagnetic wave impinges on irregularities in a medium, the wave may be randomly dispersed. This phenomenon is called scattering. In sensor systems such as radar and RFID, a transmitter will transmit a radio-frequency electromagnetic wave and a receiver will detect an object's scattered response. When the receiver is collocated with the transmitter, scattering is referred to as monostatic, or backscatter. Modulated backscatter RFID systems achieve communications through controlled changes of a tag's backscatter response

A useful representation of an object's monostatic scattering characteristics is its backscattering cross section or radar cross section (RCS). After briefly reviewing the definition of scattering and radar cross-section, I will describe the basic principles of modulated backscatter in the context of RFID systems. I will then discuss, in more detail, the factors influencing RFID tag and reader design.

2.5.2.4.2 Scattering and Radar Cross Section

The IEEE defines scattering in radio wave propagation as "a process in which the energy of a traveling wave is dispersed in direction due to interaction with inhomogeneities of the medium" [16]. It results when electromagnetic waves impinge upon an object and induce oscillating charges and currents within the object and on its surface and hence an electromagnetic field. We can determine these scattered fields through numerical or analytical evaluation of the induced surface charges and currents, or through the use of the tangential field approximation as used in physical optics [17]. In general, the spatial distribution of the scattered energy depends on size, shape, and composition of the object, and the waveform, and direction of its arrival.

Radar cross section (RCS), has been defined as a measure of power scattered in a given direction. It is expressed as an area much like an antenna's effective aperture. The IEEE defines it as,

> …4π times ratio of the power per unit solid angle scattered in a specified direction to the power per unit area in a plane wave incident on the scatterer from a specified direction. More precisely, it is the limit of that ratio as the distance from the scatterer to the point where the scattered power is measured approaches infinity… [16]

Three cases are given: monostatic or backscattering RCS, where incident and pertinent scattering directions are coincident but opposite in sense, forward scattering RCS where the two directions and senses are the same, and bistatic RCS where the two directions are different. If not specified, RCS is assumed to be monostatic. It is further mentioned that RCS is a function of frequency, transmitting and receiving polarizations, and target aspect angle. It can be represented symbolically as,

$$\sigma = \lim_{R\to\infty} 4\pi R^2 \frac{\left|\mathbf{E}^{\mathbf{scat}}\right|^2}{\left|\mathbf{E}^{\mathbf{inc}}\right|^2} \tag{2.57}$$

where $\mathbf{E}^{\mathbf{scat}}$ is the scattered electric field, $\mathbf{E}^{\mathbf{inc}}$ is the incident electric field, and R is the distance from the target. In another form, RCS is given as

$$\sigma = 4\pi R^2 \frac{P_s}{P_i} \tag{2.58}$$

where P_s is the power scattered and P_i is the incident power.

Scattering can be characterized by three regimes based on the ratio of the wavelength λ to body size L. These regimes are the Rayleigh region (not to be confused with the far-field plane wave region given by Eq. (2.21)), the resonant region, and the optics region. In the Rayleigh region, the wavelength is much greater than the body size so there is little variation in phase over the length of the body. As the body essentially sees a quasistatic field, a dipole moment is induced resulting in a scattered field. In the resonant region, the wavelength is on the order of the body size – typically the body size is taken to be between 1 and 10 wavelengths. In this region, the electromagnetic energy shows a tendency to stay attached to the body's surface creating surface waves including traveling waves, creeping waves, and edge traveling waves. In the optics region, the wavelength is much less than the body size. Here the dominating scattering mechanisms are specular scattering, end-region scattering, diffraction, and multiple reflections [17].

As we are concerned with backscatter from antennas with size on the order of a wavelength, we are principally concerned with scattering in the resonant region where surface wave scattering dominates. We are unable to use the tangential plane approximation, as it is essentially only useful for evaluation of specular scattering. Instead we must solve for the scattered field based on the induced surface currents and charges. As an analytical solution can be difficult, the various methods, including the method of moments can provide a numerical solution [17].

2.5.2.4.3 Antenna Scattering

In analyzing modulated backscatter tags, we are most concerned with the scattering characteristics of the tag's antenna. Because antennas are designed to transmit and receive radiation, they are generally regarded as having two modes of scattering. One, the structural mode, is the scattering that occurs because a given antenna is a certain shape, size, and material. The other, the antenna mode, is the scattering that occurs because the antenna was designed to transmit RF energy and has a specific radiation pattern. It should be noted that there are no formal definitions of these modes, and variations exist [17][13].

There have been several models given for the field scattered from an antenna [18][19]. I will use the one presented by Green [19]. Referring to Figure 2.4, the scattered field as a function of load impedance Z_L is given by

$$\mathbf{E}(Z_L) = \mathbf{E}(Z_A^*) + \Gamma I(Z_A^*)\mathbf{E}^r \tag{2.59}$$

where the modified reflection coefficient gamma is given by

$$\Gamma = \frac{Z_A^* - Z_L}{Z_A + Z_L} \tag{2.60}$$

and E_r is the far-field electric field when the antenna is excited by a unit current source, given by

$$\mathbf{E}^r = -j\frac{\eta\beta}{4\pi r}e^{-j\beta r}\mathbf{h} \tag{2.61}$$

Comparing this equation to the far-field electric field for an ideal dipole Eq. (2.13), we see that $\mathbf{h}$ = $dl\sin\theta\boldsymbol{\theta}$ in the case of an ideal dipole. This will be different for practical antennas. The negative sign represents the reflection. $\mathbf{E}(Z_A^*)$ and $I(Z_A^*)$ in Eq. (2.59) represent the scattered field and antenna current when the load impedance is a conjugate match of the antenna impedance, respectively. The second term in Eq. (2.59) is the antenna mode scattering. It is related directly to the load impedance connected to the antenna. When the load impedance is matched with the antenna impedance, this term goes to zero, leaving the first term called the structural mode scattering.

Analytical derivation of the structural mode scattering can be difficult and will typically require methods such as the method of moments. If we choose to neglect it, the RCS of the antenna mode scattering can be given in terms of the familiar antenna parameters,

$$\sigma_{ant} = p\Gamma^2 G^2(\theta,\phi)\frac{\lambda^2}{4\pi} \tag{2.62}$$

where p is the polarization factor from Eq. (2.52) accounting for a mismatch in polarization between the incident field and the scattering antenna. Additionally, Green considers the case where the load is switched between two different loads with modified reflection coefficients Γ_1 and Γ_2. In effect, this cancels the static structural mode. The modulation of the scattered field becomes

$$\Delta\mathbf{E} = -j\frac{\eta_0}{4\lambda R_A}\mathbf{h}\left(\mathbf{h}\cdot\mathbf{E}^{inc}\right)\left(\Gamma_1 - \Gamma_2\right)\frac{e^{-j\beta r}}{r} \tag{2.63}$$

The associated radar cross section becomes

$$\sigma_\Delta = p^2\left|\Gamma_1 - \Gamma_2\right|^2 G^2(\theta,\phi)\frac{\lambda^2}{4\pi} \tag{2.64}$$

These equations show the basis of backscatter modulation. By varying the load impedance, tags are able to modulate the amplitude and phase of the backscattered field.

2.5.2.4.4 Power Scattered and Power Received

As we may only vary the antenna mode scattering through varying the effective load impedance of the tag, we may ignore the structural mode scattering. It would be desirable to limit its scattering so as to reduce the noise within the channel. This, though, is difficult as such matters might either complicate the tag, or reduce the ability of the tag to receive efficiently.

From Eq. (2.56), we found that some of the received power is delivered to the load, while the rest is scattered. The fraction delivered is q, while that reflected back to the antenna is $1 - q$. The quantity Γ from Eq. (2.60) is related to q by

$$1-q=|\Gamma|^2 \tag{2.65}$$

Thus we should see the relationship between Eq. (2.62) and Eq. (2.56). From the antenna mode RCS, we can find the power radiated back to the initial transmitting antenna, in the case of RFID, the reader. Essentially, the power transmitted is captured by the antenna mode RCS and then radiated isotropically. Using a basic form of the radar range equation,

$$P_{tr} = p_t \frac{P_t G_t^2 \lambda^2}{(4\pi)^3 R^4} \sigma \tag{2.66}$$

where p_t is a polarization mismatch factor between the scattered wave and the final receiver, the backscattered power received by the transmitting antenna becomes

$$P_{tr} = pp_t \Gamma^2 \frac{P_t G_t^2 G_r^2 \lambda^4}{(4\pi R)^4} \tag{2.67}$$

Through Eq. (2.65), we find

$$P_{tr} = pp_t (1-q) \frac{P_t G_t^2 G_r^2 \lambda^4}{(4\pi R)^4} \tag{2.68}$$

Thus we see how power delivered to the tag load and the tag-scattered power received by the reader is related.

2.6 Environmental Influences

Until now, we have purely considered the behavior of electromagnetic fields and waves in free space. Under such conditions, the atmosphere is uniform and non-absorbing, and no objects surround or interfere with transmission and reception. In practice, however, the environment will differ dramatically from free space. Properties of the channel media, such as temperature and humidity, and interactions with various materials are particularly important. We will briefly discuss these issues here. Interference and noise, which are also environmental influences, will be discussed briefly here and further in Chapter 7.

2.6.1 *Loss*

Because the electromagnetic field behavior differs radically between the near-field and the far-field, we will consider them separately.

2.6.1.1 Near-field Loss

Because of the field strength decay of $1/r^3$ or $1/r^2$ in near-field systems, as opposed to $1/r$ in far-field systems, near-field cells are relatively well defined [20]. This, in a sense, precludes long range operation, yet it also in some ways lessens affects on field strength, patterns, and overall performance. Near-field inductive and capacitive systems, however, are affected strongly by surrounding materials – in particular conductive materials.

Inductively coupled systems rely on the magnetic fields, which are sensitive to currents. When large pieces of conducting materials are exposed to a time varying-magnetic field, currents are induced in the material. These currents, called eddy currents, oppose the variation of the magnetic field that induced them and result in ohmic losses. Eddy currents can severely weaken the magnetic fields, and in the case of RFID systems, can reduce range and de-tune antennas. Systems with high Q are particularly sensitive. To prevent eddy currents, powdered-iron or a class of materials called ferrites are used to shield large conductors [6]. Ferrites have a low electrical conductivity and are commonly used in transformer cores as their eddy currents and ohmic losses are reduced.

Capacitively coupled systems also have problems related to surrounding materials. Because capacitive systems rely on the electric field, however, effects are caused by charge differentials. Essentially any grounded object will attract the electric field lines of a reader's antenna and radically alter the cell coverage pattern.

2.6.1.2 Far-field Loss and Multipath

In the far-field, radiated electromagnetic waves propagate through the environment. Field strengths decay as $1/r$, so long ranges are more easily achieved. The consequence of the free radiation and low decay is increased susceptibility to interference from radiation from the same and/or other sources due to scattering, reflection, and diffraction.

Far-field losses can be characterized by large-scale affects and small-scale affects. Large-scale affects are those variations of field strength over larger distances. Small-scale affects generally include the multipath phenomenon where waves from a single source, having traveled different paths, can constructively and destructively interfere. Small-scale affects are characterized by rapid fluctuations over short distances. Large-scale path loss models and small scale fading models, respectively, describe these affects [21].

Large-scale path loss models describe how power is attenuated with distance from a transmitter. They modify the typical inverse square law relationship for free space (Eq. (2.53)) and provide for the attenuation due to atmosphere and material interaction. Setting

$$PL(R) = \left(\frac{\lambda}{4\pi R} \right)^n \tag{2.69}$$

where for free space, n is 2. Equation 2.41 becomes

$$P_r = pP_tG_tG_rPL(R) \tag{2.70}$$

A commonly used path loss model for estimating path loss in indoor environments is given by the Log-distance Path Loss Model [21],

$$PL(R)[dB] = PL(R_o) + 10n\log\left(\frac{R}{R_0}\right) + X_\sigma \tag{2.71}$$

where n depends on the surrounding and building and X_σ represents a normal random variable with standard deviation of sigma dB. The quantity R_0 is a reference distance from which to base the path loss measurements and is usually taken to be 1 meter in indoor environments. Table 2.1 shows values for n and X_σ in various environments at various frequencies. Low X_σ values represent a more accurate model.

Table 2.1: Log-distance Path Loss Model parameters n and σ for indoor wave propagation [23].

Building	Frequency (MHz)	n	σ
Retail Stores	914	2.2	8.7
Grocery Store	914	1.8	5.2
Office, hard partition	1500	3.0	7.0
Office, soft partition	900	2.4	9.6
Office, soft partition	1900	2.6	14.1
Factory LOS			
Textile/chemical	1300	2.0	3.0
Textile/chemical	4000	2.1	7.0
Paper/cereals	1300	1.8	6.0
Metalworking	1300	1.6	5.8
Suburban home			
Indoor to street	900	3.0	7.0
Factory OBS			
Textile/chemical	4000	2.1	9.7
Metalworking	1300	3.3	6.8

Small-scale fading models describe the multipath phenomenon. Multipath can cause large fluctuations in both amplitude and phase over short distances, random frequency modulations, and time dispersion caused by delays. In environments containing metals and other reflective objects, multipath can be severe. Various statistical models are used to describe multipath, and a common measure is the root-mean-square (RMS) delay spread. Buildings with few metals and hard partitions typically have small RMS delay spreads in the range of 30 to 60 ns. Larger buildings with more metal and open aisles can have delay spreads as large as 400 ns [22]. Techniques employed to lessen the effects of multipath are equalization, diversity and channel coding. In the case of RFID, readers may employ antenna diversity and equalization. Tags, however, due to severe constraints on size, complexity, and cost, will typically not employ any of these techniques.

2.6.2 Other Antennas

When multiple tags are at close range, coupling between antennas can have detrimental effects on the transfer of power. In the near-field, closely located inductive tags can cause detuning of neighboring antennas. Tags with high quality factors are particularly sensitive. One common technique is to tune antennas of tags expected to be in close proximity with other tags to a higher frequency. In the far-field, radiation patterns can be severely distorted, and power transfer efficiency can subsequently be reduced.

2.6.3 Temperature and Humidity

Variations in temperature can cause variations in the parameters of matching circuits and subsequent inefficiencies in transferring power. Systems with high quality factors can suffer serious performance degradation and detuning due to a shift in the resonant frequency. Components with low temperature coefficients should be used when possible. Humidity can also degrade performance. Effects are generally more detrimental at higher frequencies.

2.7 Summary

In this chapter, we have seen how RFID systems must achieve power and communications coupling through electromagnetic fields and waves. We have seen how electromagnetic fields and waves are created by antennas. We have studied the characteristics of the near-field and far-field zones associated with an antenna. We have studied the means of energy coupling and communications coupling in each of these regions. Finally, we have examined some the environmental influences on field and wave behavior. In the next chapter, we will see how information can be exchanged through variations of electromagnetic waves and how this constrains the design and function of RFID systems.

Chapter 3

COMMUNICATIONS

3.1 Introduction

Communications between reader and tag is essential to any RFID system. I have previously discussed the electromagnetic principles by which communications must be achieved. Now I will discuss the principles of communications themselves.

After briefly reviewing communications systems analysis techniques, I will discuss the principles of coding and modulation as applied to RFID communications. I will conclude with a look at data integrity concerns, including bit error rates, and error detection methods.

In the transmission of information, we are generally concerned with three main variables, bandwidth, probability of error, and detector cost and complexity [23]. I will focus on these three variables throughout this chapter.

3.2 Communications Process

The process of communications involves the transmission and reception of information. Information is processed for transmission over a noisy channel, mapped onto a carrier signal, and finally transmitted. Once the noisy signal has been received, it is de-modulated, and finally processed to recover the original information.

In the case of RFID, communicated information includes both control commands and binary data. Commands, though often assigned to particular strings of binary data may also be assigned to some unique modulation signature. In such cases, limited signal processing is necessary. In the case of data, however, it is typically beneficial to employ a coding scheme.

Once data has been coded, it must be mapped onto a carrier signal. This is called modulation. Modulation is necessary to achieve successful transmission through the medium of a channel and compliance with frequency spectrum regulations.

3.3 Signal and Spectra Fundamentals

In this section we will consider the fundamentals of signal analysis. Wewill review the concepts of time averaging, power, and root-mean-square value of a waveform. We will then consider the spectra and power spectral densities of various waveforms. This will be useful in understanding coding and modulation techniques covered in subsequent sections.

3.3.1 *Time Averaging, Power, and RMS*

Before reviewing the spectral characteristics of waveforms, one should understand the various ways of describing their magnitude with respect to time. I will briefly review the concepts of time averaging, normalized power, and root-mean-square value of a waveform.

The time average or DC value of a waveform *w(t)* is given by

$$\langle w(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} w(t) dt \tag{3.1}$$

For a periodic waveform, this simplifies to

$$\langle w(t) \rangle = \frac{1}{T_0} \int_{-T_0/2}^{T_0/2} w(t) dt \tag{3.2}$$

where T_0 is the period of the waveform.

Whether the time-domain waveform *w(t)* carries units of volts, amps, volts/m, or amps/m, squaring it gives its normalized power in watts (any resistive term is assumed to be one). The normalized average power of a waveform is

$$P = \langle w^2(t) \rangle = \lim_{T \to \infty} \frac{1}{T} \int_{-T/2}^{T/2} w^2(t) dt \tag{3.3}$$

The square root of the average normalized power gives the root-mean-square (RMS) value of the waveform

$$W_{rms} = \sqrt{\langle w^2(t) \rangle} \tag{3.4}$$

The concepts of time averaging, normalized power, and RMS will be important when considering the spectra of signals.

3.3.2 *Fourier and Spectra*

Every waveform is made up of a spectrum of frequencies. The Fourier transform allows us to identify these frequencies. Such identification is particularly useful, not only for determining compliance with regulations, but also compatibility with system components. For example, antennas have some characteristic spectrum over which they may efficiently radiate or receive. Complete transmission or reception of a waveform with frequency components outside of this spectrum will not be possible. In a sense, antennas often have their own intrinsic filters. In this section, I will briefly review the principles and terms used to describe signals and their associated frequency spectra.

3.3.2.1 Rectangular Pulse

If we consider a rectangular pulse *w(t)* with some pulse width T and amplitude A in the time-domain (Figure 3.1a), we can use the Fourier transform to convert it to the frequency domain [24],

$$W(f)=\int_0^T Ae^{-j2\pi ft}dt=\frac{jA}{2\pi f}\left(e^{-j2\pi fT}-1\right)=ATe^{-j\pi fT}\left(\frac{\sin\pi fT}{\pi fT}\right) \tag{3.5}$$

This function carries units of the units of the time-domain waveform per hertz. In the case where *w(t)* is a voltage waveform, *W(f)* carries units of volts/hertz. The magnitude of the spectrum of the rectangular pulse in the frequency domain simplifies to

$$|W(f)|=\frac{A}{2\pi|f|}\left[2-2\cos\left(2\pi fT_r\right)\right]^{\frac{1}{2}} \tag{3.6}$$

This is called the magnitude spectrum and it is plotted in Figure 3.1a

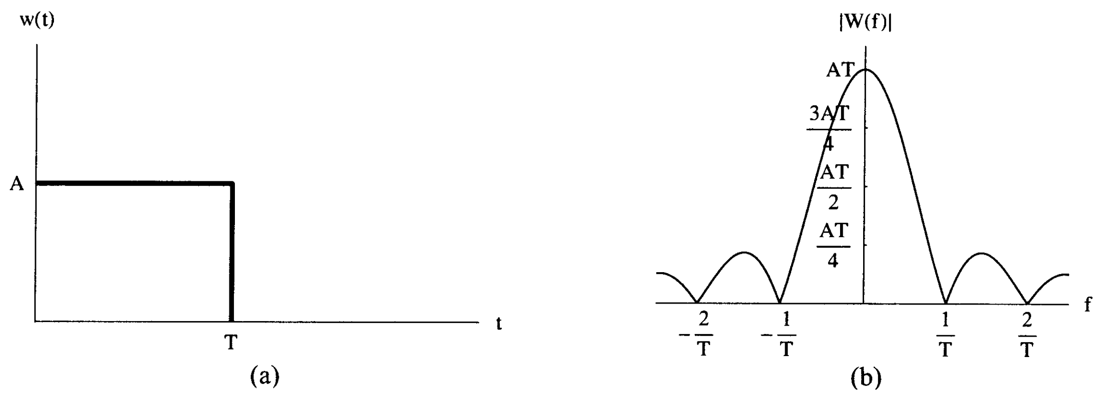


Figure 3.1: (a) Time domain plot of rectangular pulse. (b) Magnitude spectrum of rectangular pulse.

There are frequency nulls at every multiple of the inverse of the pulse width, and peaks half-way between the nulls. The peak at zero has a magnitude of AT_r, the other peaks have magnitudes of

$$\frac{2AT_r}{n\pi} \qquad |n|=3,5,7,... \tag{3.7}$$

The total area under the magnitude spectrum gives the average waveform unit.

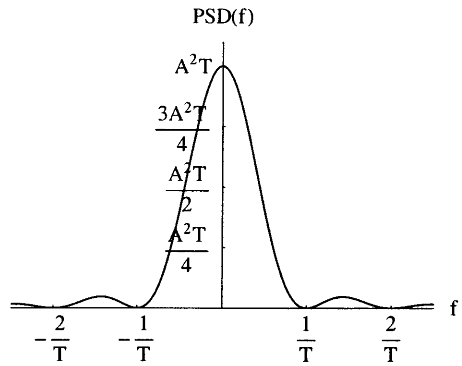


Figure 3.2: Power spectral density plot of a rectangular pulse.

From the magnitude spectrum, we can find the power spectral density of the waveform truncated over a period T,

$$PSD \triangleq \lim_{T \to \infty} \left(\frac{|W_T(f)|^2}{T} \right) \tag{3.8}$$

The power spectral density for the rectangular pulse is shown in Figure 3.2. The total area under the power spectral density curve gives the time-averaged power of the waveform.

3.3.2.2 Rectangular Pulse Train

If we consider a train of rectangular pulses with pulse width T_r and period T_0 over all time as shown in Figure 3.3, we can find its spectrum through using the Fourier series.

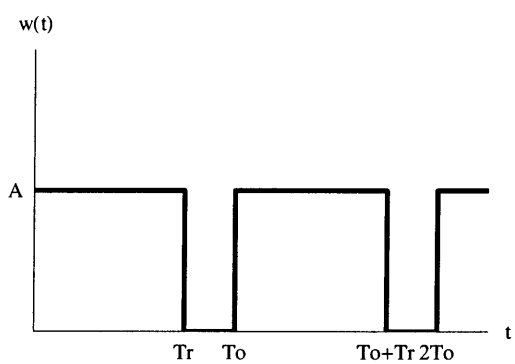


Figure 3.3: Rectangular pulse train with $T_r = \frac{3}{4} T_0$.

For a periodic waveform, the spectrum is

$$W(f) = \sum_{n=-\infty}^{n=\infty} c_n \delta\left(f - nf_0\right) \tag{3.9}$$

where $f_0 = 1/T_0$, and we relate T_0 to T_r with some constant r by $T_r = rT_0$. The phasor Fourier coefficients, c_n, of the waveform become

$$c_n = \frac{1}{T_0}\int_0^{rT_o} Ae^{-j2\pi f_0 nt}dt = j\frac{A}{2\pi n}\left(e^{-j2\pi nr} - 1\right) \tag{3.10}$$

These reduce to

$$c_n = \begin{cases} Ar & n = 0 \\ j\dfrac{A}{2\pi n}\left(e^{-j2\pi nr} - 1\right) & n \text{ otherwise} \end{cases} \tag{3.11}$$

The magnitudes become

$$|c_n| = \begin{cases} Ar & n = 0 \\ \dfrac{A}{2\pi |n|}\left[2 - 2\cos\left(2\pi nr\right)\right]^{\frac{1}{2}} & n \text{ otherwise} \end{cases} \tag{3.12}$$

The magnitude spectrums when $r = 0.25$, $r = 0.5$, and $r = 0.75$ are plotted in Figure 3.4. Spectral lines appear at identical frequencies (multiples of f_0), yet vary in magnitude based on the envelope of the continuous spectrum of the single rectangular pulse with width T_r.

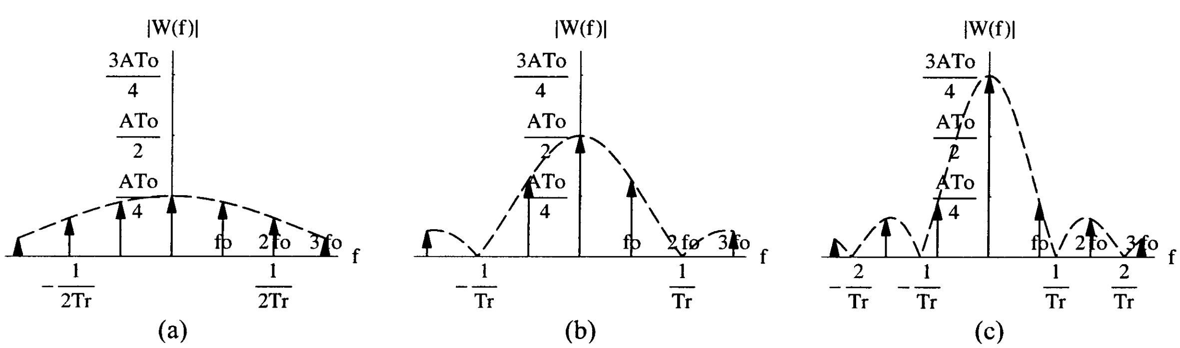


Figure 3.4: Magnitude spectra for three ratios of T_r to T_0: (a) ¼ (b) ½, (c) ¾.

The band of frequencies that the spectral lines cover also varies with the envelope of the continuous spectrum. We see that when the pulse width narrows, the magnitude of the spectral lines drop, yet they disperse themselves over a wider band of frequencies. This may present problems for reception by narrow band antennas, reduced signal to noise power, and problems with regulatory compliance.

The power spectral density for these periodic pulses can be found by squaring the magnitude spectrum coefficients.

$$PSD(f) = \sum_{n=-\infty}^{n=\infty} |c_n|^2 \delta\left(f - nf_0\right) \tag{3.13}$$

The sum of these spectral lines from negative to positive infinity gives the average power.

3.3.2.3 Filtering

Another very important principle in evaluating communications systems is the principle of linear-time invariance [24]. Linear-time invariant (LTI) systems are those that are linear, meaning that the principle of superposition holds, and time invariant in that for a given delay in the input, the output is delayed by the same amount. With these properties, we can evaluate more complex waveforms including filtering effects and more advanced modulation techniques. I will briefly describe the application of LTI principles to filtering.

When the input into a linear system, such as a filter, is an impulse (a delta function), the output will be the impulse response of that system. The impulse response of a system is useful as it allows us to directly determine the output of a system given the input. Convolving the input $x(t)$ with the impulse response $h(t)$ gives the output $y(t)$. In the frequency-domain the convolution operation becomes multiplication, so

$$Y(f) = X(f)H(f) \tag{3.14}$$

where we call $H(f)$ the frequency response, or transfer function of the system as it transfers input $X(f)$ to output $Y(f)$.

Now, if we consider our rectangular pulse train from Eq. (3.9) and pass it through some system with frequency response $H(f)$, we find

$$Y(f) = \sum_{n=-\infty}^{n=\infty} c_n H(nf_o)\delta(f - nf_o) \tag{3.15}$$

In the case of a RC low-pass filter, we can find the frequency response through Kirchoff's circuital laws, Fourier transformation, and some simplification. The frequency response becomes

$$H(f) = \frac{1}{1 + j(2\pi RC)f} \tag{3.16}$$

Substitution into Eq. (3.14) gives

$$Y(f) = \sum_{n=-\infty}^{n=\infty} c_n \frac{1}{1 + j(2\pi RC)nf_o} \delta(f - nf_o) \tag{3.17}$$

where coefficients c_n from Eq. (3.10) still hold. Considering the rectangular pulse train with $r = 0.5$, and RC time constant of $T_r/6$, we can plot the magnitude spectrum of the output $|Y(f)|$ as shown in . We see that the higher frequency components have dropped below the envelope of an unfiltered rectangular pulse.

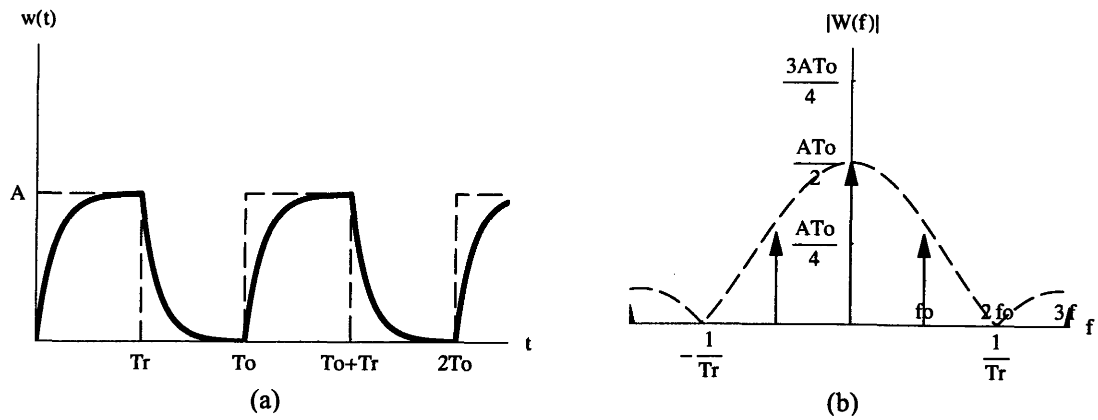


Figure 3.5: Low-pass filtered rectangular pulse: (a) waveform (b) magnitude spectrum.

3.3.2.4 Modulation

The signals we have considered thus far have been baseband signals in that their spectra have been centered around zero frequency. When transmitting information over wireless channels it is most often necessary to map the baseband signal onto a carrier signal through modulation. This creates a bandpass signal with spectra centered around the carrier frequency.

Though I will discuss modulation in a subsequent section, it is useful to examine the general spectral characteristics of a modulated waveform. For a general baseband waveform *g*(t), the modulated bandpass signal with carrier frequency ω_c is [24]

$$v(t) = \mathrm{Re}\left\{g(t)e^{j\omega_c t}\right\} = g(t)\cos\left(\omega_c t\right) \tag{3.18}$$

The spectrum is

$$V(f) = \tfrac{1}{2}\left[G(f-f_c)+G^*(-f-f_c)\right] \tag{3.19}$$

and the PSD is

$$PSD_v(f) = \tfrac{1}{4}\left[PSD_g(f-f_c)+PSD_g(-f-f_c)\right] \tag{3.20}$$

We notice that the act of modulation centers the spectrum of the baseband waveform around both the positive and negative carrier frequency while halving the magnitudes of each. As we would expect there is a similar effect on the power spectral density.

3.3.2.5 Bandwidth

The concept of bandwidth is of particular importance from the standpoint of data rates, signal power relative to noise power, and regulatory compliance. In general, bandwidth refers to the range of positive frequencies that a signal occupies, but there is no single formal definition. I will give definitions of some types we will encounter in further discussions [24]:

- Absolute bandwidth is the range of frequencies outside of which the magnitude spectrum is zero.
- N-dB bandwidth, where n is some value, typically 3, 20, or 60, is the range of frequencies where the magnitude spectra fall no lower than n-dB below the maximum value.
- Null-to-null bandwidth for bandpass systems is the range of frequencies from the first null above and below the center frequency. For baseband systems it is typically from 0 to the first null frequency.

When discussing signals on relative terms, it may not be necessary to specify the type of bandwidth. However, when discussing signals on absolute terms, the type of bandwidth should be specified.

3.4 Line Coding

In digital communications systems, digital data symbols must be converted into some transmittable form. This is typically done by sending trains of pulses that are formatted to represent the data symbols. This pulse formatting is often called line coding. I will briefly discuss line coding in the context of wireless transmission and RFID.

Line coding originated in telephony, where transmission was over a copper wire line [25]. It applies, however, to wireless transmission as well. Figure 3.6 shows a number of the more common coding techniques. Depending on implementation, there may be two or three levels – positive and zero, or positive, negative, and zero. Those signals using both positive and negative levels are referred to as bipolar or polar, while those using only positive and zero are referred to as unipolar. Because line coding is used in wired baseband transmission, these levels have significant meanings. As RFID systems transmit bandpass signals, however, the coded pulse train will be modulated onto a carrier. Depending on the modulation scheme used, it may be necessary to convert a bipolar scheme to a unipolar scheme, as is the case with amplitude shift key. Because coding and modulation together determine bandwidth, probability of error, and receiver complexity, I will save much analysis until I discuss modulation. I will however, derive and plot the power spectral densities of the coding schemes of interest.

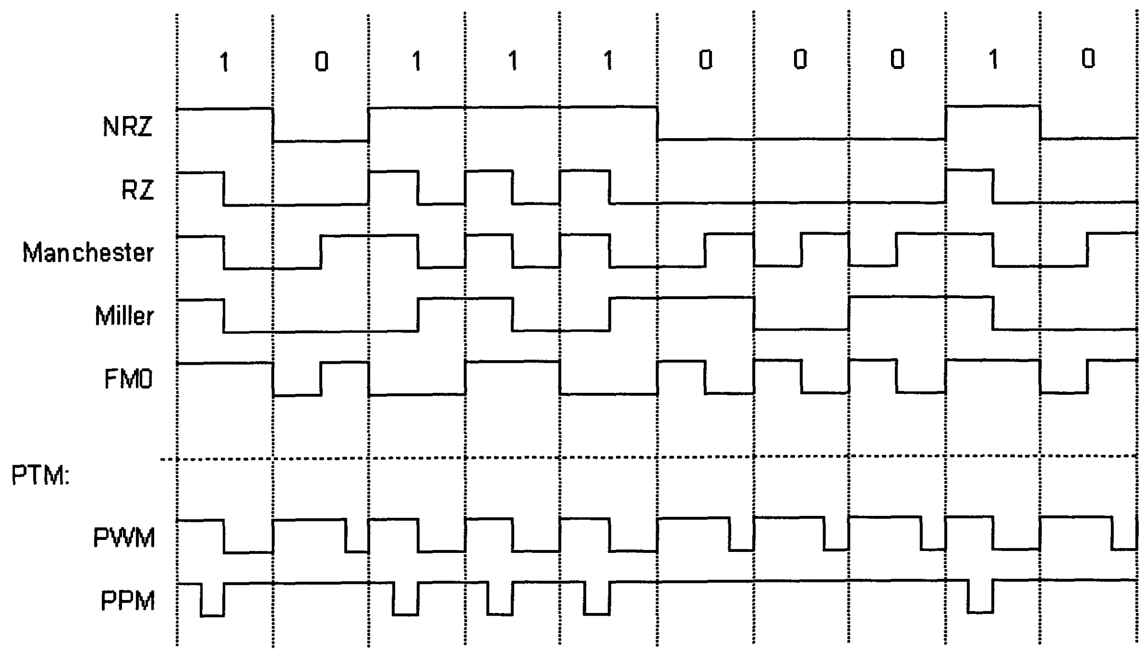


Figure 3.6: Some commonly used coding schemes in RFID systems.

There are two main classes of line codes: level codes and transition codes [25]. Level codes carry information in the signal level and are usually instantaneous in that they typically encode data independent of previously coded data. Non-return-to-zero (NRZ) and return-to-zero (RZ) are common examples. Transition codes on the other hand, carry information in the signal transitions. They may be instantaneous, but will often exhibit memory by encoding data based on previously coded data. Manchester (Split Phase), Miller (Delay Modulation), and FM0 (Biphase-Space) are all line codes commonly used in RFID. Miller and FM0 exhibit memory whereas Manchester does not. Pulse time modulation (PTM) schemes, including pulse-width modulation (PWM) and pulse position modulation (PPM) vary pulse width or position to convey information. They are not line codes, yet they are also a class of baseband coding schemes.

Manchester coding, also known as split phase, or digital biphase is an instantaneous transition code. The name diphase arises from the fact that it is essentially a square wave where 0 degree phase represents a binary 1, and a square wave with a phase of 180 degrees represents a 0. This may present problems for some receivers. Manchester is used in Ethernet local area networks where it is useful because in bipolar form it has a null at DC. The mid-bit transitions are helpful for timing. It however has a higher bandwidth than many other common codes and no inherent error detection capability [25].

Miller coding is another transition code, yet unlike Manchester, it exhibits memory. It provides good timing without the bandwidth increase of Manchester. It also is more suited for modulation since there are no phase reversal problems.

FM0, also known as Biphase-Space, is another transition code that exhibits memory. It is similar to Miller coding, yet it specifies a transition at the beginning of every bit period. This allows for easier synchronization at the cost of higher bandwidth.

Pulse time modulation (PTM) is a class of coding schemes that typically encodes an analog signal onto the time scale of a digital signal [24]. It includes pulse width modulation (PWM) and pulse position modulation (PPM). It is often used in RFID systems by encoding digital signals onto the time scale of a transmission. In pulse position modulation, a modulation for some interval within a frame might represent a 1 while no modulation might represent a 0. In pulse width modulation, binary 1s and 0s would each be assigned different pulse widths defined from a common edge. Such schemes can be beneficial from the standpoint of bandwidth, synchronization, and probability of error.

These are only some commonly used codes, most of which were designed for wired systems. As the transmission, channel, and reception is much different for wireless systems, it may be advantageous to define a coding scheme unique to the application. Pulse time modulation schemes, in general, allow this flexibility. I will analyze these various coding schemes in combination with modulation in the next section.

3.4.1 Baseband Power Spectral Densities

Though the coded signal will ultimately modulate a carrier, it is still useful to examine its baseband spectral characteristics. Depending on the modulation scheme used, the spectral characteristics will likely remain the same, only the spectrum will be centered around the carrier frequency and decreased in magnitude as described by Eq. (3.19). As mentioned, many of these coding schemes were originally designed for bipolar operation. Consequently most analyses of their spectral characteristics are based on their unipolar form. For reasons outlined in Section 3.4, I will derive their power spectral densities in unipolar form.

There are two approaches to computing the power spectral density of a signal – deterministic methods and probabilistic methods. Deterministic methods, as those presented in Section 3.2.2, require prior knowledge of the exact waveform. Probabilistic methods, however, assign probabilities to certain elementary waveforms of which the signal is composed. As I will assume a random signal, probabilistic methods are appropriate.

Often a random process, in our case, a signal, will have some special structure that allows it to be characterized with relatively few specifications. The Markov process is one of these special processes. Markov processes describe random processes where the future state depends only on the present state and not on the past history. This is true, in general, of the coded signals we will consider. When composed of discrete states, a Markov process is called a Markov chain [26].

Following the approach of [27] based on [28] we will consider a Markov signal composed of N elementary waveforms of the set $\{s_i(t);\ i = 1, 2, \ldots, N\}$ with common period T_s. A signal's sequence is characterized by a set of stationary probabilities $\{p_i;\ i = 1, 2, \ldots, N\}$ and a set of transition probabilities $\{p_{ik};\ i = 1, 2, \ldots, N\}$. Stationary probabilities indicate the probability of a particular elementary waveform occurring, while transition probabilities indicate the probability of waveform s_k following waveform s_i. Formally, transition probabilities are defined by

$$p_{ik} = P\{s_{n+1} = k \mid s_n = i\} \quad i \geq 1, k \geq 1 \tag{3.21}$$

They can be arranged in a transition probability matrix **P** defined by

$$\mathbf{P} \triangleq \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{bmatrix} \tag{3.22}$$

Where the elements must satisfy

$$p_{ik} \geq 0 \quad \sum_{k=0}^{\infty} p_{ik} = 1 \quad i = 1,2,\ldots,N \tag{3.23}$$

Based on this statistical description, the power spectral density of the signal is given by [28]

$$\begin{aligned} PSD(f) &= \frac{1}{T_s^2} \sum_{n=-\infty}^{\infty} \left| \sum_{i=1}^{N} p_i S_i\left(\frac{n}{T_s}\right) \right|^2 \delta\left(f - \frac{n}{T_s}\right) \\ &+ \frac{1}{T_s} \sum_{i=1}^{N} p_i \left|S_i(f)\right|^2 \\ &+ \frac{2}{T_s} \mathrm{Re}\left[\sum_{i=1}^{N} \sum_{k=1}^{N} p_i S_i^*(f) S_k(f) p_{ik}\left(e^{-j2\pi f T_s}\right) \right] \end{aligned} \tag{3.24}$$

where S_i is the Fourier transform of s_i as given by W in Eq. (3.5), S_i^* is the complex conjugate of S_i, and

$$p_{ik}(z) \triangleq \sum_{n=1}^{\infty} p_{ik}^{(n)} z^n \tag{3.25}$$

The quantity $p_{ik}(n)$ is defined as the probability that s_k is transmitted n intervals after s_i. It is the ikth element of the matrix $\mathbf{P}^n$. Also, by definition, $p_{ik}^{(1)} \triangleq p_{ik}$.

For signals coded with line codes exhibiting no memory, the waveform sequence is purely random and the future waveform has no dependence on the present waveform. In this case, $\mathbf{P}^n = \mathbf{P}$ for all $n \geq 1$, and equation simplifies to

$$\begin{aligned} PSD(f) &= \frac{1}{T_s^2} \sum_{n=-\infty}^{\infty} \left| \sum_{i=1}^{N} p_i S_i\left(\frac{n}{T_s}\right) \right|^2 \delta\left(f - \frac{n}{T_s}\right) \\ &+ \frac{1}{T_s} \sum_{i=1}^{N} p_i(1-p_i) \left|S_i(f)\right|^2 \\ &- \frac{2}{T_s} \sum_{i=1}^{N} \sum_{\substack{k=1 \\ i \neq k \\ i < k}}^{N} p_i p_k \, \mathrm{Re}\left[S_i(f) S_k^*(f) \right] \end{aligned} \tag{3.26}$$

Further, when $N = 2$, and $p_1 = p$, $p_2 = 1 - p$, Eq. (3.26) becomes

$$PSD(f) = \frac{1}{T_s^2}\sum_{n=-\infty}^{\infty}\left|pS_1\left(\frac{n}{T_s}\right)+(1-p)S_2\left(\frac{n}{T_s}\right)\right|^2\delta\left(f-\frac{n}{T_s}\right)$$
$$+\frac{1}{T_s}p(1-p)\left|S_1(f)-S_2(f)\right|^2 \tag{3.27}$$

Through these relations, I will derive and plot the power spectral densities for unipolar NRZ, RZ, Manchester, Miller, FM0, and PWM. I will first consider those codes without memory, NRZ, RZ, Manchester, PWM, and PPM, followed by those with memory, Miller, and FM0.

3.4.1.1 NRZ

NRZ has two elementary waveforms given by

$$\begin{aligned} s_1(t) &= A \qquad 0 \le t \le T_s \\ s_2(t) &= 0 \qquad 0 \le t \le T_s \end{aligned} \tag{3.28}$$

The Fourier transforms are

$$\begin{aligned} S_1(f) &= AT_s e^{-j\pi f T_s}\frac{\sin(\pi f T_s)}{\pi f T_s} \\ S_2(f) &= 0 \end{aligned} \tag{3.29}$$

Substituting into Eq. (3.27) and normalizing to energy $E_s = A^2T$, we get

$$\frac{PSD(f)}{E_s} = p(1-p)\frac{\sin^2(\pi f T_s)}{(\pi f T_s)^2} \tag{3.30}$$

When the binary signals are equiprobable, p = ½, and Eq. (3.30) becomes

$$\frac{PSD(f)}{E_s} = \frac{\sin^2(\pi f T_s)}{(2\pi f T_s)^2} \tag{3.31}$$

3.4.1.2 RZ

Unipolar RZ is similar to NRZ except the binary 1 waveform returns to zero half-way through the bit period. The two elementary waveforms are given by

$$\begin{aligned} s_1(t) &= A \qquad 0 \le t \le T_s/2 \\ s_2(t) &= 0 \qquad 0 \le t \le T_s \end{aligned} \tag{3.32}$$

The Fourier transforms are

$$S_1(f) = \frac{AT_s}{2} e^{-j\pi f T_s/2} \frac{\sin(\pi f T_s/2)}{\pi f T_s/2}$$
$$S_2(f) = 0 \tag{3.33}$$

Substituting into Eq. (3.27) and normalizing to energy $E_s = A^2T$, we get

$$\frac{PSD(f)}{E_s} = \frac{p^2}{4T_s}\delta(f) + \frac{p^2}{4T_s}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty}\left(\frac{2}{n\pi}\right)^2 \delta\left(f - \frac{n}{T_s}\right) + \frac{p(1-p)}{4}\frac{\sin^2(\pi f T_s/2)}{(\pi f T_s/2)^2} \tag{3.34}$$

When the binary signals are equiprobable, $p = ½$, and Eq. (3.34) becomes

$$\frac{PSD(f)}{E_s} = \frac{1}{16T_s}\delta(f) + \frac{1}{16T_s}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty}\left(\frac{2}{n\pi}\right)^2 \delta\left(f - \frac{n}{T_s}\right) + \frac{1}{16}\frac{\sin^2(\pi f T_s/2)}{(\pi f T_s/2)^2} \tag{3.35}$$

3.4.1.3 Manchester

In unipolar form, Manchester's two elementary waveforms are given by

$$\begin{aligned} s_1(t) &= A \qquad 0 \le t \le T_s/2 \\ s_2(t) &= A \qquad T_s/2 \le t \le T_s \end{aligned} \tag{3.36}$$

The Fourier transforms are

$$\begin{aligned} S_1(f) &= \frac{AT_s}{2} e^{-j\pi f T_s/2} \frac{\sin(\pi f T_s/2)}{\pi f T_s/2} \\ S_2(f) &= \frac{AT_s}{2} e^{-j3\pi f T_s/2} \frac{\sin(\pi f T_s/2)}{\pi f T_s/2} \end{aligned} \tag{3.37}$$

Substituting into Eq. (3.27) and normalizing to energy $E_s = A^2T$, we get

$$\frac{PSD(f)}{E_s} = \frac{\pi}{64T_s}\delta(f) + \frac{1-4p(1-p)}{16T_s}\sum_{\substack{n=-\infty \\ n\neq 0, n=odd}}^{\infty}\left(\frac{1}{n\pi}\right)^2 \delta\left(f-\frac{n}{T_s}\right)$$

$$+\frac{1}{16T_s}\sum_{\substack{n=-\infty \\ n\neq 0, n=even}}^{\infty}\left(\frac{1}{n\pi}\right)^2 \delta\left(f-\frac{n}{T_s}\right) \tag{3.38}$$

$$+p(1-p)\frac{\sin^4\left(\pi f T_s/2\right)}{\left(\pi f T_s/2\right)^2}$$

When the binary signals are equiprobable, $p = ½$, and Eq. (3.38) becomes

$$\frac{PSD(f)}{E_s} = \frac{\pi}{64T_s}\delta(f) + \frac{1}{16T_s}\sum_{\substack{n=-\infty \\ n\neq 0, n=even}}^{\infty}\left(\frac{1}{n\pi}\right)^2 \delta\left(f-\frac{n}{T_s}\right)$$

$$+\frac{1}{4}\frac{\sin^4\left(\pi f T_s/2\right)}{\left(\pi f T_s/2\right)^2} \tag{3.39}$$

Normalized power spectral densities for NRZ, RZ, and Manchester are plotted in Figure 3.7.

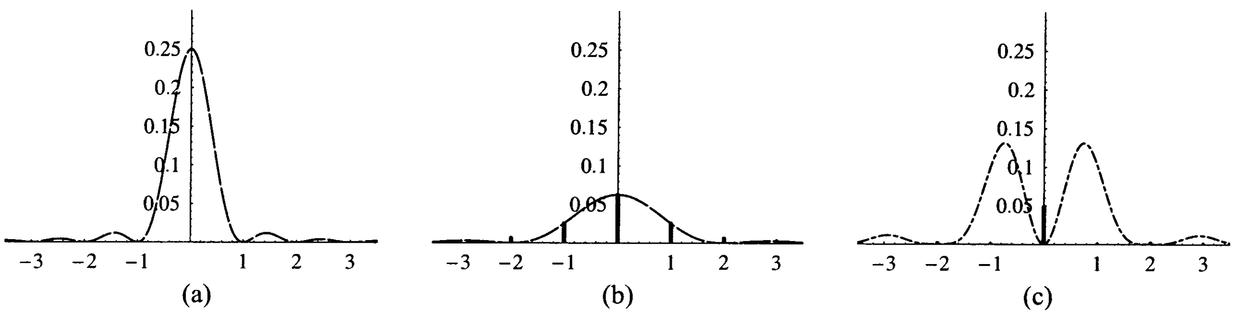


Figure 3.7: Normalized power spectral densities vs. $1/T$ for unipolar (a) NRZ, (b) RZ, (c) Manchester.

3.4.1.4 PWM and PPM

Thus far, each coding scheme has specified equal high and low pulse lengths. Depending on the link, whether it be reader to tag, or tag to reader, we may either wish to minimize or maximize the energy in the coded signal. Transmissions from the reader should maximize energy, whereas transmissions from the tag should typically minimize the energy. Because the various line schemes use equal pulse widths, energy in the signal is not optimal for powering the tag. PWM and PPM coding methods offer more flexibility. I will consider the case where energy in the signal is relatively high.

Because of the flexibility offered by PWM and PPM, depending on the application, their definitions may vary. I will consider pulses coded as pictured in Figure 3.6. This form differs from its use in other applications, yet similar forms are often used in RFID systems. I will compute the PSD only for PWM as the PSD for PPM can be quite similar.

We can find the power spectral density of a pulse width modulated signal with pulse width $T_1 = r_1 T_s$ representing a logic 1, and $T_2 = r_2 Ts$ representing a logic 0. The signal is composed of two separate signals s_1 and s_2, with probabilities of occurrence $p_1 = p$ and $p_2 = 1\text{-}p$, respectively.

$$
\begin{aligned}
s_1(t) &= \begin{cases} A & 0 \le t \le T_1 \\ 0 & T_1 \le t \le T_s \end{cases} \\
s_2(t) &= \begin{cases} A & 0 \le t \le T_2 \\ 0 & T_2 \le t \le T_s \end{cases}
\end{aligned}
\tag{3.40}
$$

Taking the Fourier transform of these rectangular pulses, we get

$$
\begin{aligned}
S_1(f) &= A r_1 T_s e^{-j\pi f r_1 T} \frac{\sin(\pi f r_1 T)}{\pi f r_1 T} \\
S_2(f) &= A r_2 T_s e^{-j\pi f r_2 T} \frac{\sin(\pi f r_2 T)}{\pi f r_2 T}
\end{aligned}
\tag{3.41}
$$

Substituting into Eq. (3.27) and normalizing to energy $E_s = A^2 T$, we get

$$
\begin{aligned}
\frac{PSD(f)}{E_s} &= \frac{\pi\left[p(r_1 - r_2) + r_2\right]^2}{2T_s}\delta(f) \\
&+ \frac{1}{T_s}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{1 - p(1-p) - p\cos(2\pi n r_1) + p(1-p)\cos\left[2\pi n(r_1 - r_2)\right] - (1-p)\cos(2\pi n r_2)}{2(n\pi)^2}\delta\left(f - \frac{n}{T_s}\right) \\
&+ p(1-p)\frac{\sin^2\left[\pi f T_s(r_1 - r_2)\right]}{(\pi f T_s)^2}
\end{aligned}
\tag{3.42}
$$

When the binary signals are equiprobable, $p = ½$, and Eq. (3.42) becomes

$$
\begin{aligned}
\frac{PSD(f)}{E_s} &= \frac{\pi\left[\frac{1}{2}(r_1 - r_2) + r_2\right]^2}{2T_s}\delta(f) \\
&+ \frac{1}{T_s}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{3 - 2\cos(2\pi n r_1) + \cos\left[2\pi n(r_1 - r_2)\right] - 2\cos(2\pi n r_2)}{8(n\pi)^2}\delta\left(f - \frac{n}{T_s}\right) \\
&+ \frac{1}{4}\frac{\sin^2\left[\pi f T_s(r_1 - r_2)\right]}{(\pi f T_s)^2}
\end{aligned}
\tag{3.43}
$$

Arbitrarily setting $r_1 = 1/2$ and $r_2 = 7/8$, Eq. (3.43) becomes,

$$\frac{PSD(f)}{E_s} = \frac{121\pi}{256T_s}\delta(f)$$

$$+\frac{1}{T_s}\sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} \frac{3+\cos\left(\frac{3\pi n}{4}\right)-2\cos(\pi n)-2\cos\left(\frac{7\pi n}{4}\right)}{8(n\pi)^2}\delta\left(f-\frac{n}{T_s}\right) \tag{3.44}$$

$$+\frac{1}{4}\frac{\sin^2\left(\frac{3\pi f T_s}{8}\right)}{(\pi f T_s)^2}$$

This particular case in plotted in Figure 3.8.

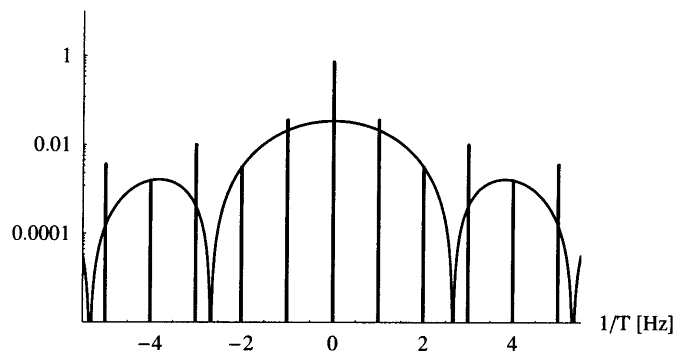


Figure 3.8: Log-scale plot of PWM normalized PSD with high pulses of width 7/8 and ½ the bit period T.

From Figure 3.8 we can see that the 20 dB bandwidth lies at the 3rd spectral line, thus the 20 dB bandwidth is *6/T*.

3.4.1.5 Miller and FM0

Derivation of the power spectral densities of the Miller and FM0 codes is more complex as these codes exhibit memory. Because the future elementary waveform is based on the present elementary waveform, we will use Eq. (3.24). I will present a deriviation of the Miller code given in [27] and [29]. Derivation of power spectral densities for unipolar Miller and FM0 codes will follow a similar form.

Figure 3.9 shows the four transitions possible in the Miller code. We can see that there are four elementary waveforms that may compose the signal.

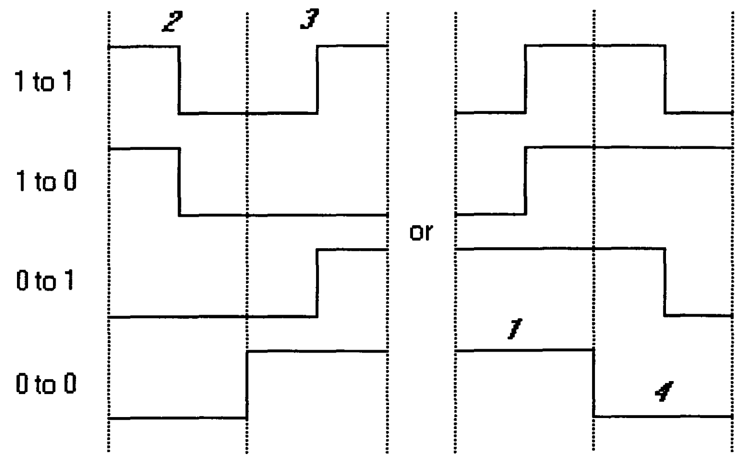


Figure 3.9: Miller code transitions and elementary waveforms denoted by *1*, *2*, *3*, and *4*.

The elementary waveforms can be given mathematically as,

$$
\begin{aligned}
s_1 &= -s_4 = A \qquad 0 \le t \le T_s \\
s_2 &= -s_3 = \begin{cases} A & 0 \le t \le T_s/2 \\ -A & T_s/2 \le t \le T_s \end{cases}
\end{aligned} \tag{3.45}
$$

In order to solve Eq. (3.24), we must find the transition probability matrix as given in Eq. (3.22). Noting from Figure 3.9 that each of the elementary waveforms has a ¼ chance of occurring and only certain transitions are possible, the transition probability matrix becomes

$$
\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix} \tag{3.46}
$$

In order to simplify the infinite series given by Eq. (3.25), we find a recursive relationship based on the autocorrelation function. The autocorrelation function for this discrete set of signals is given by

$$
s_{ik} \triangleq \frac{1}{T_s}\int_0^{T_s} s_i(t)s_k(t)dt \qquad i,k = 1,2,3,4 \tag{3.47}
$$

The signal correlation matrix is given by

$$
\mathbf{S} = \begin{bmatrix} s_{11} & s_{12} & s_{13} & s_{14} \\ s_{21} & s_{22} & s_{23} & s_{24} \\ s_{31} & s_{32} & s_{33} & s_{34} \\ s_{41} & s_{42} & s_{43} & s_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix} \tag{3.48}
$$

The recursive relationship is found to be

$$\mathbf{P}^{4+l}\mathbf{S} = -\tfrac{1}{4}\mathbf{P}^{l}\mathbf{S} \quad l \geq 0 \tag{3.49}$$

Simplifying (3.24) through these relations, the energy normalized power spectral density becomes

$$\frac{S(f)}{E_s} = \frac{1}{2\theta^2\left(17+8\cos 8\theta\right)}\begin{pmatrix} 23-2\cos\theta-22\cos 2\theta-12\cos 3\theta+5\cos 4\theta \\ +12\cos 5\theta+2\cos 6\theta-8\cos 7\theta+2\cos 8\theta \end{pmatrix} \tag{3.50}$$

This is plotted in Figure 3.10.

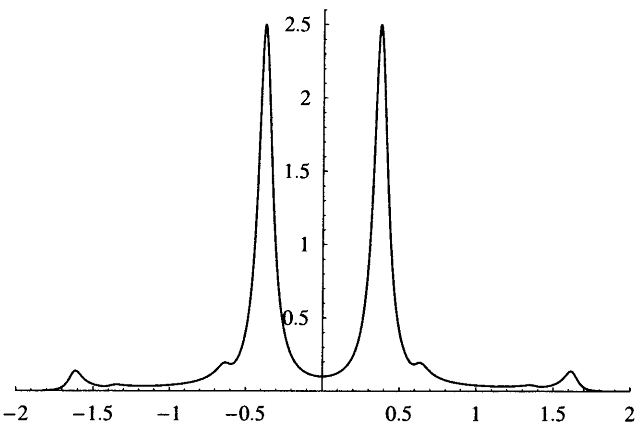


Figure 3.10: Normalized power spectral density of bipolar Miller coded signal versus $1/T$.

3.5 Modulation

The imparting of source information onto a bandpass signal with a carrier frequency f_c while introducing amplitude variations, phase variations, or both, is called modulation [24]. It is necessary to transmit information over a wireless channel. A number of analog and digital modulation techniques exist, yet they all in some way vary the carrier signals amplitude or phase. In the following discussion I will consider only binary digital modulation techniques.

There are three main types of binary digital modulation, amplitude shift key (ASK), binary phase shift key (BPSK), and frequency shift key (FSK). In all three techniques, the amplitude, phase, or frequency is shifted in accordance with an information carrying unipolar binary signal. The line coding method used determines how the unipolar binary signal represents digital data. In amplitude shift key, a carrier is simply modulated between two amplitudes. In binary phase shift key (BPSK), a carrier is shifted in phase by 180 degrees. In frequency-shift key, (FSK), the carrier is shifted between two frequencies.

In RFID systems, information is transmitted from the reader to the tag, and from the tag to the reader. As the physical constraints on the system are different, so to are the modulation techniques used. The reader must provide a signal capable of both powering the tag and dominating the noise. As the tag is limited by cost and size constraints to only simple receivers, the reader-modulated signal must also be simple. ASK is used as it allows for relatively simple detection. Detection can be coherent where the signal phase is recognized, or non-coherent where the signal phase is not recognized. Non-coherent detection is simpler to implement and less expensive. Envelope detection is a form of non-coherent detection where the envelope of the amplitude modulated signal is detected. I will discuss ASK modulation and envelope detection further in Section 3.5.1.

The performance of the tag differs radically from that of the reader. As the tag is not equipped with a transmitter, it relays information through load and backscatter modulation based on the principles discussed in Chapter 2. Both of these modulation techniques allow for variation of amplitude and phase depending on their implementation. Because the power of the tag-modulated signal is low relative to that transmitted by the reader, a subcarrier is often used to move the spectra associated with the tag's modulation further from the center frequency. Load or backscatter modulation with a subcarrier allows both FSK and BPSK. Given these differences in reader to tag and tag to reader modulation, I will treat their communications separately.

3.5.1 Reader to Tag Coding and Modulation

In reader to tag communications, we must ensure that the tag receives sufficient power, that it easily detects the signal, and that the signal meets regulatory constraints on field strength and bandwidth. The reception of power is obvious. In passive RFID, tags must receive power, if a coding and modulation scheme does not contain enough energy within its signal, the tag may not function. Signal detection has multiple concerns. The tag receiver must be simple. More complex receivers though likely more robust will add cost. The tag must also be able to synchronize to the signal. Coding schemes should allow some form of synchronization, whether it be from the trailing edge of a PWM signal, or the transitions inherent in a Manchester coded signal. Finally, the probability of error should be low. If integrity of the data is not high, no matter how optimized the data rates, communications will be worthless. In meeting regulatory constraints, not only must the coding and modulation scheme be optimized, but wave shaping and filtering should be considered as well.

To allow the simplest and most inexpensive receiver on the tag, low-cost RFID systems most often use amplitude shift-key modulation.

3.5.1.1 Amplitude Shift Key (ASK)

In binary amplitude shift key modulation, a unipolar information carrying baseband signal causes a change between two amplitudes of a carrier signal. If the lesser of the two amplitudes is zero, it is called on-off shift key (OOK). If we consider two amplitudes, A_c and A_{min}, we can define a modulation index u, where

$$u = 1 - \frac{A_{\min}}{A_c} \tag{3.51}$$

When A_{min} equals A_c, u is zero and there is no modulation. When A_{min} is zero, u is 1 and there is 100% amplitude modulation, or OOK. We can describe the ASK modulated signal $g(t)$ by

$$v(t) = A_c\left[1 + u\left(s(t) - 1\right)\right]\cos\left(\omega_c t\right) \tag{3.52}$$

where s(t) is the unipolar binary signal (1 or 0), and ω_c is the carrier frequency in radians/second. Using Eqs. (3.18) - (3.20), we can find the spectrum and power spectral densities of the modulated signal. The modulating signal $g(t)$ based on the coded signal $s(t)$ is

$$g(t) = A_c\left[1+u\left(s(t)-1\right)\right] = uA_c s(t) + A_c(1-u) \tag{3.53}$$

The spectrum of $g(t)$ is

$$G(f) = uA_c S(f) + A_c(1-u)\delta(f) \tag{3.54}$$

The power spectral density is

$$PSD_g(f) = \left(uA_c\right)^2 PSD_s(f) + \left[A_c(1-u)\right]^2 \delta(f) \tag{3.55}$$

Substituting into Eq. (3.20), the power spectral density of the modulated signal, for positive frequencies, becomes,

$$PSD_v(f) = \tfrac{1}{4}\left[\left(uA_c\right)^2 PSD_s(f-f_c) + \left[A_c(1-u)\right]^2 \delta(f-f_c)\right] \quad f \gg 0 \tag{3.56}$$

We note that the bandwidth is dependent purely on the power spectral density of the baseband signal while its magnitude is dependent on the carrier amplitude and the modulation index. At a modulation index of 1, the power spectral density of the modulated signal is simply the power spectral density of the baseband signal times ¼ of the square of the maximum carrier amplitude. There is no spectral line introduced at the carrier frequency. Decreasing the modulation index decreases the magnitude of the sidebands, yet introduces and increases the weight of the delta function at the carrier frequency.

Using a modulation index of less than 1 is a technique commonly used in RFID systems, particularly for those operating in frequencies with strict bandwidth limitations. A negative consequence of this, however, is a reduction in the difference between the high and low levels and an increased probability of error.

3.5.1.2 Envelope Detection

Envelope detection is a type of non-linear detection often used to detect amplitude-modulated signals. Essentially the positive envelope of the real portion of the incoming bandpass signal is captured and eventually processed. Figure 3.11 shows a simplified model of an envelope detector.

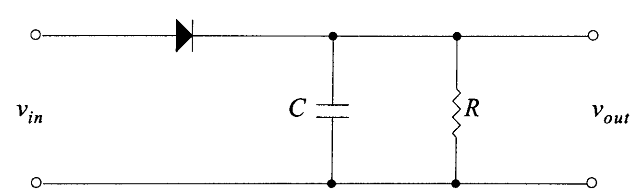


Figure 3.11: Envelope detector.

The resistor and capacitor form a low-pass filter where the time constant is set so that the output signal follows the signal envelope. The upper bounds of the low-pass filter should be much less than the carrier frequency, but much larger than the bandwidth of the incoming modulated signal [24].

$$B \ll \frac{1}{2\pi RC} \ll f_c \tag{3.57}$$

The envelope detector output is essentially the envelope of the incoming signal multiplied by a proportionality constant *K*. For the envelope of the ASK modulated signal from Eq. (3.53), the output voltage is given as

$$v_{out}(t) = K\left|g(t)\right| = KuA_c s(t) + KA_c(1-u) \tag{3.58}$$

One should note that a DC value exists for modulation indices of less than 1.

3.5.2 Tag to Reader Modulation and Coding

In the tag to reader link, conditions and requirements are somewhat different from the reader to tag link. The tag communicates through either load modulation or backscatter modulation. Depending on the implementation, amplitude and phase or simply amplitude are varied. As opposed to the reader to tag link, coding schemes should be low energy and bandwidth is of much less concern. Because the signal levels are so low, they are not constrained by the stringent regulations at higher signal levels. One consequence of the low signal, however, is difficulty in reception by the reader. The reader signal can far outweigh that of the tag, so at times, detection can be difficult. For this reason, a subcarrier is often employed. A subcarrier effectively allows BPSK or FSK modulation over the amplitude and phase modulation inherent in load or backscatter modulation. I will briefly describe the theory of subcarriers and then further discuss the issues in tag to reader coding.

3.5.2.1 Subcarrier

A subcarrier is created by modulating the load or backscatter modulation at a higher frequency than the data rate. Varying the phase or frequency of the subcarrier modulations allows BPSK or FSK modulation, respectively. Some RFID systems, for example, those operating at 13.56 MHz, use a subcarrier of 212 kHz. This moves the sidebands associated with the tag's modulations 212 kHz above and below the carrier frequency of 13.56 MHz. Subcarriers are useful for isolating the tag's sidebands from noise due to the readers signal and that inherent in the reader's receiver and amplifiers. They however require that the tag have a faster clock or an oscillator.

3.5.2.2 Coding

In tag to reader coding, unlike reader to tag coding, energy in the signal should be minimized. Less energy in the signal means more energy available to power the tag. The amplitude however should be high enough so that the reader can detect the signal. Often Manchester, FM0 or other line coding schemes are used. One problem with Manchester is that due to its effective phase variation, it can confuse certain receivers as the 180 degree phase shifts can be ambiguous. Also, coding schemes with mid-bit transitions can either require a faster clock or a slower data rate – both negative effects.

3.5.3 Spread Spectrum

Beyond bandwidth and energy contained in a signal, there are other concerns in communications such as multiple-access, anti-jamming, and interference rejection [24]. Spread spectrum (SS) modulation, by spreading signals over a bandwidth much larger than that require by the message signal, can allow multiple-access, anti-jamming, and interference rejection. RFID systems often use spread spectrum modulation techniques at UHF and microwave frequencies. At these frequencies, regulations typically allow greater output powers for spread spectrum systems relative to narrow-band systems. Power is spread over a wide bandwidth, reducing the chance of harmful interference to and from other devices.

The two most common types of spread spectrum modulation are direct sequence (DS), and frequency hopping (FH). In DS modulation, a pseudo-noise coded signal with period less than the message bit period is mixed with the message signal, creating an output signal with wide bandwidth. In order to retrieve the message, a receiver must de-spread the code with the same pseudo-noise sequence used to transmit the signal. In frequency hopping systems, multiple narrow-band channels at frequencies chosen from a pseudorandom list, are hopped between. This has the effect of averaging the power over the entire width of the hopping frequencies. Typically receivers in frequency hopping systems must hop with the same pseudorandom list of frequencies in order to receive the modulated signal.

In low-cost passive RFID systems, spread spectrum is often used differently than originally intended. In order for a tag to achieve true DS or FH modulation it would require excessive complexity of the tag receiver. Instead, a common strategy is to design the tag's antenna to have a bandwidth covering the entire width of the spread spectrum channel. In this way, the tag may receive all power, whether it be confined in a narrow frequency hopping channel, or spread out over the entire width through direct sequence. To achieve communications, the signals, whether they be frequency hopped or direct sequenced, may still be amplitude modulated by the reader, and subsequently backscattered by the tag. This strategy is useful in RFID as regulations typically allow higher output powers for spread spectrum systems and narrow-band systems.

3.6 Probability of Error

In general, there are two important factors important in the evaluation of a communication system – bandwidth, and performance in the presence of noise [24]. We have already considered bandwidth, now we will consider performance in the presence of noise. A figure of merit for this performance in analog systems is the signal power to noise power ratio. In digital systems it is the probability of bit error, or bit error rate BER.

The BER is the probability of error arising when the system must make a decision between, in binary systems, two signals. In ASK modulated signals, levels will be either high or low. In OOK signals, the low will be zero. The receiver has a decision threshold above which signals will be considered high, and below which they will be considered low. The total probability of error or BER, is the sum of the probability of error associated with each signal.

When detection is coherent, more information is used – not only the amplitude information, but also the phase information. When detection is non-coherent, however, only the amplitude information is used. Error performance thus decreases with non-coherent detection. RFID systems typically use non-coherent envelope detection because it can be far simpler and more inexpensive to implement.

With OOK or ASK modulated signals, there will either be two signals levels. How these levels represent a particular bit, depends on the coding scheme used. Regardless of the coding scheme however, if errors are made in detecting these levels, there will be an error in the bit. Certain coding schemes may inherently be able to detect such an error, while others may not. If not, other error detection or error correction techniques may be employed. Noise in the wireless channel can take many forms. Two common forms are burst-noise and gaussian noise. Gaussian noise is typically that resulting from background radiation, thermal noise, and shot noise. In short-range device communications, particularly in low-cost passive RFID communications, burst noise resulting from other interfering transmissions, in particular, is most prevalent. I will consider burst noise for this reason, and gaussian noise to illustrate a calculation of the BER.

3.6.1 Burst Noise

Burst noise can be caused by interference from other sources or affects such as multipath. It can be particularly harmful to RFID transmissions where signals can be low and detectors simple. Gilbert [30] models errors created by burst noise as a Markov chain where states can take one of two states – a good state where errors occur with low probability or a bad state where errors occur with a high probability. Elliott has further generalized this model [31], and others have applied it [32][33]. Burst errors can occor on single bits, or large blocks of bits.

3.6.2 Gaussian Noise

In considering the affect of additive gaussian noise, we will assume two general cases of signal levels, those that have a nonzero amplitude and those that have zero amplitude. In the presence of noise, probability density functions will vary with these two levels. More generally, they will vary with any variation in level. The simpler case is when the signal level is zero and noise is the only received component. In OOK, this will occur. In ASK, however, with a modulation index less than 1, it will not occur. I will describe the differences seen in probability density functions for these variations in signal levels.

First we consider the case where the transmitted signal amplitude is zero. Thus, the input c to an envelope detector, is simply equivalent to the noise n. If we assume noise is gaussian, we can decompose it into an in-phase and quadrature form, represented by

$$c(t) = n(t) = x(t)\cos(\omega_c t) - y(t)\sin(\omega_c t) \tag{3.59}$$

where both x and y are independent gaussian random variables. When the signal is zero, we simply must transform these variables from Cartesian form to Polar form. This will give us an magnitude component r and a phase component θ. Because we assume envelope detection, we will only use the amplitude component to find the probability density function associated with the signal.

Transforming the in-phase/quadrature noise to polar form and finding the probability density function of r and θ, and then integrating over all θ, gives the probability density function of the amplitude r alone. The distribution function found is called the Rayleigh distribution.

Now, we can consider the case where the transmitted signal has a nonzero amplitude, whether it be the full amplitude, or the lower level when the modulation index is less than 1. This introduces an additional value to the in-phase component of the received signal. This additional component can be given by v from Eq. (3.52) Now the detector input c becomes

$$c(t) = v(t) + n(t) = v(t) = \left[A_c\left[1 + u\left(s(t) - 1\right)\right] + x(t)\right]\cos(\omega_c t) - y(t)\sin(\omega_c t) \tag{3.60}$$

This complicates the situation somewhat. Just as for the simple case with a zero signal amplitude, we transform the cartesisan inphase/quadrature components to polar form. We find the probability density function of r and θ together, however in this case they are no longer independent. We find that integrating over theta presents us with an integral that can only be solved through a modified Bessel function of the first kind and zero order [34].

The resulting distribution function is called a Rician distribution. We should note that setting the signal amplitude to zero, the probability density function becomes a Rayleigh distribution; setting the amplitude much greater than the noise in phase component, produces a Gaussian distribution. At points in between, it is Rician distribution.

Given the probability density functions of the two signal levels, we can find the probability of error. As mentioned, it is simply the sum of the areas under the individual probability density functions corresponding to errors at those two levels. This is represented as

$$P_e = P_1\int_{-\infty}^{b} f(r \mid v_1)dr + P_2\int_{b}^{\infty} f(r \mid v_2)dr \tag{3.61}$$

where P_1 and P_2 are the probabilities of signal 1 and 2 occurring, respectively. The quantity b is the threshold level. In OOK modulation, the first probability density function is simply the Rayleigh distribution, and the second, the Rician distribution function. In the case of ASK with a modulation index of less than 1, both density functions will be Rician distribution functions. Unfortunately, for Rician distributions, we find that the integration of the Bessel function is not possible in closed form. However it can be solved numerically. In general, we find that the less the difference in signal levels, the greater the error. In addition, the lower these levels with respect to RMS noise, the greater the BER. The threshold level must also be wisely chosen so as to minimize the probability of error. This of course requires prior knowledge of the signal amplitude relative to noise, and is certain to be variable in RFID systems. Due to simple detectors, low signal levels, and likely relatively high BERs, it can be useful to incorporate some form of error detection or correction in the transmissions both from reader to tag and tag to reader. I will review such methods next.

3.7 Error Detection and Correction

There is an interesting tradeoff in RFID systems when considering the reader to tag and tag to reader links. The reader-transmitted signal is relatively strong, yet the tag is limited to very simple detectors and signal processing. This leaves room for increased error. In the tag to reader link, the tag-transmitted signal is relatively low. Even though the RFID reader may have far more capable and robust receivers and signal processing, due to the low signal level, errors may still be highly probable. Error detection or correction of some extent should be considered in both links.

Though there are exceptions, error detection and correction coding schemes require the addition of bits to the original message. Depending on the number of bits and the method in which they are added, these codes may be able to correct or just simply detect errors. In general, error correction codes require more bits, and thus, though robust, are more inefficient and require more bandwidth and processing. In RFID systems, transmissions are relatively short and processing requirements are relatively stringent. For the most part, error detection codes are more appropriate.

Common types of error detection include parity checking, longitudinal redundancy checking (LRC), and cyclic redundancy checking (CRC) [6]. In parity checking, an extra bit is adding to a string of bits, which indicates if there are an even or an odd number of binary 1 bits depending on implementation. Such checking is very simple, yet useless if there is an even number of errors within the string. Longitudinal redundancy checks require recursive exclusive OR (XOR) operations on each byte within a block of data. Results of the operation are appended to the string and transmitted. On reception, the same procedure is performed. If the result of a check is not zero, an error has occurred. Similar to parity checking, however, multiple errors can cancel each other out. Cyclic redundancy checks are more robust, yet slightly more demanding on processing. The CRC operation provides a way of nearly uniquely identifying a string of bits. The larger the CRC, the more data can be reliably checked. Because transmissions are relatively short in RFID systems, 8-bit CRCs are commonly used.

3.8 Summary

In this Chapter we have studied the fundamentals of communications as performed by low-cost passive RFID systems. We should note that there is flexibility in coding and modulation both from the reader to the tag and the tag to the reader. We will see in subsequent chapters that coding and modulation influences more than just bandwidth and spectral density, and is influenced by more than just limitations on bandwidth and spectral density. In the next chapter we will see how regulations constrain bandwidth, and in subsequent chapters we will see how this affects the overall performance of the tag.

Chapter 4

REGULATIONS

4.1 Introduction

We have discussed the electromagnetics and communications principles that govern the operation of RFID systems. In particular, we have seen that operating frequency, power and bandwidth affect the more important performance specifications – namely range, data rate and communications integrity. In this chapter, I will describe the regulations that govern operating frequency, power, and bandwidth and how they affect worldwide compatibility.

4.2 Necessity of Common Allocations

Of the large variety of RFID applications, object tracking in global supply chains represents one of the largest potential markets. It is thus paramount that tags be able to operate in and between different countries. For this to be possible, tags must be compatible with foreign readers and both readers and tags must be compatible with foreign radio-frequency regulations. This compatibility can be accomplished by 1) tags that function, without modification, at multiple frequencies, and 2) by allocation of common frequency bands worldwide.

Design of tags that function, without modification, over extremely wide bands (i.e. from 125 kHz to 13.56 MHz), would require excessive complexity and cost. Generally, certain antennas can operate with relatively uniform performance within certain bands (i.e. 902-928 MHz in the US) and between closely neighboring bands (868 MHz in Europe and 915 MHz in US). While this is possible in certain bands, for possible operation at a greater range of frequencies, it is necessary to locate frequency bands common worldwide.

4.3 Worldwide Structure of Regulatory Organizations

There is a well-established worldwide structure of regulatory organizations consisting of international, regional, and national organizations. I will briefly describe these various organizations.

4.3.1 ITU

The International Telecommunication Union (ITU), of which virtually all independent nations are members [35], is involved with worldwide radio frequency spectrum management and publishes a table of international frequency allocations (S5)[36]. Of its many functions and purposes, it seeks to maintain cooperation between its members and coordinate efforts and allocate bands for the elimination of interference between various services that use the radio-frequency spectrum [ITU]. Its Radiocommunication sector (ITU-R) is responsible for dealing with all telecommunications using radio waves. Though the ITU seeks uniform frequency allocations worldwide, for situations where a country might require band allocations that differ from the rest of the world, exceptions are made and noted in the frequency allocation table. For the purposes of designating region specific allocations, the ITU uses its division of the world into three regions (S5.2)[36].

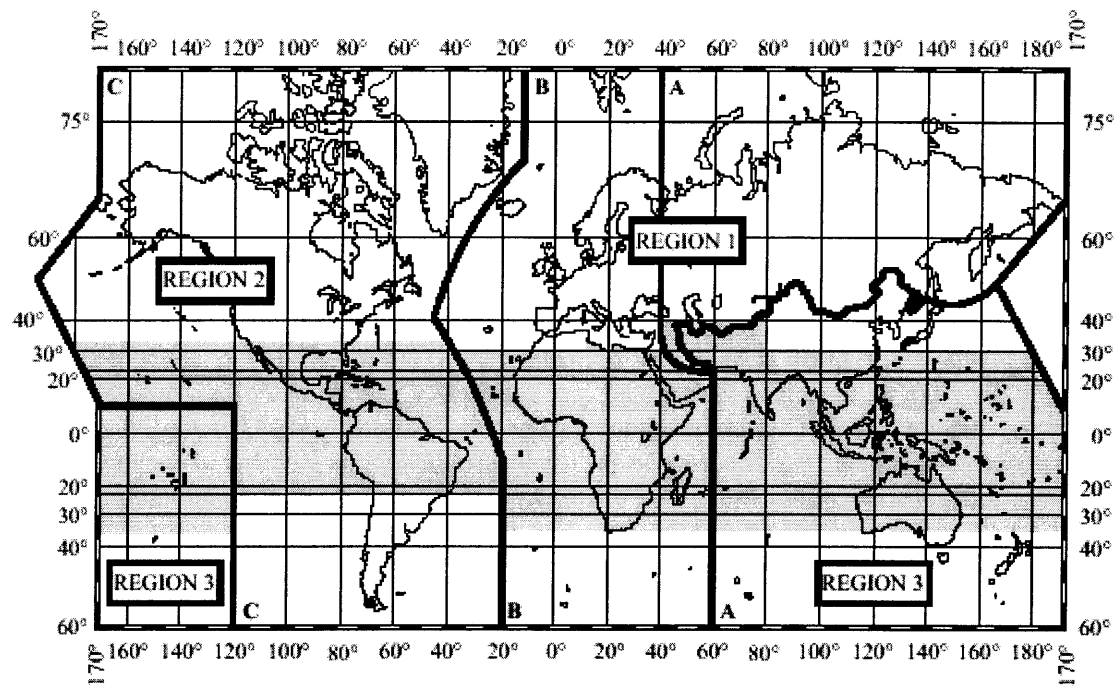


Figure 4.4.1: ITU regions (S5.2)[36]

4.3.2 Regional/National

While the ITU fosters cooperation between and makes recommendations to regional and national bodies, it is the various regional and national bodies that set and administer the actual band allocations and emission limits. As both the world's largest economies, and largest potential markets for RFID exist in Europe, the United States, and Japan, I will focus on the standards of these countries' respective telecommunications standards organizations.

The European Conference of Postal and Telecommunications Administrations (CEPT) deals with sovereign and regulatory matters in the areas of post and telecommunications. Members include 43 European counties including those in Central and Eastern Europe. It consists of three committees, one on postal matters, the CERP (Comitè europèen des règulateurs postaux) and two on telecommunications issues: the ERC (European Radiocommuication Committee) and the ECTRA (European Committee for Regulatory Telecommunications Affairs). These committees work on harmonization issues and adopt recommendations and decisions, normally prepared by their working groups and project teams. The ERC has established the European Radiocommunications Office (ERO), which supports the committees' activities and conducts studies for it [37].

The European Telecommunications Standards Institute (ETSI), established in 1988 by CEPT (along with CEN and CENELEC), is recognized as one of the three European Standards Organizations (ESO). It is responsible for developing telecommunications related standards in Europe. Thirty-seven National

Standards Organizations of thirty-four countries as far east as Russia and Turkey, provide elaboration, approval, and implementation of the standards [38].

The organizations responsible for telecommunications standards and regulations in the United States and Japan, are the Federal Communications Commission (FCC) [39], and the Ministry of Posts and Telecommunications (MPT) [40], respectively. In Japan, the Association of Radio Industries and Businesses (ARIB) has been designated as the Center for Promotion of Efficient Use of the Radio Spectrum. Among its many tasks, ARIB develops detailed standards that fit within the less detailed requirements of the MPT regulations [41].

4.4 Frequencies Available for RFID Use

CEPT has adopted the term "Short Range Device" (SRD) to describe "radio transmitters which provide either unidirectional or bi-directional communication and which have low capability of causing interference to other radio equipment." Antennas may be integral, dedicated, or external, and all types of modulation may be used [42]. The ITU, in their Recommendation 213/1 has supported this term and further recommends to national administrations that they adopt the recommended parameters and that use of SRDs should not be unnecessarily restricted. Efforts to harmonize frequency allocations worldwide are ongoing.

RFID systems are considered to be short-range devices, and they may operate in certain frequency bands allocated in most countries. I will describe these bands now.

4.4.1 ISM

In their frequency allocation tables, the ITU designates certain bands as industrial, scientific, and medical (ISM) bands. These bands are for industrial, scientific, and medical applications of radio frequency emitting devices. Typically radiation from these devices, though often either intended, or not intended for limited range applications, can interfere with other radio frequency communications services. For this reason, it is necessary to allocate specific bands for their use [43]. The principal ISM bands (RR S5.150)[36] are,

13.553-13.567 MHz
26.957-27.283 MHz
40.66-40.70 MHz
902-928 MHz (Region 2)
2.4 – 2.5 GHz
5.725-5.785 GHz
24-24.25 GHz

Certain bands are available for ISM use only on approval of administrations that might be affected by their use (RR S5.138)[36],

6.765 – 6.795 MHz
61.0 – 61.5 GHz
122-123 GHz
244-246 GHz

The 433.05-434.79 MHz band is also available for ISM use, but only in ten European countries (S5.280)[36]. It is available for use in other countries in Region 1 subject to the terms of RR S5.138. The ISM bands are also allocated for other radio services, so such services must tolerate ISM device generated radiation. The ITU recommends that all countries (administrations) seek to ensure that ISM device radiation is minimal (S15.13)[36], yet there are currently no constraints on the radiation levels. Discussions between the ITU and the International Special Committee on Radio Interference of the International Electrotechnical Commmission (IEC) regarding limiting interference to radio services were initiated with RR Resolution 63.

4.4.2 Other (LF/UHF)

In addition to the ITU ISM bands, the low frequency (LF) band from 9kHz – 135 kHz is available for unlicensed use in most regions.

As recommended in CEPT Recommendation 70-03, referenced by the applicable ETSI standards, additional UHF bands are available for use in most of the CEPT countries by non-specific short range devices [44]. These bands are within 868 MHz to 870 MHz. Efforts are under way to add an allocation from 865 MHz to 868 MHz.

Table 4.1 Summary of available LF, UHF, and ISM bands worldwide.

Frequency	Bandwidth	Near/Far	$\lambda/2$	Type	Region	Restrictions
9 – 135 kHz		382 m	1200 m			Sub-divided
6.78 MHz	± 15 kHz	7.04 m	22.1m	ISM	All*	
13.56 MHz	± 7 kHz	3.52 m	11.1m	ISM	All	
27.12 MHz	± 163 kHz	1.76 m	5.53m	ISM	All	
40.68 MHz	± 20 kHz	1.17 m	3.69 m	ISM	All	
433.92 MHz	± 870 kHz	11.0 cm	34.6 cm	ISM	1**	
869.525 MHz	± 125 kHz***	5.5 cm	17.3 cm	NS - SRD	CEPT	865-868 MHz proposed
915 MHz	± 13 MHz***	5.2 cm	16.4 cm	ISM	2	250 kHz/500 kHz FH
2.45 GHz	± 50 MHz***	1.9 cm	6.12 cm	ISM	All	FH in Europe and US
5.8 GHz	± 75 MHz***	8.2 mm	2.59 cm	ISM	All	
24.125 GHz	± 125 MHz	2.0 mm	6.2 mm	ISM	All	

* On approval of administrations that might be affected by its use, (RR S5.138)[36]

** Only in ten European countries (RR S5.280)[36]. Available for use in other countries in Region 1 subject to the terms of RR S5.138 [36].

*** May allow frequency hopping and direct sequence spread spectrum.

4.5 Radiation, Bandwidth, and Detection Specifications

Before describing the radiation level and bandwidth limitations for the various frequencies, I will briefly review the units used for their specification, and compare the methods used to describe them.

Radio regulations as specified by the various administrations vary on three levels: power and field strength limits, bandwidth limits, and detection methods. Power and field strength may be specified in terms of near-field or far-field electric or magnetic fields, or radiated power. Bandwidth may be 6 dB, 20 dB, 30 dB, or more. Detection method may either be through a quasi-peak detector or an average field level detector. To understand how the regulations are related, we must convert them to common units. I will derive the conversion relations.

4.5.1 *Field Strength and Power Conversion*

In order to compare the field strength and power levels limits set by the different administrations, we must convert them to common units. For most frequencies, the FCC specifies emission limits in terms of radiated power or the far-field electric field. CEPT/ETSI, on the other hand, specifies the limits in terms of the near-field magnetic field, the far-field magnetic field, or the radiated power. The MPT specifies limits in terms of either the far-field electric field or radiated power.

4.5.1.1 Far-field

As we have seen in Chapter 2, the electric and magnetic field in the far-field are simply related by the free space impedance η_0. Given the electric field E_{max} at some distance R_E, we can find the magnetic field H_{max} at some distance R_H. In free space, the conversion can be done as follows,

$$E_{\max}^{ff} = \frac{Id\ell}{4\pi R_E}\eta_0 k \tag{4.1}$$

$$H_{\max}^{ff} = \frac{Id\ell}{4\pi R_H} k \tag{4.2}$$

$$H_{\max}^{ff} = E_{\max}^{ff}\frac{R_E}{\eta_0 R_H} \tag{4.3}$$

Given the electric or magnetic field at some distance in the far-field, we can convert to the radiated power. First we find the average power density at some distance R_{EH},

$$S = \frac{E_{ff}^2}{2\eta_0} = \frac{H_{ff}^2\eta_0}{2} = \frac{P_t G_t}{4\pi R_{EH}^2} \tag{4.4}$$

EIRP, which is simply the antenna input power P_t multiplied by the antenna gain relative to isotropic G_t, can be expressed as

$$EIRP = \frac{E_{ff}^2}{2\eta_0}4\pi R_{EH}^2 \tag{4.5}$$

Through Eq. (2.48) we can convert to ERP. One should note that if E_{ff} and H_{ff} in Eq. (4.4) and Eq. (4.5) are RMS values, then the factor of 2 should be dropped.

4.5.1.2 Near-field

In the near-field, however, the relationship between the electric and magnetic field is more complicated. We must consider the structure of the antenna in order to make the conversion. Because near-field systems are either capacitive or inductive, we must consider two different types of antennas. I will only consider inductive systems here. For inductive systems, we can approximate the antenna by a small loop antenna as long as the distance from the antenna is much greater than the loop radius (Eq. (2.15 - 2.20)). Given the far-field electric field E^{ff}_{max} at some distance R_E (Eq. (4.1)), we wish to find the near-field magnetic field H_{max} at some distance R_H. From Eq. (2.18),

$$H^{nf}_{\max} = \frac{Id\ell}{4\pi k {R_H}^3} 2 \tag{4.6}$$

Combining Eq. (4.1) and Eq. (4.6) we find an approximate relationship between the peak far-field electric field and the peak near-field magnetic field,

$$H^{nf}_{\max} = E^{ff}_{\max} \frac{2R_E}{\eta_0 k^2 R_H^3} \tag{4.7}$$

4.5.2 Bandwidth

Unless otherwise specified, the bandwidth is generally taken to mean the range of frequencies for which the spectral components are above a certain level. Administrations typically specify general base radiation levels, under which spectral components may exist. Certain bands may be restricted and allow only spurious emissions.

In some cases, bandwidth is expressed specifically as n-dB bandwidth, however, in most cases this must be determined from comparing the fundamental field strength or power limit with the limits for neighboring spurious emission or sidebands. The n-dB bandwidth is simply the ratio, converted to dB, of the fundamental or center frequency limit to the spurious emission or sideband limit.

Spurious emissions are typically defined as those not including the carrier and modulation sidebands. However, they are also often used as the limits for the out of band modulation sidebands [45].

4.5.3 Detection

In order to understand how regulations determine performance of RFID systems, it is imperative that one understand how the measurements are performed. Most administrations test equipment in either an open field test site or an anechoic chamber, depending on the frequency and device. Devices are tested in most orientations and varying heights in order to determine the maximum field strength or power density at a given distance (FCC 15.31)[46]. Typically, devices are tested in normal operating modes. What is of particular interest, are the types of detectors used in the measurements. I will briefly discuss these here.

In general, signals are either continuous or pulsed. Continuous signals, such as those produced by an oscillator or clock signal, have a narrow spectrum. Pulsed signals, such as momentary spikes at some pulse-repetition frequency (PRF), have a broadband spectrum. Though peak magnitudes of these signals might be equal, the continuous signal is generally considered to produce more harmful interference. Using an instantaneous peak detector to measure the pulsed signal will inaccurately reflect its degree of potential interference.

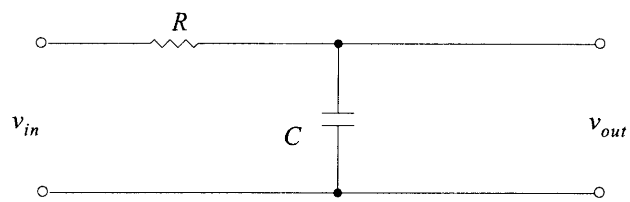


Figure 4.2: Average detector.

One solution to this problem is to average the signal over time. This is done by passing the signal through an RC filter (Figure 4.3) [47]. Average detectors are specified for various frequency bands between 9kHz and 1 GHz, and typically above 1 GHz [46][45]. The problem with the average detector though, is that for pulses with very low duty cycles, the average would be excessively low, and likely underestimate the degree of potential interference.

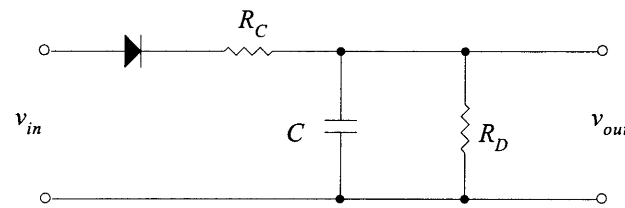


Figure 4.3: Quasi-peak detector.

For this reason, the quasi-peak detector is often used. The quasi-peak detector essentially combines a peak detector with an average detector. It consists of a charge and discharge circuit set by two resistors and a capacitor (Figure 4.4). Such a detector adjusts for the peak magnitude, the energy of the pulse, and its repetition rate. The quasi-peak detector is often specified for use in frequencies below 1 GHz.

Specifications for the design and use of these detectors are given in the IEC International Special Committee on Radio Interference (CISPR) Publication 16 [48]. The specifications, which include time constants and bandwidths, vary with frequency band of operation. There are four bands: 9 – 150 kHz (A), 150 kHz – 30 MHz (B), 30 MHz – 300 MHz (C) and 300 MHz – 1 GHz (D). Above 1 GHz spectrum analyzers with peak detectors are often used.

For the quasi-peak detector, the 6 dB bandwidth for band A is 220 Hz, for band B, 9 kHz, and for band C and D, 120 kHz. If the pulse repetition frequency exceeds these bandwidths, the detector acts as an average detector, and no correction factor is necessary. Those systems occupying frequency bands with large bandwidths will have pulse repetition frequencies greater than these bandwidths. However, systems operating within frequency bands with relatively narrow bandwidths, such as below 135 kHz, 6.78 MHz, and 13.56 MHz, may have pulse repetition frequencies less than 9 kHz or 220 Hz. In these cases, we can adjust for the differences through a correction factor based on pulse response curves and relations given in CISPR-16 (Appendix E.3) [48]. Correction factors can be significant (> 10 dB), but as the actual factors vary with the bandwidth of the average detector, I will not make any conversions here.

4.6 Frequencies and Limits

Because RFID systems operating at the lower frequencies rely on manipulation of the near-field electric or magnetic field, I will consider these fields as opposed to the far-field radiated power. When analyzing far-field systems on the other hand, I will consider the far-field radiated power.

In Table 4.1, the near-field limit and the half-wavelength (resonant antenna) dimension are given for each frequency. We can see a suitable grouping for near-field and far-field systems. Those systems from 125 kHz to 40.6 MHz should be considered near-field systems, as the near-field limit is suitably large, while the half-wavelength dimension is unsuitably large. The systems operating from 483 MHz through 24 GHz should operate in the far-field, as the half-wavelength dimension drops to suitable size, while the near-field becomes unsuitably small.

The various regulatory agencies typically specify general limits for spurious emissions and other limits for specific frequency bands. I will describe both the limits for the specific frequency bands of interest and the general limits in the vicinity of these bands.

One should note that in certain configurations, such as tunnel systems, fields may be shielded and thus higher radiation levels than regulations allow may be "seen" inside. When considering RFID systems, we will consider only unshielded systems.

Regulations that will frequently be referred to include European standards ETSI EN300330 (Inductive loops from 9 kHz to 25 MHz) [45], EN 300220 (SRDs from 25 MHz to 1000 MHz) [49], EN300440 (SRDs from 1 GHz to 40 GHz) [50], and CEPT Rec. 70-03 (SRDs at all radio frequencies) [44], the United States regulations, FCC Part 15 [46], and Japan's MPT's Radio Law.

4.6.1 Near-field Operation Frequencies

Each administration defines the field limits for these frequencies differently. The FCC specifies the electric field at some distance in the far-field. ETSI/CEPT specifies the magnetic field at 10 m which lies either in the near-field or the far-field depending on the frequency. The MPT specifies the electric field at 3 m down to approximately 16 MHz, where the electric field is then specified at the near to far-field transition point for a given frequency. This will be discussed further in the following section.

4.6.1.1 9 kHz - 135 kHz

Though not an ISM band, frequencies in this band are commonly used for short-range radio-communication applications and ISM devices. In addition, several long range services including radio-navigation, time signal, and military radio services use this band [ITU-R SM 1056/ RR]. Field strength limits here are generally higher than those at other bands.

CEPT has divided this band into five smaller bands, largely to protect the other radio-communication services from harmful interference. They are allocated for inductive applications (70-03 Annex 9). These bands and their associated power levels are given in Table 4.2 [44][45].

Table 4.2: Frequencies available for inductive SRDs under CEPT/ETSI regulations from 9 – 135 kHz.

Frequency Band	Magnetic Field Strength	Detector
9 – 59.75 kHz	72 dBμA/m @ 10 m*	Quasi-peak
59.75 - 60.250 kHz	42 dBμA/m @ 10 m**	Quasi-peak
60.25 – 70.0 kHz	72 dBμA/m @ 10 m* **	Quasi-peak
70 – 119 kHz	42 dBμA/m @ 10 m	Quasi-peak
119 – 135 kHz	72 dBμA/m @ 10 m* **	Quasi-peak
135 kHz – 1 MHz (General)	37.7 dBμA/m @ 10 m ***	Quasi-peak

* At 30 kHz descending 3.5 dB/octave (equivalent to 72 – $3.5\log_2$(f [kHz]/30) dBμA/m @ 10 m)
** For loop coil antennas with an area of < 0.05 m^2, the limit is 10 dB below that given. For loop coil antennas with an area between 0.05 m^2 and 0.16 m^2, the limit is the given value + 10log(Area/0.16 m^2)
*** At 135 kHz descending 3 dB/octave (equivalent to 37.7 - $3\log_2$(f [kHz]/135) dBμA/m @ 10 m)

The FCC gives general emission limits in section 15.209. For the frequencies between 9 kHZ and 490 kHz, the electric field strength may be 2400/f(kHz) μV/m at 300 m, where f is the frequency. The band 90 kHz to 110 kHz is a restricted band limited only to spurious emissions. At 160 kHz – 190kHz there is a separate allocation. Thus, the available band for RFID includes 9 – 90 kHz, and 110 – 160 kHz. In the 9-90 kHz and 110-490-kHz bands, measurements are made using an average detector [46].

Table 4.3: Frequencies available for general use under FCC regulations from 9 – 160 kHz.

Frequency Band	Electric Field Strength	~Magnetic Field Strength	Detector
9 – 90 kHz	2400/f [kHz] μV/m @ 300 m	198.8 - 60Log(f [kHz]) dBμA/m @ 10 m	Average
110 – 160 kHz	2400/f [kHz] μV/m @ 300 m	198.8 - 60Log(f [kHz]) dBμA/m @ 10 m	Average

In Japan, regulations are given in the MPT's Radio Law. According to paragraphs 1 and 3 of Article 4, extremely low power radio stations may be used without a license. Article 6 of the Regulations for the Enforcement of the Radio Law, specifies the limits [42]. For frequencies below 322 MHz, the electric field at 3 m should be below 500 μV/m. At approximately 16 MHz, 3 m is the near-field to far-field transition point. Below 16 MHz, 3 m lies within the near-field. Thus below 16 MHz, the 500 uV/m limit is enforced at the frequency's near-field to far-field transition point. To compute the magnetic field at some distance, in our case 10 m, we can use Eq. (4.3) or (4.7) depending on whether the distance lies in the near-field or far-field. Because 10 m is in the near-field at frequencies less than approximately 4.8 MHz, I have made the conversion using Eq. (4.7) for these frequencies. For frequencies between 4.8 MHz and 15.9 MHz, 10 m lies in the far-field and we can convert using Eq. (4.3). One should note that this is an assumption based on a regulation comparison given in [11] and should be validated with the appropriate MPT and ARIB standards.

Table 4.4: Frequencies available for use by extremely low power radio stations under MPT regulations.

Frequency Band	Electric Field Strength	~Magnetic Field Strength
9 kHz – 4.8 MHz	500 μV/m at ($c/2\pi f$) m	229.2 - 60Log(f[kHz]) μA/m @ 10 m
4.8 MHz – 15.9 MHz	500 μV/m at ($c/2\pi f$) m	76 – 20Log(f[kHz]) μA/m @ 10 m
15.9 MHz – 322 MHz	500 μV/m at 3 m	-8 dBμA/m @ 10 m

4.6.1.2 6.78 MHz

The band 6.765 – 6.795 MHz is an ISM band available in countries subject to RR S5.138 [36]. Beyond other ISM devices, it is used for broadcasting and aeronautical radio. Regulations in Europe are favorable for RFID systems. The United States, however, is very stringent.

For inductive applications, limits are given in CEPT 70-03 Annex 9 [44]. The magnetic field strength limit is 42 dBμA/m at 10 m. Up to +/- 150 kHz from the center frequency, 6.78 MHz, the magnetic field strength limit is 9 dBμA/m at 10 m. Beyond, the magnetic field strength limit falls to –1 dBμA/m at 10 m. Measurements are made with a quasi-peak detector.

Table 4.5: Bandwidth and field strength limits at 6.78 MHz under CEPT/ETSI regulations.

Frequency Band	Magnetic Field Strength	Detector
6.78 MHz +/- 30 kHz	42 dBμA/m @ 10m	Quasi-peak
6.78 MHz +/- 30 – 150 kHz	9 dBμA/m @ 10m	Quasi-peak
6.78 MHz +/- 150 kHz –	–1 dBμA/m @ 10m	Quasi-peak

According to FCC 15.223, for the band 1.705 – 10 MHz, the FCC dictates that if the 6 dB bandwidth of the emission is greater than or equal to 10% of the center frequency, the electric field limit is 100 μV/m at 30m. If the 6 dB bandwidth, however, is less than 10%, the electric field limit at 30 m is the greater of 15 μV/m, or the bandwidth of the device in kHz divided by the center frequency of the device in MHz μV/m at 30 m. If the system confined itself to operation within the ISM band, the 6 dB bandwidth would most certainly be less than the 30 kHz allocated for the band. The electric field limit would be 15 μV/m at 30 m [46]. However, this may not be necessary.

Table 4.6: Bandwidth and field strength limits at 6.78 MHz under FCC regulations.

Frequency Band	Electric Field Strength	~Magnetic Field Strength	Detector
6.78 MHz +/- 30 kHz, or	15 μV/m @ 30 m	-18.5 dBμA/m @ 10m	Average
6.78 MHz +/- 678 kHz	100 μV/m @ 30 m	-1.98 dBμA/m @ 10m	Average

In Japan, the approximations given in Table 4.4 apply [42].

4.6.1.3 13.56 MHz

The band 13.553 – 13.567 MHz is available as an ISM band worldwide (RR S5.150). In addition to a number of short-range and ISM devices, transcontinental radio services use this band. Regulations are generally favorable for RFID worldwide.

For inductive applications, limits are given in CEPT 70-03 Annex 9 [44]. The magnetic field strength for the band is 42 dBμA/m at 10 m. Up to +/- 150 kHz from the center frequency 13.56 MHz, the magnetic field strength limit is 9 dBμA/m at 10 m. Beyond, the magnetic field strength limit falls to –3.5 dBμA/m at 10 m. Measurements are made with a quasi-peak detector.

Table 4.7: Bandwidth and field strength limits at 13.56 MHz under CEPT/ETSI regulations.

Frequency Band	Magnetic Field Strength	Detector
13.56 MHz +/- 7 kHz	42 dBμA/m @ 10 m	Quasi-peak
13.56 MHz +/- 30 – 150 kHz	9 dBμA/m @ 10 m	Quasi-peak
13.56 MHz +/- 150 kHz →	–3.5 dBμA/m @ 10 m	Quasi-peak

The FCC specifies an electric field limit of 10,000 μV/m at 30 m within the ISM band (15.225). Outside of the band, the emissions must not exceed 30 μV/m at 30 m. There is a restricted radioastronomy band at and below 13.41 MHz where only spurious emissions are allowed. In band measurements are made with a quasi-peak detector while out of band measurements are made with an average detector [46].

Table 4.8: Bandwidth and field strength limits at 13.56 MHz under FCC regulations.

Frequency Band	Electric Field Strength	~Magnetic Field Strength*	Detector
13.56 MHz +/- 7 kHz, or	10,000 μV/m at 30 m	38.0 dBμA/m @ 10 m	Quasi-peak
13.56 MHz +/- 7 kHz →	30 μV/m at 30 m	-12.4 dBμA/m @ 10 m	Average

In Japan, wireless card systems (RFID) operating in this band may occupy a bandwidth of 7 times the modulation rate with an antenna input power of 10 mW and a gain less than or equal to 30 dBi. This is equivalent to an EIRP of 10 W. At 10 m, this corresponds to a magnetic field of 73.2 dBuA/m at 10m. If the antenna gain is 2.15 dBi, then the corresponding magnetic field is 45.4 dBμA/m at 10 m [42] These limits should be verified in the appropriate documents, which were unavailable at the time of publication.

4.6.1.4 27.12 MHz

The band 26.957 – 27.283 MHz is recognized as a worldwide ISM band. It is used by a number of short-range and ISM devices, including welding systems, medical devices, remote controlled models, and pagers. In North America and Europe, it lies within a Citizens Band (CB) radio band. It is not frequently used by RFID systems, though bandwidth is wide.

CEPT Rec. 70-03 allows a magnetic field strength of up to 42 dBμA/m at 10 m within the ISM band, as measured with a quasi-peak detector (Annex 9), or 10 mW ERP output power (Annex 9) [44].

Table 4.9: Bandwidth and field strength limits at 27.12 MHz under CEPT/ETSI regulations.

Frequency Band	Magnetic Field Strength	Detector
27.12 MHz +/- 163 kHz	42 dBμA/m @ 10 m	Quasi-peak
27.12 MHz +/- 163 kHz →	-3.5 dBμA/m @ 10 m	Quasi-peak

The FCC specifies an electric strength of up to 10,000 μV/m at 3 m within the ISM band [15.225]. Outside of the band the electric field limit is 30 uV/m at 30 m [15.209]. Measurements are made with an average detector [46].

Table 4.10: Bandwidth and field strength limits at 27.12 MHz under FCC regulations

Frequency Band	Electric Field Strength	~Magnetic Field Strength	Detector
27.12 MHz +/- 163 kHz	10,000 μV/m at 3 m	18.0 dBμA/m @ 10 m	Average
27.12 MHz +/- 163 kHz →	30 μV/m at 30 m	-12.4 dBμA/m @ 10 m	Average

In Japan, the general limit of 500 uV/m at 3m (-8 dBμA/m at 10 m) applies [42].

4.6.1.5 40.68 MHz

The band 40.66 – 40.7 MHz is also recognized as a worldwide ISM band. It is rarely used by RFID systems, as propagation characteristics and regulations are less favorable than both the 13.56 MHz and 27.12 MHz bands.

CEPT/ETSI specifies a maximum output power of 10 mW ERP within the ISM band. Outside of the band, the limit falls to 250 nW ERP [44][49].

Table 4.11: Bandwidth and field strength limits at 40.7 MHz under CEPT/ETSI regulations.

Frequency Band	Power (ERP)	~Magnetic Field Strength	Detector
40.7 MHz +/- 20 kHz	10 mW	45.4 dBμA/m @ 10 m	Quasi-peak
40.7 MHz +/- 20 kHz →	250 nW	-0.62 dBμA/m @ 10 m	Quasi-peak

FCC limits vary depending on the nature of the device. For a general radiator, the electric field limit is 1000 μV/m at 3 m as measured with a quasi-peak detector. Outside the band, the limit is 100 μV/m at 3 m (15.229), as measured with an average detector [46].

Devices only radiating periodically may transmit up to 2250 μV/m at 3 m at the fundamental. Spurious emissions must not exceed 225 μV/m. Transmission must cease 5 seconds after beginning. Measurements may be made either with a quasi-peak or average detector.

Table 4.12: Bandwidth and field strength limits at 40.7 MHz under FCC regulations.

Frequency Band	Electric Field Strength	~Magnetic Field Strength	Detector
40.7 MHz +/- 20 kHz	1,000 μV/m at 3 m	-1.98 dBμA/m @ 10 m	Average
40.7 MHz +/- 20 kHz →	100 μV/m at 3 m	-22.0 dBμA/m @ 10 m	Average

In Japan, the general limit of 500 μV/m at 3m (-8 dBμA/m at 10 m) applies.

4.6.2 Far-field Operation Frequencies

At UHF and microwave frequencies, limits are usually specified in terms of radiated power (ERP or EIRP). Limits may also be expressed in terms of a transmitter output power and antenna gain. From these quantities, EIRP or ERP can simply be computed.

In the 900 MHz region, 2.45 GHz, and 5.8 GHz bands, spread spectrum operation including frequency hopping and/or direct sequence may be permitted. Typically, administrations will allow both narrow-band and spread spectrum operation within the band. I will describe both these modes.

It should be noted that in the 900 MHz region and the 2.45 GHz bands, a number of changes are being proposed worldwide. I will describe both the current regulations and the proposed changes.

In addition, unless otherwise specified, limits in Japan are those for extremely low power radio stations. For the frequencies between 322 MHz and 10 GHz, the electric field strength limit is 35 μV/m at 3m [42].

Table 4.13: Power limit for extremely low power radio stations at 322 MHz – 10 GHz under MPT regulations.

Frequency Band	Electric Field Strength	Power (EIRP)
322 MHz – 10 GHz	35 μV/m at 3 m	0.37 nW

4.6.2.1 433.92 MHz

The band 433.050 – 434.790 MHz is not a worldwide ISM band. It is allocated in 10 European countries according to (RR S5.280) and in Region 1 according to (RR S.138) [36]. It is heavily used by various short range and ISM devices and lies within an amateur radio band.

In Europe, CEPT/ETSI has allocated this band for non-specific short range devices [44]. Power is limited to 10 mW ERP and devices must have a duty cycle of less than 10%.

Table 4.14: Bandwidth and radiated power limits at 433.92 MHz under CEPT/ETSI regulations.

Frequency Band	Power (ERP)	Power (EIRP)	Duty Cycle	Detector
433.92 MHz +/- 870 kHz	10 mW	16.4 mW	< 10%	Quasi-peak
433.92 MHz +/- 870 kHz →	250 nW	410 nW	< 10%	Quasi-peak

There is neither an ISM allocation in the United States or Japan, though in Japan the general limits for extremely low power radio stations apply.

4.6.2.2 862-870 MHz

The 868-870 MHz band is not an ISM band, but has been allocated by CEPT for non-specific short range devices. Neighboring bands include cordless and cellular phones, and other devices sensitive to interference. The state of this band is currently in flux, and additional proposals have been made to allocate additional bands within the 862-868 MHz region. I will describe both the current and proposed limits.

Currently the 868-870 MHz band is divided into five sub-bands 868 – 868.6, 868.7 – 869.2, 869.3 – 869.4, 869.4-869.65, and 869.7 - 870 MHz. Of these sub-bands, the 869.4-869.65 MHz band allows the highest power level, 500 mW ERP, and thus is most suitable for RFID. Bandwidth is 250 kHz and devices must have a duty cycle of less than 10%. Outside this band, power should not exceed 250 nW ERP [44].

Table 4.15: Current bandwidth and radiated power limits at 869.525 MHz under CEPT/ETSI regulations.

Frequency	Power (ERP)	~Power (EIRP)	Duty Cycle	Detector
869.525 MHz +/- 125 kHz	500 mW	820 mW	< 10%	Quasi-peak
869.525 MHz +/- 125 kHz →	250 nW	410 nW	< 10%	Quasi-peak

Because the current limits are lower than those allowed in the United States, a proposal has been made to open the 865 – 868 MHz band to non-specific SRDs. It is proposed that RFID devices operate within this band either fixed frequency or with frequency hopping in fifteen 200 kHz channels. To reduce interference to neighboring bands, it has been suggested that power levels be lower at the edges of the band. These proposed levels are summarized in Table 4.16.

Table 4.16: Proposed bandwidth and radiated power limits for 865– 868 MHz under CEPT/ETSI regulations.

Frequency	Power (ERP)	~Power (EIRP)	Detector
865-865.6 MHz	100 mW	164 mW	Quasi-peak
865.6-867.6 MHz	2 W	3.28 W	Quasi-peak
867.6-868 MHz	500 mW	820 mW	Quasi-peak
865-868 MHz →	250 nW	410 nW	Quasi-peak

4.6.2.3 915 MHz

The 902 – 928 MHz band is recognized as an ISM band in Region 2. There is neither a similar allocation in Europe nor Japan. However, Canada, Austrailia, New Zealand, and much of Latin America allow operation in or around this band. It is used by a number of devices including motion sensors and cordless phones.

In the United States, there are three modes under which devices may operate within this band. One is limited to field disturbance sensors, another to either frequency-hopping or direct sequence spread spectrum operation, and the third to general narrow-band operation. I will only consider the frequency hopping mode and the general use modes, though the direct sequence mode may be used as discussed in Chapter 3.

The general requirements for frequency hopping systems are given in 15.24 and apply as well to 2.45 GHz and 5.8 GHz. Channels must be separated by a minimum of the greater of 25 kHz or the 20 dB bandwidth of the channel. The hopping frequencies should be chosen from a pseudorandomly ordered list and each frequency should be used equally on average. Receivers must have an input bandwidth matching the hopping channel and shift in accordance with the transmitter. Any 100 kHz outside of the band, the power must be less than 20 dB below the maximum power in band. Spectral components in restricted bands must comply with the general limits given in 15.209. Frequency hopping systems are not required to hop through all frequencies during every transmission, but if transmissions were continuous, they should comply. Employing intelligence to recognize and avoid other users in a channel is permitted, though other coordination for avoiding simultaneous occupancy of a channel is not [46].

In the 902 – 928 MHz band, channels may occupy a 20 dB bandwidth of up to 250 kHz, or between 250 kHz and 500 kHz. Systems using channels with up to 250 kHz bandwidth must use at least 50 hopping frequencies and maintain an average time per frequency of 0.4 seconds in 20 seconds. Systems using channels with between 250 kHz and 500 kHz of bandwidth, must use at least 25 hopping frequencies and maintain an average time per frequency of 0.4 seconds in 10 seconds.

Table 4.17: Bandwidth and power limits for frequency hopping at 902-928 MHz under FCC regulations.

Frequency	Bandwidth (20 dB)	Channels	Power (EIRP)
902 – 928 MHz	< 250 kHz	≥50	4 W
902 – 928 MHz	≥ 250 kHz, <=500 kHz	≥25	1 W

Systems employing at least 50 channels may operate with a peak antenna input power of up to 1 W. Those employing between 25 and 50 channels may operate with a peak antenna input power of up to 0.25 W. If transmitting antennas with gain greater than 6 dBi are used, the peak antenna input power should be reduced by the difference. Essentially, the EIRP will remain constant.

Other devices may also operate in the 902-928 MHz band in accordance with FCC 15.249 without frequency hopping. The fundamental must not exceed 50 mV/m and the harmonics must not exceed 500 µV/m at 3 m. Outside of the band, the emissions should not exceed the 15.209 general limit of 500 µV/m at 3 m.

Table 4.18: Bandwidth and power limits for narrow-band use at 902-928 MHz under FCC regulations.

Frequency	Electric Field Strength	~Power (EIRP)	Detector
902 – 928 MHz (fund)	50 mV/m at 3 m	750 µW	Quasi-peak
902 – 928 MHz (harm)	500 µV/m at 3m	75 nW	Quasi-peak
902 – 928 MHz →	500 µV/m at 3m	75 nW	Average

4.6.2.4 2.45 GHz

This band is recognized worldwide as an ISM band, and thus is used frequently by a number of different devices. Microwaves, other ISM devices, wireless networking, including Bluetooth and Home RF, are major sources of interference [51].

In Europe, parameters are given specifically for RFID systems operating in this band [50]. According to CEPT 70-03, non-specific SRDs operating between 2.4 and 2.4835 GHz may not exceed 10 mW EIRP [44], however, an additional Annex 11 for RFID applications is expected in the fall of 2001. The current draft of ETSI EN 300440-1, gives a spectrum mask specifically for RFID systems operating within this band. In the band 2.446 MHz to 2.454 MHz, EIRP must not exceed 27 dBm (~500 mW). At this level, equipment may be used both indoors and outdoors with frequency hopping or continuous wave. In this same band, for frequency hopping systems operating indoors, the power limit is 36 dBm EIRP (~4 W). It is also specified that at the higher power level, the antenna should not exceed a 90 degree horizontal beamwidth and should have at least 15 dB sidelobe attenuation. Frequency hopping should be performed with at least 20 channels with an average time per channel of 0.4 s [50].

Table 4.19: Bandwidth and power limits for frequency hopping at 2.45 GHz under CEPT/ETSI regulations.

Frequency	Power (EIRP)	Duty Cycle	Detector
2.45 GHz +/- 4 MHz	4 W*	<= 15% in 200 ms	Average
2.45 GHz +/- 4 MHz	500 mW**	<= 100%	Average
2.45 GHz +/- 4 MHz →	1.64 µW		Average

*Only frequency hopping indoors

**Frequency hopping or continuous wave indoors and outdoors

In the United States, RFID systems may operate within this band with or without frequency hopping. For frequency hopping, the general requirements remain the same as specified for the 902-928 MHz band. However, the bandwidth, number of channels, and other details differ. In this band, devices should operate with at least 75 channels, or in special cases, at least 15 channels. When operating with at least 75 channels, the 20 dB bandwidth should not exceed 1 MHz and the average time per frequency should not exceed 0.4 seconds in 30 seconds. The 20 dB bandwidth may be greater than 1 MHz, if the system uses at least 15 non-overlapping channels and their total span is 75 MHz. Under this mode, the average time per frequency should be 0.4 seconds in the time it takes to hop through all channels. If the system uses at least 75 channels, the antenna input power should not exceed 1 W. If using fewer channels, it should not exceed .125 W. If an antenna with gain greater than 6 dBi is used, the input power should be reduced by the amount of the increase in dB. Fixed point-to-point operations are slightly different [46].

Table 4.20: Bandwidth and power limits for frequency hopping at 2.45 GHz under FCC regulations.

Frequency	No. Channels	Bandwidth (20 dB)	Power (EIRP)	Detector
2.4 – 2.4835 GHz	<= 75	1 MHz	4 W	Average
2.4 – 2.4835 GHz	<= 15	> 1 MHz (75 MHz span)	500 mW	Average
2.4 – 2.4835 GHz →			75 nW	Average

Systems may also operate in the band without the frequency hopping requirement according to 15.249. The fundamental should not exceed 50 mV/m at 3 m and the harmonics should not exceed 500 μV/m at 3 m. These measurements are made with an average detector, though it is specified that the peak field strength should not exceed the limits by more than 20 dB.

Table 4.21: Bandwidth and power limits for narrow-band use at 2.45 GHz under FCC regulations.

Frequency	Electric Field Strength	~Power (EIRP)*	Detector
2.4 - 2.4835 GHz (fund)	50 mV/m at 3 m	750 μW	Average
2.4 - 2.4835 GHz (harm)	500 μV/m at 3m	75 nW	Average
2.4 - 2.4835 GHz →	500 μV/m at 3m	75 nW	Average

The standards for radio station systems, RCR STD, administered by the Association of Radio Industries and Businesses (ARIB), specifies limits for RFID systems operating at 2.45 GHz in RCR STD-1 and STD-29. STD-1 gives limits for vehicle identification systems while STD-29 gives limits for specified low power radio stations. STD-29 specifies narrow-band operation with an antenna input power of 10 mW and an antenna gain of 20 dBi. It has been proposed that RCR STD-29 allow frequency hopping as well. Limits may be equivalent to the current limits for Wireless LANs – power of 10 mW and an antenna gain of 2.14 dBi [42].

Table 4.22: Bandwidth and power limits for 2.45 GHz under MPT regulations.

Frequency	Power/Gain	Power (EIRP)
2.4 - 2.4835 GHz (narrow)	10 mW / 20 dBi	1 W EIRP
2.4 - 2.4835 GHz (FHSS)	270 mW / 2.14 dBi	.44 W

4.6.2.5 5.78 GHz

This band is recognized worldwide as an ISM band. It has recently been occupied by wireless LANs; however, it is not nearly as crowded as the 2.45 GHz area band.

In Europe, non-specific SRDs should not exceed 25 mW EIRP. There are no special allocations for RFID. Spurious emissions above 1 MHz should be less than 1 μW [50].

Table 4.23: Bandwidth and power limits for non-specific SRDs at 5.78 GHz under CEPT/ETSI regulations.

Frequency	Power (EIRP)	Detector
5.725-5.875 GHz	25mW	Average
5.725-5.875 GHz →	1 μW (or ERP)	Average

In the United States, as in the 915 MHz and 2.45 GHz bands, devices may operate using spread spectrum or general narrow-band modes. The general frequency hopping requirements mentioned for 915 MHz apply here, as well. Frequency hopping systems in this band should operate with at least 75 hopping channels, a 20 dB bandwidth not exceeding 1 MHz, and an average time per frequency of 0.4 seconds in 30 seconds. The antenna input power should not exceed 1 W, and if the antenna gain exceeds 6 dBi, the input power should be reduced by the corresponding amount in dB. Fixed point to point operations are different [46].

Table 4.24: Bandwidth and power limits for frequency-hopping at 5.78 GHz under CEPT/ETSI regulations.

Frequency	No. Channels	Bandwidth (20 dB)	Power (EIRP)	Detector
5.725-5.875 GHz	≥ 75	1 MHz	4 W	Average
5.725-5.875 GHz →			75 nW	Average

For general narrow-band use, according to 15.249, the electric field of the fundamental should not exceed 50 mV/m and the harmonics should not exceed 500 μV/m at 3 m. These are based on average measurements, but the peak should not exceed the specified limits by more than 20 dB.

Table 4.25: Bandwidth and power limits for narrow-band use at 5.78 GHz under FCC regulations.

Frequency	Electric Field Strength	~Power (EIRP)*	Detector
5.725-5.875 GHz (fund)	50 mV/m at 3 m	750 μW	Average
5.725-5.875 GHz (harm)	500 μV/m at 3m	75 nW	Average
5.725-5.875 GHz →	500 μV/m at 3m	75 nW	Average

In Japan, the general radiation limit in this region is an electric field strength of 35 μV/m at 3 m.

4.7 Summary

In examining the available frequency bands and their associated field strength and power limits in Europe, the United States, and Japan, the most suitable bands are the <135 kHz, 13.56 MHz, and 2.45 GHz area bands. In Europe and the United States, the 125 kHz and 900 MHz area bands are also favorable. Other bands are typically only favorable under one administration, for example, 6.78 MHz in Europe as opposed to the United States, or 5.78 GHz in the United States as opposed to Europe.

In selecting an operating frequency, in addition to available bandwidth, field strength and power levels, and interference, a number of other factors should be considered. These include not only propagation and behavior in various environments, as discussed in Chapter 2, but also influences on system design and implementation. We will discuss these factors in Chapter 5.

Chapter 5

PHYSICAL IMPLEMENTATION

5.1 Introduction

In Chapter 2 we have seen how electromagnetic fields and wave behave. In Chapter 3, we have seen the theory that describes how tags are able to communicate through modulations of the electromagnetic fields. In Chapter 4, we have seen how the both electromagnetic fields and their modulations are governed by regulations. Now we will study how the tag's basic functions, with their overriding constraints, are physically implemented.

The physical implementation of tags is extremely important on the basis of cost and performance. Though both factors are important to the adoption and benefit of the technology, it is cost, which is of primary importance. Passive tags, just like everyday consumable objects we buy, have a finite life. They will be fabricated and destroyed just as often as consumable items. At the huge volumes of tags expected to be used and produced, minute increases in the cost of a single tag will create massive increases in the overall cost of their use. A tag's performance is determined not only by the tag, but by the readers and associated infrastructure. A tag's cost, on the other hand, is determined only by the tag itself – its design and manufacture, and the volume at which it is produced. For this reason, we must understand the physical components necessary to make a tag function and their influence on cost.

A tag consists of an antenna attached to an electronic circuit. The antenna acts as a transducer between electromagnetic fields and electric energy. A transmission line transfers this energy to circuitry and vice versa. The circuitry processes the energy, stores it, uses it, and redirects it back through the transmission line and antenna, causing controlled responses.

First I will briefly discuss some of the antennas applicable to RFID tags. I will then discuss the tag's integrated circuit, including a review of its various components and overall power consumption characteristics. My focus in this chapter is on cost and performance issues associated directly with the design and implementation of the tag. Manufacturing is without doubt, extremely important, and deserves extensive treatment. However, here I will merely touch on some of the more important issues.

5.2 Antennas

As we know there are two main types of RFID tags – those that operate in the near-field and those that operate in the far-field. Just as the way they interact with the electromagnetic fields and waves is fundamentally different, so is the design of the antennas they use. I will first describe near-field antennas, followed by far-field antennas.

5.2.1 Near-field

Systems in the near-field are either inductively coupled or capacitively coupled. Inductively coupled systems interact with the magnetic field, while capacitively coupled systems interact with the electric field. As we have seen, loop antennas are the desired configuration for inductively coupled systems as their radiated magnetic field dominates the electric field. For capactively coupled systems, however, other antennas such as dipoles and electrode antennas are preferred. I will focus on inductively coupled systems and loop antennas.

Several characteristics affect a loop antenna's cost and performance. Loop area and number of turns relate directly to the magnetic flux through the antenna, the resulting induced voltage and current, and the achievable range. In addition, a larger area and more turns mean more material and higher cost. Materials, due to variations in impedance and quality factor, also affect range. Due to variations in their cost and manufacture, materials also affect cost. I will briefly examine the affects of loop size and materials on the basis of performance and cost.

If we examine Eq. (2.29), we see the direct relationship between induced voltage, area enclosed by the loop, number of turns, and frequency.

$$V_{1\to 2} = j\omega N_2 \mu_0 H_1 A_2 \cos\psi$$

Other factors include orientation, which we can assume to be maximized, and permeability, which is a constant for free space. Permeability may be increased through the incorporation of ferrites as a core, yet this will increase directivity.

The relationship between frequency, number of turns, and area is of particular importance. For maximum induced voltage, the tag's loop should typically be matched to the area of the reader loop. However, due to application size constraints and material cost, the size of the tag's loop is usually limited. We also note that tags operating at lower frequencies, 125 kHz for example, need almost 110 times the number of turns used at 13.56 MHz, for similar performance.

The number of turns of the loop antenna also affects the inductance and the resistive losses of the loop. As an example, we can consider a circular loop (rectangular loops are also common). The inductance for a circular loop is given by Eq. (2.31)

$$L_2 = \mu b N^2 \left[\ln\left(\frac{8b}{a}\right) - 1.75 \right]$$

We also note that inductance increases with the coil radius, or area. Resistance for a circular loop is given by Eq. (2.33)

$$R_{L_2} = N^2 \frac{b}{a}\sqrt{\frac{\omega\mu}{2\sigma}}$$

where σ is the material conductivity.

How the extra area and the number of turns affect cost, is determined largely by the material used.

Loop antennas are typically made either by wound wire or printed conductive ink. Both of these methods have a direct affect on both performance and cost. Inherent in the material used in a loop antenna is an inductance and a resistance. For a maximized quality factor, resistive losses should be minimized, while inductivity should be maximized. Though it may not always be desirable for a maximum quality factor due to detuning effects when surrounded by other antennas and conductors, it will provide maximum voltage.

When large numbers of turns are required, copper wire is generally used. Because resistive losses in wire increase with decreasing diameters, Q decreases. Resistivity of a copper wire is approximately 0.017 $\Omega mm^2/m$.

Printed conducting ink and paste may be used as well, though their use is typically limited to loops with low numbers of turns. Various inks may be used, though silver is used most frequently for its high conductivity relative to cost. Similar to copper wire, printed antenna losses increase with decreasing line widths. Resistivities for silver conductive paste typically vary between 0.01 and 0.05 Ω/square for a thickness of 1 mil. Resistivities for carbon-based conductive paste typically vary between 10 and 50 Ω/square for a thickness of 1 mil [52]. Actual resistivity is sensitive to curing temperature and rate, and other factors.

Another concern is the size of the tuning capacitor necessary to create the appropriate resonance. At lower frequencies where numbers of coils are larger, inductance will likely be higher as well. As such, capacitances will also need to be larger. We will see that capacitors can consume relatively large areas on an integrated circuit, and thus radically increase cost.

One must also consider the impedance used for the load modulation. It may be a capacitor or resistor, or a combination of the two. To decrease power consumption and component size, this impedance should be minimized while ensuring that the signal is still detectable by the reader.

5.2.2 Far-field

Systems in the far-field interact directly with electromagnetic waves. As we have seen, radiation and reception characteristics improve drastically for antennas with at least one dimension on the order of half of a wavelength. At this size antennas are resonant and have a real input impedance. This makes for improved radiation and reception efficiency, easier matching, and more power delivered to and from the attached circuitry.

We should note that systems operating in the far-field do not necessarily have to incorporate half-wavelength dimensioned antennas. Smaller dimensioned antennas may be used, however, input reactance will increase and input resistance will decrease. This may create difficulties in matching, and will result in lower radiation efficiency, meaning less power available to the tag, and less power available to be scattered back to the reader. It certain cases, though, this may be beneficial, as the structural mode scatter that contributes directly to noise, will be reduced.

There are several types of resonant antennas. Commonly used types include half-wave and folded dipoles, and various shapes of microstrip patch antennas. Generally, dipole antennas may use wires, or they may be printed onto a substrate. Microstrip patch antennas are printed onto a dielectric, often a printed circuit board material, with a conducting ground plane underneath. I will briefly discuss some of the performance and cost related issues of these antennas.

5.2.2.1 Dipole Antennas

Wire dipoles can be more costly than printed dipoles due to manufacturing processes, so for very low cost applications, printed dipoles may be preferred. It is important to reduce ohmic losses in the antenna, so generally, the dipole width should be relatively wide. The material used should also have a low resistivity and a high quality factor.

Though they provide more efficient power transfer, half-wave dipoles, by virtue of their resonance, can have a relatively narrow bandwidth. Bandwidth can be increased, though, by increasing the widths, or effective diameters [13].

5.2.2.2 Microstrip Patch Antennas

Microstrip patch antennas are planar antennas that are typically printed directly onto a dielectric circuit board. They consist of three layers: the conductive patch on top, a conductive ground plane on the bottom, and a dielectric material in between. The patch is usually some appreciable fraction of a wavelength in dimensions, while the ground plane extends further in dimensions. The dielectric material is typically some printed circuit board material – often, PTFE, or Alumina. Because they have a ground plane, their radiation pattern is limited to the area above the ground plane [53].

The patch can take any number of shapes. Rectangular, circular, and elliptic patches are commonly used. Different shapes are often explored to vary the various parameters, including input impedance, radiation efficiency, and polarization. Patches are fed through a small transmission line, either directly, coupled electromagnetically, or through an aperture. Each feed-line method has its own degree of freedoms, but like the patch shape, the main factors that they influence are input impedance, radiation efficiency, and polarization.

Bandwidth is an important concern with patch antennas as basic designs typically have a fractional bandwidth relative to the center frequency of only a few percent. For a rectangular patch, the fractional bandwidth can be given by [13]

$$B = 3.77 \frac{\varepsilon_r - 1}{\varepsilon_r^2} \frac{W}{L} \frac{t}{\lambda} \tag{5.1}$$

where ε_r is the relative permittivity of the dielectric, W is the patch width, L is the patch length, and t is the thickness. We see that increasing thickness increases bandwidth, while increasing the dielectric permittivity decreases bandwidth. Selection of dielectric is important not only for bandwidth but for overall performance of the patch. Dielectrics should have minimal losses, high temperature stability, high dimensional stability, and uniformity.

Microstrip patch antennas can be costlier than printed dipoles, as they require the ground plane and dielectric material. Printed dipoles of course need a substrate, yet the substrate material does not have as great as an affect on performance as does the patch dielectric.

In general, performance offered by dipoles and patch antennas can be comparable, and their use in many ways depends on the application, the operating frequency, and the cost of materials and manufacture.

5.3 Integrated Circuit

For reasons outlined in the introduction, I have assumed the use of an integrated circuit – more particularly, a single integrated circuit. Often it may be desirable to use external discrete components in addition to a chip. While this may be acceptable for low volume applications, for reasons of cost, manufacturing, and performance, the tag should have only a single integrated circuit.

The goal of this section is not to design an integrated circuit, but rather to describe the various components necessary for its function, including their effects on performance and cost. The more important variables controlled by the chip include – power consumption, size, and cost. Because the tag relies on electromagnetic fields for power, energy is limited. Power consumption must be reduced and managed in a way so as not to detrimentally affect the performance of the tag, principally its operating range. Size directly affects cost. In general, the more silicon used, the more expensive the chip. The number of components used, and their size should be reduced to minimize cost. It should be noted, however, that more efficient use of gates can lead to more switching per gate, which leads directly to higher power consumption. We will discuss this tradeoff in a later section.

5.3.1 Overall Architecture

Before describing the various components used in integrated circuits, I will briefly discuss the overall architecture of RFID tags. Whether they operate in the near-field or far-field, their overall architecture is similar.

A tag, which is simply a carrier for a unique identification number must perform several basic functions. It must receive and rectify the incoming signal for the extraction of energy and information. It must store and manage the extracted energy to power the tag. From the extracted information it must establish a clocking signal with which to drive its digital circuitry. Through this circuitry, it must process the information and make the appropriate modulations of the incoming signal through load or backscatter modulation as described in Chapters 2 and 3.

The RF front end is responsible for bi-directional interfacing between the antenna and other functional blocks of the tag. In the RF front end, energy and data are extracted from the input signal and sent to power supply, biasing, clock recovery, and data processing circuitry [54][55]. Because of the large variations in field strength with respect to distance, over-voltage protection is also located in the front-end. As directed by the data processing circuitry, information is modulated back through the antenna to the reader.

The power supply, biasing, clock recovery, and data processing circuitry are implemented in a combination of analog and digital circuits. Memory is also included. In the case of read-only tags, it is typically read-only memory (ROM) or write once read many (WORM).

5.3.2 Components

The components necessary to perform the tag's functions include analog and digital circuits with passive components, active components, and memory. They all may be incorporated and interconnected in a single integrated circuit. Because many of the current fabrication processes were designed for digital circuitry, implementation of analog circuits next to digital circuits can be difficult. One problem is reduced yield, as analog circuits tend to suffer more variations [56]. In addition, as we will see, important passive components can require relatively excessive amounts of area, and given their loose tolerances, still offer only satisfactory performance.

After briefly describing the physics of semiconductors, I will describe the various pertinent components. I will begin with transistors, followed by passive components, diodes, and other important devices, including oscillators, phase-locked loops, digital circuits, and memory. Finally I will describe the overall power consumption characteristics. I will not describe the fabrication process and interconnect properties in any depth.

5.3.2.1 Semiconductors and Transistors

Semiconductor materials are those that are neither good insulators nor good conductors. They are made by introducing controlled amounts of impurity atoms in a process called doping. By doping a region of a material with another material having more valence electrons, the region becomes an "n-type" region. Alternatively doping with a material having fewer valence electrons, the region becomes a "p-type" region. Applying a voltage causes excess electrons to move. Silicon is the most common material used for semiconducting devices for various reasons. It is abundant in nature and is relatively insensitive to temperature various due to its large bandgap. Gallium Arsenide is used for its speed advantages – it can be over twice as fast. Gallium Arsenide, however, is more expensive than standard Silicon, and not nearly as abundant [57][58].

Transistors, which are semiconducting devices, have been in development since the late 1940's. Bipolar technologies were developed first, followed by metal-oxide semiconductor (MOS) technologies. MOS technologies offered reduction in size and power consumption. P types (PMOS) were developed first, followed by N types (NMOS). Complimentary MOS (CMOS) devices were developed for further reduction in power consumption at a cost of an increase in size. BiCMOS devices combine fast bipolar technologies with the power efficiency of CMOS. The CMOS market however, dwarfs those of other technologies, and for the foreseeable future, it is expected that CMOS will have a significant cost advantage [59].

A MOS device, in some respects, can be thought of as a parallel plate capacitor. A simple example is an NMOS device. The gate, which is typically heavily doped polysilicon, represents one plate, which is separated from the semiconductor, the other plate, by a dielectric material. The gate lies above and between two heavily doped n-type regions – one the source, and the other, the drain. All other regions on the semiconductor are p-type regions.

When there is no voltage at the gate, there is no conductivity on the device. When voltage is applied at the gate however, an electric field is created which induces charges of opposite polarity in the semiconductor. When enough voltage is applied, charges are pulled from the surface of the semiconductor and electrons from the source and drain create an "inversion" layer. This inversion layer is a conductive path between the source and the drain. Charges moving at a velocity derived from the electric field and the electron mobility produce a current. Variation of the relative sizes, distances, and voltages produces a number of different effects.

These devices have both static and dynamic effects. The dynamic effects are dominated by capacitances at various places in the device. The source and drain form a capacitance from a reverse bias junction with the substrate. In addition, parallel plate capacitances are formed between the gate and the source, and the gate and the drain, as well as from the gate to the channel and the channel to the bulk material.

Complimentary metal oxide (CMOS) devices combine PMOS devices with NMOS devices. Though larger, they provide reduced power consumption. Transistors can be regarded as the main building block of an integrated circuit because they may be used as passive components, such as resistors and capacitors, as well as active components.

5.3.2.2 Passive Components

In RFID tags, passive components are particularly important for the RF front end and power management systems. I will briefly review their use in integrated circuits.

5.3.2.2.1 Resistors

There are relatively few good options for creating resistors in standard CMOS technology. Interconnect material such as polysilicon, metal, or other more specialized materials, may often be deposited in bending lines. Polysilicon is typically more resistive than metal and offers approximately 5-10 Ω per square with a tolerance of around 35%. It also has a low parasitic capacitance per area, a moderate temperature coefficient, and a low voltage coefficient. Metal typically gives on the order of 10 MΩ per square [60].

Wells are another option when high resistances are needed. They offer resistivities in the range of 1-10 kΩ per square, yet they suffer from large parasitic capacitances due to the junction between the well and the substrate. They have an initial tolerance of 50-80%, a large temperature coefficient, and a large voltage coefficient.

A MOS transistor may also be used as a compact resistor. It, however, has a loose tolerance and high temperature coefficient due to its dependence on electron mobility. N-type or p-type source-drain diffusions are another option, which provide similar resistivity and temperature coefficients to polysilicon. Parasitic capacitance can be significant, and the voltage coefficient is noticeable [60].

5.3.2.2.2 Capacitors

Capacitors can be formed in a number of ways. One common method is through creating parallel plate capacitances with interconnect material. Because the inter-level dielectric is relatively thick, capacitance per area is typically small (5×10^{-5} pF/μm^2). Parasitic capacitance from layers directly below, can be detrimental to performance and may have 10-30% of the main capacitance.

Interconnect material can also be used to create capacitance through lateral lines, along with parallel plate methods. As the capacitance depends on perimeter, some have used fractals for their large perimeter with respect to area. Such methods may offer factor of 10 increases in capacitance [60].

Because of the gate capacitance common to a transistor, a MOS device may also be used. Depending on the dielectric thickness, capacitances of 1-5 fF/μm^2 may be achieved.

Another option is to use the junction capacitance inherent in a p to n well junction. Because the junction capacitance depends on the applied bias, capacitance can be tuned in this way [60].

5.3.2.2.3 Inductors

Inductors with low noise, distortion, and power consumption are typically hard to implement in CMOS technology. The planar spiral is the most widely used inductor, yet it suffers from large areas for minor inductances. Losses due to resistance and non-uniform current distribution can be large. In addition, parasitic capacitances are formed with surrounding structures so resonance can be formed limiting the inductor's usefulness. Quality factors are very low compared to those of discrete components [56].

5.3.2.3 Charge Pumps

Charge pumps may be used for situations where a high voltage relative to the available voltage must be generated. Often charge pumps will be used to supply a high voltage for writing to EEPROM. They are created by connecting multiple stages of switched capacitors. When each stage is charged, voltages can be combined to create a higher output voltage.

5.3.2.4 Diodes

Diodes are necessary in RFID tags for rectification of incoming signals for both energy extraction and signal detection. They may be made from transistors, yet because transistors may have a larger bias and offer slower switching, Schottky barrier diodes are often preferred, particularly for higher frequency applications.

Schottky barrier diodes are formed from a metal to semiconductor contact (n-type or p-type). Depending on the configuration, they offer zero bias, fast switching, relatively low junction capacitance and low power consumption. They are often used in high-frequency signal detectors, mixers, and switching circuits and power supplies. In combination, they may be used as voltage doublers [61][54].

5.3.2.5 Oscillators

Oscillators, as their name suggest, provide a signal with, preferably, a constant frequency. Their implementation can be relatively simple, yet they may suffer from poor performance – amplitudes and frequencies may vary excessively. Oscillators may be used in RFID tags, yet depending on their application, suitable performance may demand excessive complexity, size, and cost.

5.3.2.6 Phase-Locked Loops

Phase-locked loops provide a means to synchronize with the phase of an input signal. They are used often in communications systems for a variety of functions. They may be used to provide a programmable frequency for oscillator applications. They also may be used for frequency modulation and demodulation. In digital systems, they may be used for clock recovery, and clock signal generation [60].

At the most basic level, a phase-locked loop incorporates a phase detector and a voltage controlled oscillator in negative feedback. They may be used in RFID tags, but because of the number of components they require, they may add excessively to the complexity, size, and cost of the tag.

5.3.2.7 Digital Circuits

The fundamental building block of all CMOS logic gates in the CMOS inverter, which combines an NMOS transistor with a PMOS transistor. From this pair of transistors a wide variety of logic circuits can be constructed, including such circuits as flip-flops, latches, one-shots, and shift registers.

5.3.2.7.1 Flip-Flops and Latches

Flip-flops are a type of sequential logic, which exhibit memory. From these devices, one can construct counters and other more complicated devices. Along with latches, they are often used to control the timing of computations in synchronous systems. In such an application, a flip-flop may make its output equal to its input at a clock edge. In one of its simplest forms, a flip-flop combines two NAND gates, each typically consisting of four transistors (two complimentary pairs of PMOS and NMOS devices).

A latch typically describes a device that holds a set of bits. When enabled, its outputs follow the inputs, but when disabled it holds its last value. When used for controlling timing in synchronous systems, it may be enabled on one clock phase and disabled the next.

5.3.2.7.2 Shift Registers

Connecting a series of flip-flops so that the output of one is the input of the next, and all clock inputs are driven simultaneously, creates a shift register. At each clock pulse, the data in the register shifts to one side while more data enters from the other side. Shift registers can be useful for converting parallel data to serial data and vice versa. They may also be used as temporary memories.

5.3.2.8 One-shots

Also known as a monostable multivibrator, or an RC monoflop, a one-shot is a variation of a flip-flop where the output of one gate is capacitively coupled with the input of the other gate. These devices sit in one state until forced to another state by a momentary input pulse. After a delay determined by the circuit parameters, it will return to the first state. They are triggered by a rising or falling edge and are often used to generate pulses of variable widths. Because they are combinations of analog and digital circuits, they can suffer from reliability problems, and in most cases should be used sparingly [62].

5.3.2.9 Memory

Memory can be implemented on an integrated circuit in a number of ways. For low-cost RFID applications where the tag simply has to store a static unique identification number, the requirements are not demanding. The main concern is to limit the area consumed. I will briefly review some of the main types of memory suitable for storage of a UID: ROM, WORM, EEPROM, and FRAM.

Read-only memory (ROM) is typically recorded at the manufacturing stage through laser etching or direct inclusion on the chip mask. Write-once-read-many (WORM) memory, allows data to be recorded once but read many times. This allows recording of data downstream of initial manufacture.

Electronically Erasable Read-only Memory EEPROM, allows both writing and reading yet it can require high voltages and has a relatively high power consumption. It is also larger in size than other memory technologies and may take longer to write. Ferro-electric random access memory represents a major improvement from EEPROM. Like EEPROM, it can store data for long periods of time, yet writing requires a lower voltage and much lower power consumption [6].

5.3.3 Integrated Circuit Power Consumption

An understanding of the power consumption characteristics of the tag are important not only from the standpoint of integrated circuit design, but also from the standpoint of communications protocol design. Coding, modulation, and command routines can and should be designed to take the power consumption characteristics into account. Minimization of power consumption can have a direct positive effect on range.

In an RFID tag integrated circuit, there are analog as well as digital circuits. Though a tag will often contain more digital than analog circuitry, analog circuitry power consumption may be disproportionably higher. For this reason, analog circuitry must be carefully designed and optimized for low power [63]. Antenna matching, RF front end, and power-supply circuitry should be designed for maximum power transfer efficiencies.

Digital CMOS power consumption characteristics are comparably much better, yet contribute significantly to the overall power consumption of the chip. Digital CMOS power consumption can be broken into two classes: static and dynamic. Static power consumption is that power consumed when a device is idle. Ideally it is zero. Some leakage does occur through parasitic diodes and other means, yet for the most part it is negligible. Dynamic power consumption, on the other hand, is not. I will briefly review dynamic power consumption.

Dynamic power consumption consists of two components: short circuit consumption, and capacitive consumption. Short circuit power consumption is due to a momentary direct current from the supply to ground during the switching of a gate. It has been shown that it is typically less than 10% of the capacitive consumption, so it is often neglected [64].

Capacitive power consumption is that resulting from the charging and discharging of parasitic capacitances throughout the integrated circuit. As we have seen, both devices and interconnect have multiple parasitic capacitances. The energy in the capacitances is related to the driving voltage, and the total power consumption is related to the activity and frequency of the switching. Because short-circuit dynamic power consumption is neglected, the equation for dynamic power consumption becomes:

$$P_{dyn} = \alpha f C V^2 \tag{5.2}$$

where α is the activity factor, f is the data rate, C is the total capacitance, and V is the supply voltage. The activity factor α, represents the expected number of zero to 1 transitions per data cycle. The average data rate f, represents the clock frequency in synchronous systems. The dynamic power consumption is essentially affected by three main variables: the switching activity, capacitance, and the square of the supply voltage.

Because of its quadratic relationship, reduction of the supply voltage can bring the greatest reduction in power consumption. In addition, because it is globally applied, its reduction has a significant effect. The main problem is that it can cause the overall performance to suffer through increased delays. Because of the quadratic relationship, it may often make sense to reduce voltage at the cost of increased capacitance and switching activity [65].

As we have seen, capacitance is present in both devices and interconnects. It can be reduced by decreasing size of line width and components, and distances between them. A known side effect of this, however, is decreased current drive. This can slow performance and possibly require an increase in supply voltage.

The third variable affecting power consumption is switching activity. The data rate f, essentially represents the frequency at which switching could occur. In synchronous systems f is the clock rate. The activity factor α, in the case where there is no glitching, is the probability that a transition will occur during a single cycle. Glitching refers to spurious transitions and can make α greater than 1. Through sensible design it can be minimized. For a random data signal, it has been shown that the maximum activity factor α is 0.5, for a simply logic gate, between 0.4 and 0.5, and for a finite state machine, between 0.08 and 0.18 [66].

Through reduction of these variables, power consumption can be minimized. Other techniques include adding parallelism for certain functions. This may add area, but should allow for increased performance given a lower voltage. Another suggested practice is to avoid waste, such as unnecessary clocking when idle, and unnecessary or over-performance [65]. It is also possible to implement energy-recovery CMOS, which allows the conservation of energy that would normally be dissipated as heat [67].

5.4 Attachment and Packaging

Attachment of the tag integrated circuit to the antenna and packaging of the entire tag, must be also be considered. Not only does attachment and packaging affect cost and size, but also performance. Attachment can affect power transfer between antenna and integrated circuit. Packaging can affect not only physical reliability and robustness, but also sensitivity to the environment and characteristics of tagged objects. As such, attachment and packaging concerns are a tradeoff between cost, related to materials and manufacturing, and performance.

5.5 Summary

In this chapter we have reviewed some of the fundamental concerns associated with physical implementation of the tag. The antenna, integrated circuit, and associated packaging can all have a significant effect on cost, size, and all of the performance concerns. Materials and geometry can be particularly important in antenna design. Both of these issues can affect cost significantly. In regards to the integrated circuit, size of the chip should be minimized to minimize cost. CMOS, which is currently the most widely used and developed integrated circuit technology, is consequently the most suitable for use in RFID tags. The largest component of CMOS power consumption is a function of switching activity, capacitance, and voltage. These should all be minimized within the bounds of integrated circuit performance in order to maximize range. Now that we have reviewed the hardware concerns, in the next chapter we will study the software concerns.

Chapter 6

COMMAND PROTOCOLS

6.1 Introduction

Now that we understand the various constraints on RFID systems including the electromagnetics, communications, regulations, and physical implementation details, we can consider the higher-level command protocols and algorithms that drive the actual identification of tags.

Because a tag is simply a carrier for a unique identification number, the goal is purely to read this number. If a single tag is in the interrogation range of a reader, it would not require any commands. It may simply respond with its data contents once energized sufficiently. However, when multiple tags within the interrogation range of a single reader respond simultaneously, their signals can interfere. This interference is referred to as a collision and typically results in a failed transmission. In order to prevent these collisions, RFID systems require a set of commands based around some protocol. These protocols are often referred to as anti-collision protocols or algorithms.

In considering the design of a command set and anti-collision algorithm, we must consider the various constraints imposed on the system. In particular we must consider the effects on the reliability of the communications, the required bandwidth, and the physical implementation of the tag integrated circuit, which translate directly to power consumption, range, and cost. For the most part, we wish to minimize bandwidth, maximize reliability, and minimize the demands on the integrated circuit.

In order to successfully communicate with a single tag among many within the reader's interrogation range, we must establish an exclusive link with this tag. There are a number of methods for doing this. After reviewing the various methods, I will analyze the performance and requirements of two common types. I will then discuss the basic command requirements associated with these algorithms.

6.2 Anti-Collision Overview

As tags operate at greater ranges and more and smaller objects are tagged, the number of responsive tags in a reader's field will increase. Such an increase will result in a higher probability of collision. Thus, there is a present and increasing need for effective anti-collision methods.

Anti-collision methods used in RFID systems are similar to multiple access communications conflict resolution methods and the various networking contention protocols including the Aloha and Carrier Sense Multiple Access (CSMA) family of protocols [68]. RFID anti-collision methods, though, are constrained by low computational power, and often little, if no, memory. In addition, they typically must be optimized for low-power operation to prevent reduction in range, in the case of passive tags, and promote longer battery life, in the case of active tags. In addition, as tags typically are only able to communicate with the reader, use of CSMA related methods is not possible. Furthermore, wireless channel conditions are highly variable relative to wired network systems – bursty noise can harmfully interfere with the relatively short communications transmitted by both reader and tag. Overall complexity should also be minimized, as it often increases total cost and adversely affects other aspects of overall system performance.

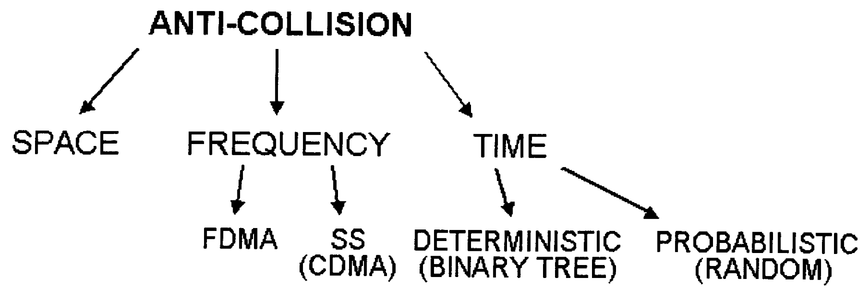


Figure 6.1: Classification of anti-collision methods.

A wide variety of anti-collision methods exist and can be categorized in several ways [69]. The broadest classification is that of domain: space, frequency, and time. In space-domain methods, tags are typically localized in space to achieve isolation and subsequent identification. This can be achieved through variation of the reader's range, or in the case of passive tags, variation of the power transferred to the tag. One method uses both techniques to identify tags based on detection of the strongest responses at varying ranges [70]. Another method uses an array of low range readers such that only one tag will be in an reader's field at a given time [6]. Triangulation using ultra-wideband communications and positioning methods might pose yet another method of tag isolation and subsequent identification [71]. The fundamental problem with spatial methods is that precise control of range can be difficult to achieve. Precise control becomes even more essential as the number of tags in the reader's field increases and the distances between them decrease. Currently the best use of these methods is in combination with frequency or time based methods.

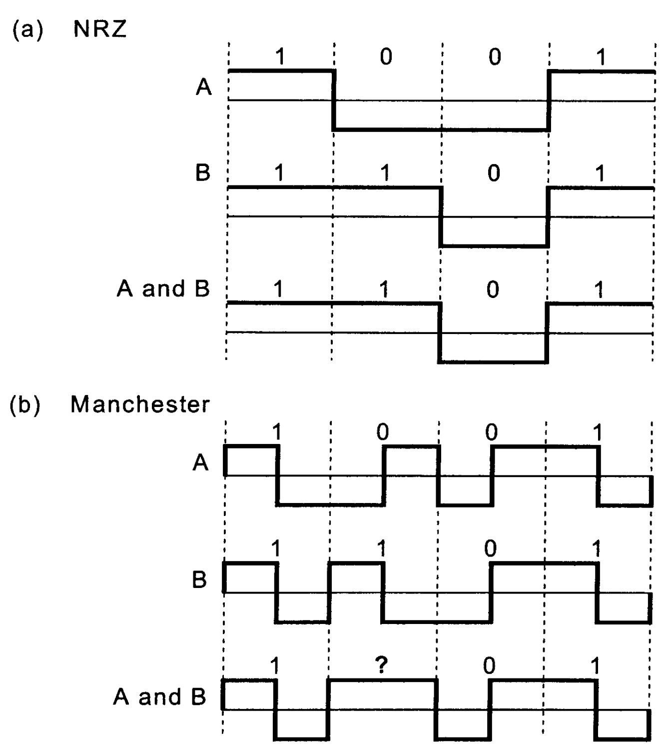
Frequency domain methods generally allow for robust wireless communications, yet can add excessive complexity and cost. Frequency Division Multiple Access (FDMA) systems divide the total available bandwidth into fixed-width channels. FDMA requires accurate frequency sources and selective band-pass filters, and can thus be difficult to implement in low-cost RFID Systems. Magellan Technology has implemented a system combining FDMA methods with Time Division Multiple Access (TDMA) methods. While its performance characteristics are no doubt positive, it is unclear how this implementation has affected complexity and cost. Code Division Multiple Access (CDMA) systems have a number advantages over FDMA, such as better adaptability to varying traffic load, increased capacity, and a processing gain [68]. CDMA and other Spread Spectrum (SS) methods, though, can also be difficult and costly to implement based on their increased complexity. In addition, their use can be limited by bandwidth limitations. For this reason, SS methods, including Frequency Hopping (FH) and Direct Sequence (DS), are generally confined to operation in UHF and microwave bands.

The vast majority of existing RFID anti-collision methods are time-domain methods. In these methods the positions of the transmissions are varied in time. These methods can be further classified into deterministic and probabilistic schemes.

Deterministic schemes are those where the reader issues a query or command instructing certain tags based on their unique identification number (UID). Based on this UID, a reader will either poll through a list of known tags in the field, or perform some variation of a binary tree search. Polling methods can be extremely time consuming unless there are few tags in the field. They also require that all UIDs are previously known. Binary tree search methods are much more common. Texas Instrument's Tag-It [72] uses a type of binary tree search. A number of other variations exist. Law et. al. [73], propose an efficient memory-less protocol. Jacomet et. al. [74] propose a related method. SCS uses yet another method [75]. Certain binary tree search methods can be performed rapidly, yet typically only work correctly if no tags enter the reader's field during the search. I will analyze a protocol based on Law's protocol.

Probabilistic schemes are those where tags in the reader's field respond at randomly generated times. There are a number of variations of these schemes where the reader exerts varying degrees of control over the tags. Many of these variations are based on the Aloha scheme for multiple access in networking [68]. In this scheme a node transmits a packet immediately after receiving a packet. If a collision occurs, the node becomes saturated and transmits the packet again after a random delay. In slotted Aloha, the transmission is done in slotted time rather than continuous time. Furuta [76] describes a variation of this protocol for contact-less IC cards. The ISO 15693 standard supports a mode similar to slotted Aloha. BTG's SuperTag [77] operates on an un-slotted Aloha principle but incorporates further additions. After reception of data, tags can be muted or their repetition rate can be slowed. Through tag muting, the SuperTag protocol allows for counting of identically coded items. Another SuperTag variation specifies muting of all tags but the one being read, to ensure that no collision occurs during data transmission. After a certain period, the muted tags are activated, and the cycle repeats. In other methods, the reader, by sending a gap or power burst, prompts the tags to respond after a randomly generated delay [78] [79]. After a tag transmits its data, it will enter a quiet mode so as to reduce the possibility of future collisions. Certain methods specify that after the reader's initial command for tags to respond, the tags will respond after the random delay with a short chirp instead of its full data contents. The reader will then quiet those tags that were slower to respond, and prompt the tag to transmit its full data contents [80]. As a representative of these protocols I will analyze the SuperTag family of protocols.

Figure 6.2: Coding scheme can affect the ability to detect a collision. NRZ (a) does not allow for collision detection whereas Manchester (b) does.



Many anti-collision methods require the ability to detect collisions. The most common method of collision detection relies on coding scheme. When simultaneously transmitted signals coded by certain schemes add, they cannot be resolved. Manchester and other transition codes inherently allow for this means of collision detection (Figure 6.2b); [6] NRZ and related level codes, on the other hand, do not (Figure 6.2a). Other methods rely on modulation schemes. Through FSK modulation in tag to reader transmission, readers can detect "wobbles" when multiple tags respond simultaneously [78].

6.2.1 Anti-Collision Analysis

In this section I will analyze two protocols in regards to performance and implementation. Though every protocol is unique, and has its own advantages and disadvantages, the two families I will consider, SuperTag and QT can, in some respects, represent their respective domains. Both SuperTag and QT achieve multiple access in the time-domain. SuperTag is a probabilistic scheme, whereas QT is a deterministic scheme.

6.2.1.1 SuperTag

The SuperTag protocol was designed by CSIR of South Africa and is currently licensed by BTG. It consists of a family of similar protocols based on the Aloha network contention protocol. As previously mentioned, the SuperTag family of protocols exploits randomness in order to achieve identification of multiple tags. There are four variants. In each variant, the routine is directed by the reader's transmission of a start command [77].

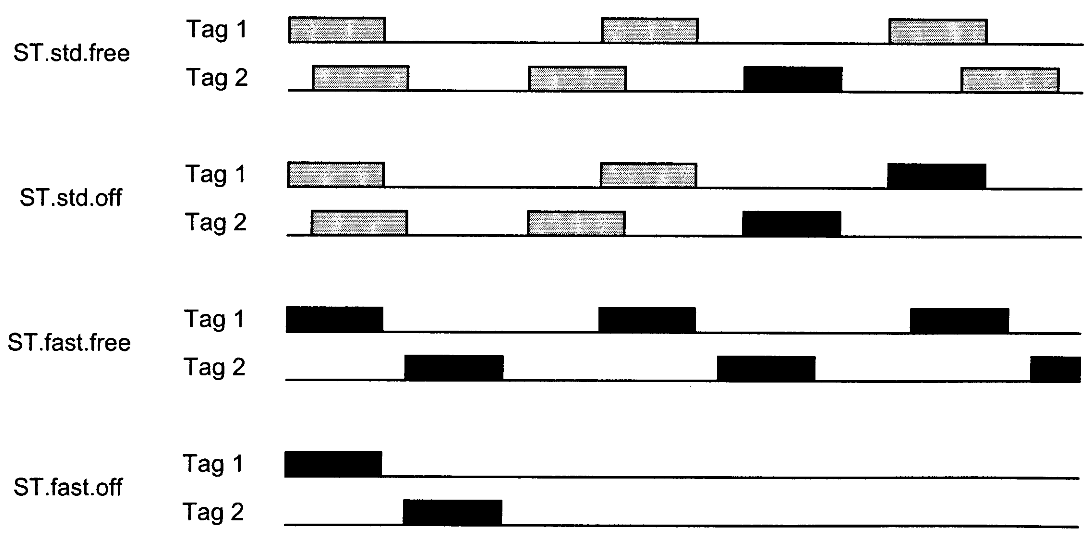


Figure 6.3: The four SuperTag variants. Black fill incdicates a successful read. Grey fill indicates a failed read.

In the simplest variant, after reception of the initial start command, tags will respond with their full unique identification after a randomly generated delay. They will continue to respond with that same delay even after being identified. I will refer to this as ST.std.free for standard free-running SuperTag.

In a slightly more complicated variant, the reader, after receiving the UID of a tag, issues a command to that tag instructing it to refrain from further transmissions. I will refer to this variant as ST.std.off for standard shut-off SuperTag.

In ST.std.free, tags may start transmission even while other tags are transmitting, this can cause a collision and failed identification. Another variant prevents this by the reader issuing a command to mute all tags but the one that first responds. I will refer to this variant as ST.fast.free.

The final variant of SuperTag, is the most complicated. As in ST.fast.free, the reader mutes all tags but the one that first responds. However, in addition, the reader mutes the tags after they have been identified. I will refer to this variant as ST.fast.off.

6.2.1.2 *Query-Tree Based*

The other broad class of time-domain anti-collision methods includes those that are deterministic in that they use binary tree searching methods to determine individual UIDs. These methods can be further classified by the transmissions required from the tag. Some schemes require the tag to transmit large portions of, or the full UID, at various points throughout the search. Other schemes use the reader to "build" a UID bit by bit while the tag simply acknowledges if a there is a match. I will focus on the later, as implementation requirements on the tag are often simpler. Many have proposed such methods [73] [75], I will analyze an algorithm based on the QT protocols.

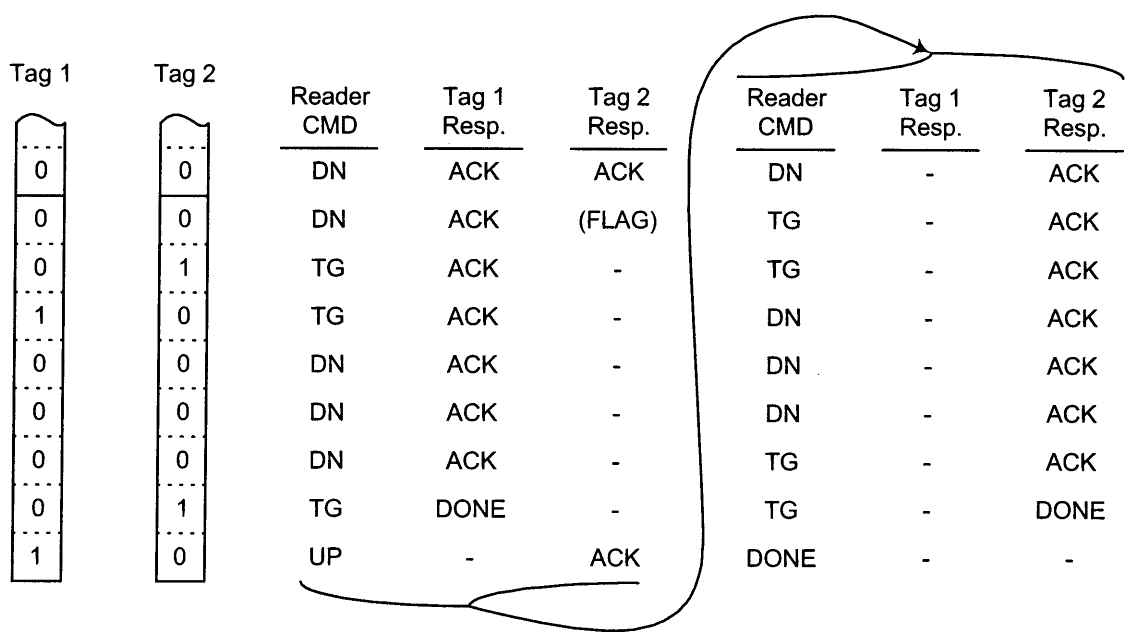


Figure 6.4: The final nine bits of a sample QT.ds cycle with two responsive tags.

In this scheme, the reader may issue two search commands: one instructing tags to walk down the tree, DN, the other instructing them to toggle and walk down the other branch of the tree, TG. If the last identified bit was a 0, then a DN command instructs tags to continue down the 0 branch. If it was a 1, then it instructs them to walk down the 1 branch. Toggle, simply instructs tags to walk down the branch opposite of the last identified bit. If tags have a matching 0 or 1 at the next bit position, they respond with a simple acknowledgment. If there is no match, they mute themselves, and leave the pointer to the last identified bit.

After an initial start command, the tag is preset to a 0 in the last bit pointer, and the reader may issue either a DN or TG command. Depending on the implementation of the algorithm in the reader, the reader may issue any sequence of DN and TG commands. After each command, it will listen for an acknowledgement from tags with matching bits. This will proceed until the entire length of the UID has been traversed, in this analysis, 64 bits, and a single tag has been matched. When a tag has been matched it will confirm its match and mute itself. In this analysis, I will assume the issuance of DN commands followed by TG commands when no acknowledgement is received.

As mentioned, when a tag does not match a 0 or 1 at the current bit position, it will mute itself and leave its pointer to the last identified bit. In addition, it will set a "recognize" flag noting failure. After the tree has been traversed and a single tag has been identified, the reader may issue an UP command. This prompts tags to un-flag their "recognize" flag and set their pointers to their nearest 8-bit boundary based on the entire 64 bits of the UID. The search may commence on this last 8-bits. If no tags lie within this branch of the tree, another UP may be issued. I will refer to this protocol as QT.ds.

In the binary search based protocols, as in this one, the distribution of UIDs can affect the performance. Best performance will occur when tags are serially distributed. Worst performance will occur when UIDs are randomly distributed. I will consider both cases here.

6.2.1.3 Results

Table 6.1: Simulation results comparing time required to reader various numbers of tags in the reader's field for the QT.ds and SuperTag protocols. QT.ds results include those for both a serial and random UID distribution.

No. of Tags in Field	**Time to Read [s]** QT.ds (ser)	**QT.ds (ran)**		**ST.Std.Free**	**ST.Std.Off**	**ST.Fast.Free**	**ST.Fast.Off**
10	0.0218	0.117		0.18	0.09	0.038	0.021
50	0.07	0.656		2.8	0.807	0.447	0.183
100	0.13	1.21		5.22	2.316	1.035	0.326
500	0.922	7.32		> 200	17.314	6.844	1.683
1000	4.01	15			> 200	18.734	3.945

Results from simulation are shown in Table 6.1. The SuperTag data was obtained from CSIR's SuperSim simulation. The SuperTag results given are averages of 10 runs for the various numbers of tags and maximum response delay times were chosen to be optimal. The QT.ds data was also obtained from simulation. In each case data rates are 64 kbps and identification code length is 64-bits. Though times may vary with different command structures and by virtue of the small number of samples, confidence intervals are large, the basic trends are illustrated here.

6.2.1.4 Comparison

These two classes of protocols can be compared on two grounds – performance and implementation. In regards to performance, we should note, in particular, identification rates, and relative behavior with varying numbers of tags in the interrogation range. Performance in the presence of noise and other unique advantages and disadvantages should also be considered. Regarding implementation, required circuitry and commands are particularly important. I will discuss the SuperTag and QT protocols on these terms.

In examining the SuperTag protocols alone, we notice the increase in linearity as we move from ST.std.free to Std.fast.off. This is due to the additional functions which either shut-off tags after they have been identified, and/or mutes tags while one is transmitting. Both of these functions have the affect of reducing the number of responding tags in the field and hence increase linearity. We see that ST.fast.off is essentially linear with the number of tags in the field.

In examining the QT.ds protocol, we notice the marked difference in performance for a serial versus random distribution of numbers. In certain applications such as those in factories or distribution centers, UIDs may very well be serially distributed. In other applications such as in store check-out lines, they may be randomly distributed. Both cases should be considered. In the SuperTag protocols there will be no difference with the distribution of codes. Also, we note that the QT.ds protocol is linear with the number of tags in the field for a random distribution of UIDs. We also note that as the number of tags in the field nears the "group" capacity of 256 tags (by virtue of the 8-bit UP), the number of tags read per second increases. This explains the non-linearities seen for a serial distribution of UIDs.

Noting the performance between the two protocols, performance is similar for low numbers of tags in the field. As the numbers increase, the non-linearites become exposed. Required commands transmitted by the reader should also be considered. Of all the SuperTag protocols and the QT.ds protocol the ST.std.free will likely require the least amount of signaling from the reader. In certain frequency bands with tight limits, this attribute may be useful. The directed search of the QT.ds protocol requires a significant amount of signaling, as does the ST.fast.off protocol. In general, the more signaling, the more bandwidth, and the higher the probability that an error may occur.

Beyond performance and commands, implementation is important. The SuperTag protocols require some means of generating a random delay before responding. This requires some form of random number generator. Many implementations may be possible, some more complex than others. Some have suggested using the tolerance inherent in circuit components as a means of generating a random number [80]. Others may use a form of oscillator. The more advanced SuperTag protocols must also be able to recognize the commands to shut off and mute. The QT.ds protocol requires some form of comparator to compare the transmitted bit with the bit stored in memory. They must also maintain pointers to bits and some state information. In any case, excessive circuitry leads to more area on the chip and higher cost.

The QT.ds and SuperTag protocols each have their own unique advantages and disadvantages. The QT.ds protocol has the inherent ability to select tags with particular UIDs. This can be useful in applications where specific tags or groups of tags may need to be addressed or ignored. The SuperTag protocols, at least he ST.std.off, require fewer transmissions from the reader, reducing bandwidth and probability for error. A unique disadvantage of the QT.ds protocol and other binary search schemes is that tags arriving during the search may not be included in the deductions. Depending on implementation, however, this may not be a problem. A problem with the SuperTag protocols and other related random response protocols is the requirement of specifying a maximum delay time. This must either be transmitted to the tag or preset at manufacture. The effect will vary depending on the number of tags in the field.

6.2.2 Channelization

Depending on the implementation and band of operation, anti-collision schemes may be "channelized" in frequency, time, and space, or a combination of all three. This can achieve improved identification rates. It also, however, can complicate the system and increase probabilities of error.

6.2.3 System Considerations

The performance of the individual protocol is important, however, it is the overall system level result which is most important. With a well chosen identification coding scheme, and an anti-collision protocol that allows addressing of particular codes or groups of codes, much of the anti-collision can in fact be moved off the tag into the host or reader. For example, in certain situations large portions of the identification code may be already known, or it may be desired to only identify tags with particular codes. In such situations, the reader can directly address the tags with those codes. Deterministic protocols allow this; most probabilistic schemes, however, do not.

6.3 Commands

There are generally two classes of RFID systems in regards to commands and protocols: tag talks first (TTF) and reader talks first (RTF). In TTF systems, tags respond as soon as sufficiently powered. In RTF systems, tags initialize to a muted state. They only respond unless commanded by the reader to do so. RTF methods are preferred largely because they inherently prevent interference from incompatible tags. For example, in an area with tags of two different vendors, by virtue of their having different start commands, only the tags compatible with the reader will understand the command and respond. If both tags were TTF, they may respond simultaneously and interfere with each other's communications.

Commands, like anti-collision protocols, influence both performance and implementation. More commands may mean greater performance, but would likely require more circuitry. At the very least, the command set should include a start command, and those commands necessary to run the anti-collision protocol and achieve the necessary performance. Issues in performance of particular importance include identification rate and data integrity.

Additional commands may be necessary for additional applications and functionality. For ultimate low-cost read-only tags though, the only functionality should be to successfully transmit the identification code.

6.3.1 Commands and Bandwidth

As we saw in the discussion on coding and modulation, the probability of transmitting a component of a signal, say zero amplitude for a given clock period versus full amplitude for half of a clock period followed by zero amplitude for the second half of a clock period, will affect the bandwidth of a signal. If the probability of transmitting no modulations is higher than the probability of transmitting half period modulations, the bandwidth will be lower. This effect is most useful when considering the reader, as its signal power is maximized to meet field strength and power regulations. Consequently its coding, modulation, and transmissions will be limited by bandwidth regulations. In situations where bandwidth relative to power is very limited, this must be considered.

6.4 Summary

In this chapter we have considered the issues related to command protocols and anti-collision algorithms. As the goal in passive RFID systems is simply to read the identification codes stored on the tag, and multiple tags will be present in the interrogation range of a reader, command protocols should be optimized around the anti-collision protocols. In considering anti-collision protocols, important issues are not only identification rates, but also required command set, performance in the presence of noise, and implementation details. In addition, affects on bandwidth can also be extremely important.

Chapter 7

INFLUENCE OF CONSTRAINTS ON PERFORMANCE SPECIFICATIONS

7.1 Introduction

Thus far we have analyzed the various constraints on RFID systems. We have seen the main components that influence the final design and performance. In this chapter, I will attempt to bring these various constraints together so as to show how they interact with each other and influence and are influenced by specifications on performance, cost, and size.

The performance specifications I outlined in Chapter 1 include, range, speed, communications integrity, and compatibility. Each individual specification is related either directly or indirectly to all constraints. Cost and size are overriding concerns, which should be viewed, in essence, when examining all of the performance specifications against all of the constraints. Most of the issues related to cost and size were covered in Chapter 6. I will reiterate them when appropriate.

7.2 Range

In essence, range is constrained by the radiated field strength and electromagnetics. Radiated field strength is directly limited by regulations, and through regulatory and hardware bandwidth constraints, influenced indirectly by communications. Very quickly we move through field strength, three-dimensional orientation and position of the tag relative to the reader, environmental influences, and finally reception, delivery, and consumption of power by the tag. I will discuss these influences and their constraints in this section.

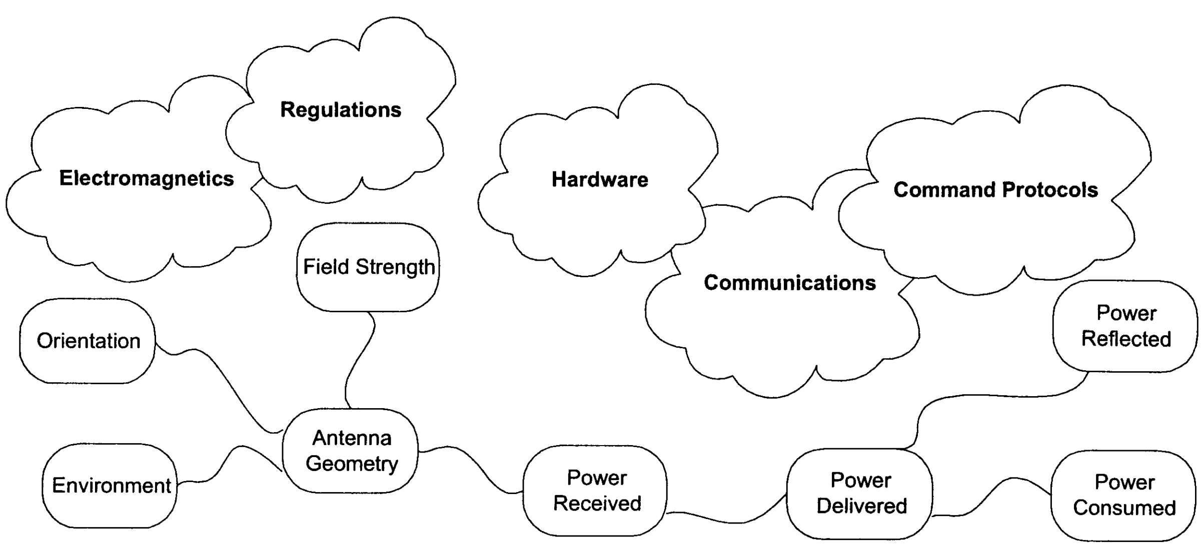


Figure 7.1: Influences on range and their overriding constraints.

7.2.1 Field Strength, Orientation, and the Environment

Range is often specified as a maximum quantity; however, this has use to only limited applications. In many applications, location and orientation will be relatively unconstrained. We saw in Chapter 2 that field strengths are highly orientation and location dependent. Depending on the geometry of the antenna, radiation/field patterns can vary radically. For example, very directional antennas can be constructed at both near-field frequencies and far-field frequencies. At the near-field frequencies, ferrite can be incorporated into the core of the coil, whereas in the far-field, antennas can be made more directional through arrays or more advanced geometries. With such directional patterns, greater range can be achieved for a given power radiated; yet the readable locations are obviously far more limited. Some applications may require highly directional characteristics whereas others may require omni-directional characteristics.

If regulations were to simply specify the maximum radiated power (i.e. EIRP), then highly directional antennas could be used. Limits, however, are typically imposed at maximum field strengths at certain distances or antenna input powers and antenna gains. Such is the case, in truth, with regulations of most administrations. Under certain administrations, however, at certain frequencies, more highly directional antennas may be used. For example, for frequency hopping use at 902 MHz, 2.45 GHz, and 5.8 GHz in the United States, radiated power is specified in terms of antenna input power and antenna gain. It is also specified that antenna gain may be increased above the specified limit, but that antenna input power must then be reduced by the same factor, effectively establishing a limit on field strength and power density at a given distance, or EIRP/ERP. The situation in Europe is similar for these frequencies – there limits are expressed in terms of ERP or EIRP. In Japan, however, limits are typically imposed directly on the antenna gain and input power. Though there is an associated EIRP, one must not increase the antenna gain or input power above the specified limit. In cases where limits are on EIRP, though allowable, increased directionality will not help increase the maximum range. It will however limit the size of the interrogation field, meaning perhaps less interference, fewer tags in the field, and likely more efficient anti-collision. A directional pattern, along with controlled variations of field strength and power (below regulations), allows a further degree of freedom in controlling the field spatially, possibly allowing for improved anti-collision.

Thus, we see that regulations essentially determine field strength and power density at a specific location, even though the electromagnetics allows tremendous flexibility. A problem, however, is that regulatory measurements of field strength and power densities at certain distances are maximum levels. Once a reader is placed in an unconstrained environment with surrounding conductors, patterns can change tremendously. This affect must be taken into account when considering specific applications and configurations.

7.2.2 Power Reception, Delivery, and Consumption

Given some field strength or power density at a specific location, the problem next becomes receiving power. For maximum induced voltage and supplied power, the tag's antenna must be oriented optimally to the transmitted field. This, of course, depending on the application, may be an extremely difficult task. Perhaps if tagged objects are moving by on a conveyor in constrained orientations, maximum available power may be received. For many applications, however, this will not be the case. Depending on the geometry and characteristics of the tag's antenna, the tag will be more or less sensitive to orientation. In addition, just as reader antennas may incorporate directionality, so to can tag antennas. Special antenna configurations such as orthogonal loops, can be used to reduce orientation nulls; however, orientation considerations must still be carefully weighed on the basis of the application and configuration.

Field strength and power density at a location is determined by regulations and environmental factors. Field strength and power density available for capture by a tag is affected by its antenna geometry and orientation. Given some antenna geometry and orientation, the power delivered to the tag's circuitry becomes a function of the tags circuit parameters. Impedance of antennas, matching circuits, and loads are particularly important. Together, they determine current through the various portions of the circuit based on an induced voltage. Induced voltage can be maximized by using various techniques specific to the frequency of operation.

In the near-field and far-field, different techniques are used. With a reactive link, the tag acts as a sink. Through resonance at both the tag's and reader's antennas, the tag can, in a sense, widen its drain. The use of resonance, by combining a capacitor in parallel with the coil of an inductive tag, and combining a capacitor in series with the coil of an inductive reader, can establish a high quality factor. The impedance of the combined resonant circuit and the equivalent impedance of the rest of the tag's circuitry determine this quality factor. The quality factor is defined as both the ratio of energy stored to energy dissipated, and the ratio of center frequency to 3 dB bandwidth. Thus, through this quality factor there is an associated filter. This filter is either equivalent to or narrower than the filter of the tag's antenna itself. At high quality factors, bandwidth is narrow and care must be taken not to filter out information. We have seen that modulation, or the imparting of information onto a carrier signal, creates sidebands. If we filter these out, we lose information. High quality factors also create sensitivity to environmental influences. Surrounding antennas or metals can change the resonant frequency and effectively shift the Q filter of center. In situations where tags will be collocated with other tags, a common strategy is to tune tags to higher than the nominal frequency of operation.

In the far field, the situation is different. Power delivered to a load is determined by the degree of matching between the antenna impedance and the equivalent impedance of the rest of the circuit. If there is a conjugate match between these two impedances, meaning the real portions are equal and the reactive portions are equal and opposite, maximum power will be delivered. If there is a mismatch, power will be reflected in the circuit and re-radiated by the antenna. This, of course, is the basis of backscatter modulation (Chapter 2). Controlled variation of the load impedance causes controlled modulations of the incident carrier signal. The more power scattered, the less absorbed. The more power absorbed, the less scattered.

Given some power delivered the load, it next becomes important how the "load" is configured. The power delivered is used not only for powering the tag, but also for providing information, a clock, and the establishment of biasing. Regarding configuration of the load, diodes are necessary to rectify the signal for both power supply and information extraction. Energy is stored, managed, and delivered to the digital circuitry. In the digital circuitry, the decision-making occurs. After reception and processing, typically through envelope detection, functions are performed and decisions are made based on the information. Tags may compare received information with stored information, decide to transmit stored information, mute themselves, or any number of other functions. These functions, in turn, require power. Power consumption is determined through the activity of the individual gates, the capacitance associated with the integrated circuit, and the supply voltage. The gate activity is determined by clock frequency and demands placed largely by anti-collision functions. These circuits also have an associated threshold voltage below which they will not function. The power delivered to the load must provide this threshold voltage and power necessary to allow the tag to achieve its functions. If these functions cannot be performed, the tag must be moved closer to the field source where field strength and power will increase relative to distance.

It is not just power delivered for powering the tag that is important, but also power available for detection. If there is not a suitable difference in signal levels, a clearly defined decision threshold, or low signal power relative to noise power, detection may be difficult. Closer range, where power is higher relative to noise, allows a more clearly defined decision threshold and subsequently a lower bit error rate (BER).

Thus far, we have just considered the link from reader to tag. We should also consider the link from tag to reader. In this link, only modulation of information occurs. This is accomplished through either load modulation or backscatter modulation. Modulation levels are relatively low compared to the reader-transmitted modulations. However, because of the much more robust and complex signal processing available to the reader, this is typically not the limiting factor in range.

Thus, we see that it is not any one factor that determines range, but many factors. Clever design can exploit these factors. Negative affects may be reduced while positive affects may be increased. Given the constraints of electromagnetics, regulations, communications, antenna and circuit design and implementation, and anti-collision protocols, there are still many degrees of freedom in the design. Range is consequently an extremely variable factor. Specification of range must consider configuration, application, and RFID system specific issues.

7.3 Speed and Integrity

The other obviously important specification is speed of identification. The importance of speed, varies with the application and configuration. In retail automatic purchase applications with few readers, high identification rates are a necessity, in low traffic shelf-based inventory management applications, requirements may not be so stringent. Speed of identification can be further divided into two components: data rate, and identification rate. Data rate is the actual speed at which bits can be transmitted, while identification rate refers to the performance of an anti-collision protocol in identifying individual tags.

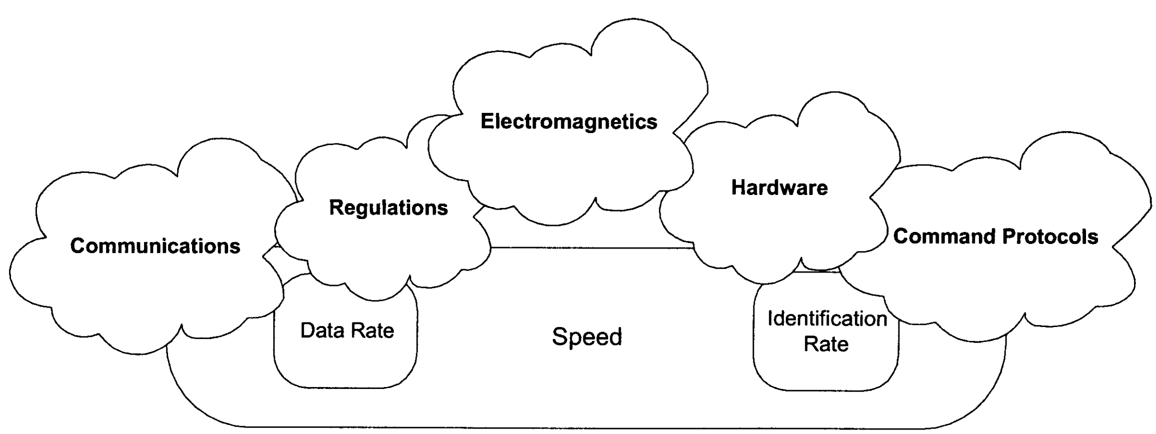


Figure 7.2: The factors that determine speed and the constraints that influence them.

7.3.1 Data Rate

Data rate is essentially a function of the bit or pulse period. Smaller periods mean faster rates. A necessary consequence of a smaller bit period is a larger bandwidth, as we saw in Chapter 3. Bandwidth is limited by two main factors, regulations, and hardware limitations. Hardware limitations involve filtering inherent in the tag's antenna and attached circuit. These factors can be especially limiting for the near-field systems relying on resonance and a high quality factor. A high quality factor indicates a narrow bandwidth and thus an upper bound on the data rate. Far-field system tag antennas also have an associated bandwidth. These hardware limitations are typically not as constraining as regulations, however.

Data rate is not necessarily a function of operating frequency, but rather the bandwidth available by regulations at that frequency, and further by the bandwidth limitations on the sidebands. By virtue of the organization of the spectrum, there is generally more available bandwidth at higher frequencies. Operating frequency typically defines the derivation source of the tags circuit clock frequency. At the near-field frequencies, tags are typically synchronous and the clock on the tag is typically divided directly from the carrier frequency. At far-field frequencies, tags typically derive the clock from a reader-transmitted modulation. Some systems my incorporate phase locked loops and oscillators, while others may constantly use an edge of the reader transmitted pulse to continuously define the clock. In any case, regardless of derivation, a faster clock means faster processing, yet increased power consumption. Thus, clocks are typically divided down to similar rates between both near-field and far-field systems. It is the bandwidth, thus, that determines speed. More particularly, in RFID systems, it is the bandwidth of the reader's signal that determines data rate of the reader to tag transmission, and depending on configuration, the data rate of the tag's clock and transmissions back to the reader.

Bandwidth is typically given by the main channel frequency span. There is also a single, or multiple associated sideband levels. Depending on the administration and operating frequency, regulations are defined differently. If we consider Europe's regulations at 13.56 MHz, the main channel bandwidth is given by the ISM designation of +/- 7 kHz. Sidebands are given at two levels. A level less than the main band limit, but higher than the lowest level, extends to +/- 150 kHz. Outside of this band is a further reduced strength level where general spurious emissions are allowed. In the US, there is simply the ISM designated +/- 7 kHz, and a low field strength level designated for general emissions. Japan further uses a different method. There, bandwidth is given as 7 times the bit rate. These regulatory issues are, of course, not only directly related to speed, but compatibility.

We should also consider the bandwidth of tag-modulated signals. Because the field strength levels of these signals are significantly lower than those of readers, they fall below regulatory limits. Bandwidth from the tag to the reader is thus limited by hardware considerations. Because readers are stationary, they may operate at the maximum capacity allowed by their administration, yet different coding and modulation schemes may be necessary to achieve similar performance between the different regions. Such is an approach adopted by some manufacturers. If tags cannot easily detect both methods, this presents a problem for compatibility.

We saw in Chapter 3 how coding and modulation influences bandwidth and bit rate. Further we saw that certain codes allow signals that contain more energy than others. In the link from reader to tag, a signal that contains more energy is desirable, as it not only must carry information, but also power. Thus, in this respect, speed and range are coupled. If hardware considerations were of concern, we should note that for a given coding and modulation scheme, increasing the bit rate would result in an increase in bandwidth. Given the hardware bandwidth limitations, this would result in less power to the tag. This represents further coupling between speed and range.

7.3.2 Identification Rate

Given some bit rate, the anti-collision protocol becomes important for determining speed. Essentially it defines a tag identification rate per bit rate efficiency. Given a bit rate R, and a identification rate, T, we can define the protocol efficiency by

$$\varepsilon_p = \frac{T\ [\frac{\text{tags}}{\text{s}}]}{R\ [\frac{\text{bits}}{\text{s}}]} \tag{7.1}$$

The better anti-collision schemes will have a higher efficiency. Anti-collision protocols, of course, influence the hardware design, which, in turn, noticeably affect cost. This is an important tradeoff – identification rate, versus cost. In addition, they affect power consumption – the greater activity demanded, the more power consumed. As noted in Chapter 6, there are three domains and several types of anti-collision within those domains. For-low cost passive RFID systems, the most suitable are those in the time domain where tags are either identified deterministically based on identification code, or probabilistically through random response times. Each has their inherent advantages and disadvantages as outlined. One important advantage of deterministic schemes is the ability to, in essence, move the anti-collision to the reader, host, and associated infrastructure by incorporation of intelligence in these units. Depending on the application and configuration, particular tags may need to be identified. Deterministic schemes can selectively limit this identification. Probabilistic schemes typically don't have this inherent advantage. They may include it, but only at the cost of extra devices on chip, extra complexity, size, and cost.

It also should be noted that anti-collision schemes can affect the data rate of the transmissions. Those requiring infrequent transmissions from the reader will have a lower associated bandwidth and thus may increase their data rate. Probabilistic schemes may have an advantage in this arena as they may require fewer transmissions from the reader. Closely related to anti-collision scheme is the integrity of the transmissions. The reader to tag fidelity is higher due to the much greater signal level. Tag to reader fidelity is much lower, and thus those protocols that require fewer transmissions from the tag to reader may be more reliable. As detailed in Chapter 4, some schemes allow only simple acknowledgements from the tag. This too, though, has its problems, as an error on this response means a failed transmission.

Data integrity is strongly related to both range and speed. Error detection or correction may be implemented; however this can reduce speed, require more complex hardware, require more power consumption, and hence reduce range. As mentioned previously the closer to the reader, the greater the field strength, a more easily discernable signal, and hence increased data integrity.

Thus we have seen that both data rate and anti-collision schemes affect the speed of operation. Anti-collision also affects the data rate. Both the anti-collision scheme and the data rate in combination with the coding and modulation scheme affect the power consumption of the tag. This of course, affects the range as discussed previously. Both data rate and anti-collision are additionally extremely important on the basis of data integrity and compatibility with worldwide regulations.

7.4 Compatibility and Standardization

As mentioned in the introduction, compatibility between different regulatory administrations, and between RFID systems of different vendors is important for the continued adoption and proliferation of RFID. I have dealt with the regulatory issues above. Here I will outline some of the important concerns related to standardization for compatibility between RFID systems of different vendors.

In general, there are two approaches one can take with respect to RFID standardization. Systems may be completely incompatible and orthogonal in coding, modulation, and commands and protocols. Or, they may be completely compatible. In either case, the reader-talks-first systems are recommended as tags default to a mute state and are only activated when specifically commanded by the reader. If these start commands are completely different and the probability of either inadvertently transmitting the command is low, either through different coding and modulation schemes, or a start command of sufficient length, then one need not worry about the additional protocols of the system. Tags will simply not interfere. If it is desired that a reader communicate with both types of tags, then the reader should accommodate for both coding and modulation schemes, and associated command protocols. Systems that are completely compatible must have identical coding and modulation schemes and command protocols. Certain commands may be omitted from certain tags, yet if omitted, they must be ignored.

Both incompatible and compatible methods for achieving compatibility, or, in essence, lack of interference, are feasible. Incompatible methods allow complete freedom of design to different vendors, yet complicate matters for the end user. If one wishes to have a reader capable of reading tags of multiple vendors, they must install readers capable of the different schemes and protocols. This can be more costly, yet with advances in software-defined radio and signal processing, this is becoming less of an issue. Further, some end users may only wish to read tags of a particular vendor. In this case the incompatible method does not present a problem. The compatible method allows for inexpensive readers and simplicity for the end users, yet impose stringent constraints on the designer. Innovation can be limited, and competition leading to decreased costs, product differentiation, and further innovation can be further harmfully limited.

Operating frequency is another important issue related to compatibility. Obviously those tags operating at different frequencies are completely orthogonal and present virtually no chance of interference. However, for end users desiring to read tags at multiple frequencies, readers must additionally have this capability. There have been efforts to designate frequencies for particular standards, however this is not recommended. Electromagnetic field and wave behavior at a given frequency can differ drastically depending on the environment for a given application. A given frequency may very well work flawlessly in one situation, yet in a different situation, even though for the same or similar application, may not work well. Other frequencies may be more appropriate. Much testing still remains to determine behavior of systems operating at different frequencies. Selection of a single frequency, even if for only a single application, may be premature. Decisions of this nature should be made on an application and configuration specific basis.

In any regard, standardization for compatibility is a sensitive issue and can have drastic affects on the proliferation and innovation of RFID technology. Achieving compatibility by complete incompatibility through reader talks first systems and completely orthogonal initialization commands, is a viable approach. Compatibility through complete compatibility may also be an option. Both methods may work in parallel, leaving the choice up to the designer and end user. As each operating frequency has its own inherent advantages and disadvantages in a particular environment, designations of specific frequencies, in general, should not be made.

7.5 Summary

In this chapter we have seen how the important performance specifications are constrained by electromagnetics, communications, regulations, hardware design, and command protocols. We have seen that the specifications are dependent both directly and indirectly on all of the fundamental constraints. We have also considered the dependence of these specifications and constraints on applications and configurations. Design of tags and standards should reflect the high degree of coupling between specifications, constraints, and desired application and configuration.

Chapter 8

CONCLUSION

8.1 Conclusion

In the recent past there has been increased adoption of RFID systems for a variety of applications requiring ever-lower costs. In parallel, there has been increased activity in RFID standards development. Already established standards include those for applications ranging from animal tagging, through highway vehicle identification. Currently, there are ongoing efforts to develop standards for item management and logistics applications. Perhaps the most demanding near term application for item management is tagging of low cost consumable items. This trend towards tagging such low-cost items on a global scale places stringent demands on the RFID systems themselves. In addition to performance related concerns, size, and particularly cost are extremely important. Absolute minimum cost tags are a necessity.

To realize such low cost tags, design must be optimized specifically for this purpose. An understanding of the fundamental constraints on the systems, along with clever design, is imperative. The fundamental constraints on these low-cost passive RFID tags include electromagnetics, communications, regulations, and physical implementation. In this thesis I have analyzed these various constraints on the basis of low-cost passive RFID. We have seen how these constraints affect cost, size, and the various important performance specifications including range, speed, communications integrity, and compatibility.

With low cost tags, we will see increased adoption, and greater volumes, which will further feed decreases in cost and a continuation of the cycle. Improvements and innovations in manufacturing techniques should allow for more efficient production and lower cost. Furthermore, increased adoption will likely lead to incorporation of additional functionality, such as inclusion of sensors, actuators, and other more advanced functions. In addition, with increased adoption, we will see the development of additional applications exploiting the automatic capture and instant machine knowledge of information about objects. Increased automation should proliferate, driving increased productivities and efficiencies.

8.2 Future Work

In studying the fundamental constraints on RFID systems and their current state of development, a number of areas deserving of future work emerge. I will briefly discuss some of these areas here.

As we have seen, power available to the tag is the limiting factor on its range. Though range is not the most important specification for all applications, it is nevertheless important to many. Currently, readers supply power to the tag through their transmitted field. Through technologies known as "energy harvesting," energy can be collected from ambient sources. This may allow increased range and cause the limiting factor on range to be receiver sensitivity for communications. Currently, much work is ongoing in energy harvesting. MEMS transducers that convert mechanical vibrations to electric energy, shoes which harvest energy while walked in, and integrated circuitry that converts dissipated heat to electric energy, have all been fabricated [82][83][67]. RFID tags may use similar schemes, or improve collection of ambient electromagnetic radiation.

In the area of electromagnetics, testing and design for specific environments is an area deserving of attention. In logistics applications, environments typically include a great deal of metal. As we have seen this can cause problems in altering field and radiation patterns, and thus affect overall performance significantly. Antennas allow many degrees of freedom in their configuration, thus improved geometries should be studied.

Though much work has been done in the area of communications related to optimization of coding and modulation, there are still areas that should be addressed. At most frequencies, there are differences in bandwidth and field strength limitations between different regions. Short of worldwide regulation harmonization, coding, modulation, and filtering techniques, which optimize systems for such operation, should be studied. In addition, design of inexpensive FSK and PSK detectors, and improvements in clock generation techniques may be useful. Space and frequency domain anti-collision schemes should also be examined.

In the area of physical implementation, novel tag antenna structures should be investigated. Fractal antennas may be promising as they can allow for reduced size while maintaining reception and radiation performance [84]. Study should also be done on exploiting the packaging of tagged objects. Packaging of certain objects may allow for integrated antenna structures or energy storage. Manufacturing techniques, particularly for fabricating and assembling the entire tag structure, including integrated circuit and antenna, should also be investigated.

Chipless tags, in which there is currently much ongoing work, should continue to be studied. Range, memory and anti-collision techniques, in particular, should be examined.

Though the focus of this thesis has been on tags, there is much to be studied in the area of readers. Reader-to-reader anti-collision techniques, in particular should be examined. As RFID systems proliferate, reader density will increase and the likelihood of interference between readers will subsequently increase. Such interference can be especially harmful to the integrity of communications.

Finally, additional functions and new applications should provide a number of interesting and compelling areas of research. Integration of sensing and actuating technology, either through MEMS devices, or simpler material structures, would be useful to a number of current and likely new applications. New applications related to robotics and automation should be studied. The identification code from the tag, in combination with a networked database allows simple perception and instant knowledge.

REFERENCES

[1] Yogesh Joshi, "Information Visibility and Its Effect on Supply Chain Dynamics," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2000, <http://auto-id.mit.edu/research/whitepapers.html>.

[2] MIT Auto-ID Center, "The Networked Physical World: Proposals for engineering the next generation of computing, commerce, and automatic-identification," MIT-AUTOID-WH-001, 2000, <http://auto-id.mit.edu/research/whitepapers.html>

[3] Jim Eagle, "RFID: The Early Years 1980-1990,"2001, <http://members.surfbest.net/eaglesnest/rfidhist.htm>

[4] Alfred R. Koelle., Steven W. Depp, Jeremy A. Landt, and Ronald E. Bobbett, "Short-Range Passive Telemetry by Modulated Backscatter of Incident CW RF Carrier Beam," *Biotelemetry*, Vol. 3, , Academic Press, New York, pp. 337-340, 1976.

[5] Thierry Roz and Vincent Fuentes, "Using Low Power Transponders and Tags for RFID Applications," EM Microelectronic Marin SA, Marin, Switzerland, <http://www.emmarin.ch/>.

[6] Klaus Finkenzeller, *RFID Handbook*, John Wiley and Sons, Inc., New York, 1999.

[7] John R. Tuttle, "Traditional and Emerging Technologies and Applications in the Radio Frequency Identification (RFID) Industry," *IEEE Radio Frequency Integrated Circuits Symposium*, 1997.

[8] S. Shuji, H. Yuzo, A. Dobashi, M. Okumura, and T. Kusuzaki, "Products Lifecycle Management System Using Radio Frequency Identification Technology," *Proceedings of the 7th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA),* Vol. 2, 1999, pp. 1459-1467.

[9] David L. Brock, "The Electronic Product Code," MIT-AUTOID-WH-002, 2001, <http://auto-id.mit.edu/research/whitepapers.html>.

[10] Peter J. Harrop, *Chipless Smart Labels: The Ultimate RFID,* IDTechx, Cambridge, UK, 1988.

[11] N. Komatsu, T. Kon, H. Mukai, K. Kamigama, "Research and Study of Elementary Technology for Achieving Greater Efficiency in Postal Processing: The Present State of RFID and their Applications to Postal Services," IPTP Monthly Report, No. 119, 1998.

[12] David H. Staelin, Ann. W. Morgenthaler, and Jin Au Kong, *Electromagnetic Waves*, Prentice Hall, New Jersey, 1994.

[13] Warren L. Stutzman and Gary A. Thiele, *Antenna Theory and Design*, 2nd ed., John Wiley and Sons, Inc., New York, 1998.

[14] David K. Cheng, *Field and Wave Electromagnetics*, 2nd ed., Addison-Wesley Publishing Co., New York, 1992.

[15] Richard Ribon Fletcher, "A Low-Cost Electromagnetic Tagging Technology for Wireless Identification, Sensing, and Tracking of Objects," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 1997.

[16] *IEEE Standard Dictionary of Electrical and Electronics Terms*, 3rd ed., ANSI/IEEE Standard 100-1984.

[17] Eugene F. Knott, John F. Shaeffer, and Michael T. Tuley, *Radar Cross Section*, 2nd ed., Artech House, Boston, 1993.

[18] R. C. Hansen, "Relationships Between Antennas as Scatterers and as Radiators," *Proceedings of the IEEE*, Vol. 77, No. 5, pp. 659-652, 1989.

[19] Robert Blair Green, "The General Theory of Antenna Scattering," Ph.D. Thesis, The Ohio State University, Columbus, OH, 1963.

[20] Alan Demers, Scott Elrod, Christopher Kantarjiev, and Edward Richley, "A Nano-Cellular Local Area Network Using Near-Field RF Coupling," Xerox Corporation, Palo Alto, 1994.

[21] Theodore S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, New Jersey, 1996.

[22] Jorgen Bach Andersen, Theodore S. Rappaport, and Susumu Yoshida, 1995, "Propagation Measurements and Models for Wireless Communications Channels," *IEEE Communications Magazine*, January, 1995.

[23] John B. Anderson, *Digital Transmission Engineering*, IEEE Press, New York, 1999.

[24] Leon W. Couch, *Digital and Analog Communications Systems*, 6th ed., Prentice-Hall, New Jersey, 2001.

[25] Joseph L. LoCicero and Bhasker P. Patel, "Line Coding," *The Mobile Communications Handbook*, Gibson, Jerry D., ed., CRC Press, Boca Raton, Florida, 1999, pp. 6-1 – 6-18.

[26] Hwei P. Hsu, *Probability, Random Variables, and Random Processes*, McGraw-Hill, New York, 1997.

[27] William C. Lindsey, and Marvin K Simon, *Telecommunication Systems Engineering,* Prentice-Hall, New Jersey, 1973.

[28] R.C. Titsworth and L. R. Welch, "Power Spectra of Signals Modulated by Random and Pseudorandom Sequences," Jet Propulsion Laboratory, Pasadena, CA, Technical Report No. 32-140, October 1961.

[29] H. Hecht and A. Guida, "Delay Modulation" *Proceedings of the IEEE*, Vol. 57., No. 7, July 1969, pp.1314-1316.

[30] E. N. Gilbert, "Capacity of a Burst-Noise Channel," *Bell System Technical Journal*, Vol. 39, pp. 1253-1266, September 1960.

[31] E. O. Elliott, "Estimates of Error Rates for Codes on Burst-Error Channels," *Bell System Technical Journal*, Vol. 42, p. 1997, September 1963.

[32] Michele Zorzi and Ramesh R. Rao, "Error Control Strategies for the Wireless Channel," IEEE 5th *International Conference on Universal Personal Communications*, Vol. 2, 1996, pp. 970-974.

[33] J. R. Yee and E. J. Weldon, "Evaluation of the Performance of Error-Correcting Codes on a Gilbert Channel," *IEEE Transactions on Communications*, Vol. COM-43, August 1995, pp. 2316-2323.

[34] Mischa Schwartz, *Information, Transmission, Modulation, and Noise*, McGraw-Hill, New York, 1980.

[35] David Withers, *Radio Spectrum Management: Management of the Spectrum and Regulation of Radio Services*, 2nd ed., The Institution of Electrical Engineers, London, 1999.

[36] ITU RR, 1998, Radio Regulations, Vol. 1,. International Telecommuications Union.

[37] ERO, European Radiocommunications Office, <http://www.ero.dk>.

[38] ETSI, European Telecommunications Standards Institute, <http://www.etsi.org>.

[39] FCC, Federal Communications Commission, <http://www.fcc.gov>.

[40] MPT, Ministry of Post and Telecommunications, <http://www.joho.soumu.go.jp>.

[41] ARIB, Association of Radio Industries and Businesses, <http://www.arib.or.jp>.

[42] ITU-R 213/1, Technical and Operating Parameters and Spectrum Requirements for Short-Range Radiocommunication Devices, 2001.

[43] ITU-R SM.1056, Limitation of Radiation for Industrial, Scientific and Medical (ISM) Equipment, 1994.

[44] CEPT 70-03, Relating to the Use of Short Range Devices (SRD), May 2001.

[45] ETSI EN 300330-1 V1.3.1 (2001-04), Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment in the frequency range 9 kHz to 25 MHz and inductive loop systems in the frequency range 9 kHz to 30 MHz; Part 1: Technical characteristics and test methods, 2001.

[46] FCC Part 15, Radio Frequency Devices, 2001.

[47] David Mawdsley, EMC Conducted Emissions Measurement: Detectors... what, why and when?," *Technical Notes from Laplace Instruments*, 2001, <http://perso.wanadoo.fr/dicomtech/article_lap_2.htm>.

[48] CISPR 16-1, Specification for Radio Disturbance and Immunity Measuring Apparatus and Methods – Part 1: Radio Disturbance and Immunity Measuring Apparatus, International Electrotechnical Commission, 1999.

[49] ETSI EN 300220-1 V.1.3.1 (2000-09), Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Short Range Devices (SRD); Radio equipment to be used in the 25 MHz to 1000 MHz frequency range with power levels ranging up to 500 mW; Part 1: Technical characteristics and test methods, 2000.

[50] ETSI EN 300440-1 V.1.3.1 (2000-07), Electromagnetic Compatibility and Radio Spectrum Matters (ERM); Short Range Devices; Radio equipment to be used in the 1 GHz to 40 GHz frequency range; Part 1: Technical characteristics and test methods, 2000.

[51] John Burns, Richard Rudd, and Zoran Spasojevic, "Compatibility Between Radiocommunication and ISM Systems in the 2.4 GHz Frequency Band," Aegis Systems Limited, 1999.

[52] Acheson Colloids Company, "Product Selection Guide," <http://www.achesoncolloids.com/>.

[53] Ramesh Garg, Prakash Bhartia, Inder Bahl, Apisak Ittipiboon, *Microstrip Antenna Design Handbook*, Artech House, Boston, 2000.

[54] D. Friedman, H. Heinrich, D-W. Duan., "A Low-Power CMOS Integrated Circuit for Field-Powered Radio Frequency Identification Tags," *IEEE International Solid-State Circuits Conference*, 1997, p. 294.

[55] Ulrich Kaiser and Wolfgang Steinhagen, "A Low-Power Transponder IC for High-Performance Identification Systems," *IEEE Journal of Solid-State Circuits*, Vol. 30, No. 3, pp. 306-310, March 1995.

[56] Stephane Donnay, Philip Pieters, Kristof Vaesen, Wim Diels, Piet Wambacq, Walter De Raedt, Eric Beyne, Marc Engels, and Ivo Bolsens, "Chip-Package Codesign of a Low-Power 5-GHz RF Front End," *Proceedings of the IEEE*, Vol. 88, No. 10, pp. 1583-1597, October 2000.

[57] M.H. Brodsky, Progress in Gallium Arsenide Semiconductors. *Scientific American*, pp.68-72, February 1990.

[58] J.D. Meindl, Chips for Advanced Computing. *Scientific American*, pp. 78-88, October 1987.

[59] Lawrence E. Larson, "Integrated Circuit Technology Options for RFIC's – Present Status and Future Directions," *IEEE Custom Integrated Circuits Conference*, 1997, pp. 169-176.

[60] Thomas H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, Cambridge University Press, Cambridge, UK, 1998.

[61] Agilent AN 1089, "Designing Detectors for RF/ID Tags," Agilent Application Note, 1999.

[62] Paul Horowitz and Winfield Hill, *The Art of Electronics*, 2nd ed., Cambridge University Press, Cambridge, UK, 1989.

[63] Eric A. Vittoz, "Micropower IC," *Low-Power CMOS Design*, Anantha Chandrakasan, Robert Brodersen, ed, IEEE Press, New York, 1998.

[64] Chenming Hu, "Device and Technology Impact on Low Power Electronics," *Low Power Design Methodologies*, Kluwer Academic Publishers, Boston, 1996, pp. 21-33.

[65] Jan M. Rabaey, Massoud Pedram and Paul Landman, "Introduction," *Low Power Design Methodologies*, Kluwer Academic Publishers, Boston, 1996, pp. 1-15.

[66] P. Vanoostende, et al., "Evaluation of the limitations of the simple CMOS power estimation formula: comparison with accurate estimation," *Proceedings of the European Workshop on Power and Timing Modeling*, 1992, pp. 16-25.

[67] Athas, William C., "Energy-Recovery CMOS," *Low Power Design Methodologies*, Kluwer Academic Publishers, Boston, 1996.

[68] Benny Bing, *Broadband Wireless Access,* Boston, Kluwer Academic Publishers, MA, 2000.

[69] Peter Hawkes, "Anti-collision and Transponder Selection Methods for Grouped "Vicinity" Cards and RFID tags," *IEE Colloquium on RFID Technology*, 1999, pp. 7/1-7/12.

[70] Josef Schuermann, U.S. Patent: System and method for reading multiple RF-ID transponders, Patent Number: US5500651, March 1996.

[71] Robert Fleming, Cherie Kushner, "Integrated, Low-Power, Ultra-Wideband Transceivers for Distributed Position Location and Communication," Aether Wire & Location, Inc., January 1998, <http://www.aetherwire.com/Aether_Wire/Technical_Abstracts_97.pdf>.

[72] Konstantin O. Aslanidis, Alain Berthon, U.S. Patent: Method for repeating interrogations until failing to receive unintelligible responses to identify plurality of transponders by an interrogator, Patent Number: US5929801, July 1999.

[73] Ching Law, Kayi Lee, Kai-Yeung Siu, "Efficient Memoryless Protocol for Tag Identification," In *Proceedings of the 4th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, August 2000.

[74] Marcel Jacomet, Adrian Ehrsam, Urs Gehrig, "Contactless Identification Device With Anticollision Algorithm," *IEEE Computer Society, CSCC'99, Conference on Circuits, Systems, Computers and Communications*, 4-8 July 1999 Athens, <http://www.microlab.ch/academics/r_and_d/research/cscc/cscc99.pdf>.

[75] Ronald Walter and Keith Vertrees, U.S. Patent: Method and apparatus for radiofrequency identification tags, Patent Number: US5856788, January 1999.

[76] Shigeru Furuta, U.S. Patent: Method and system for identifying and communicating with a plurality of contactless IC cards, Patent Number: US5698837, December 1997.

[77] EM Microelectronic-Marin SA, P4022: Multi Frequency Contactless Identification Device, <http://www.emmarin.com/pdf/RFID/P4022_C.pdf>.

[78] Microchip Technology Inc., microID 125 kHz Design Guide, <http://www.microchip.com/Download/lit/pline/rfid/guides/51115e.pdf>.

[79] Donald L. Black, Dale Yornes, U.S. Patent: Method for resolving signals collisions between multiple RFID transponders in a field, Patent Number: US6091319, July 2000.

[80] Pourang Mahdavi, European Patent: A system and method for communicating with multiple transponders, Patent Number: EP1017005A2, May 2000.

[81] John H. Bowers, John Nelson Daigle, Rade Petrovic, U.S. Patent: Anticollision protocol for reading multiple RFID tags, Patent Number: US5883582, March 1999.

[82] Scott Meninger, Jose Oscar Mur-Miranda, Rajeevan Amirtharajah, Anantha Chandrakasan and Jeffrey Lang, "Vibration-to-Electric Energy Conversion," *1999 International Symposium on Low-Power Electronic Design (ISLPED '99)*, July 1999, pp. 48-53.

[83] N. S. Shenck and J. A. Paradiso, "Energy Scavenging with Shoe-Mounted Piezoelectrics," *IEEE Micro*, Vol. 21, No. 3, May-June 2001.

[84] R. Breden and R. J. Langley, "Printed Fractal Antennas," *IEE National Conference on Antennas and Propagation*, 1999.